Molecular Coevolution among Cryptically Simple Expansion Segments of Eukaryotic 26S/28S rRNAs

John M. Hancock and Gabriel A. Dover
Genetics Department, University of Cambridge

The set of “expansion segments” of any eukaryotic 26S/28S ribosomal RNA (rRNA) gene is responsible for the bulk of the difference in length between the prokaryotic 23S rRNA gene and the eukaryotic 26S/28S rRNA gene. The expansion segments are also responsible for interspecific fluctuations in length during eukaryotic evolution. They show a consistent bias in base composition in any species; for example, they are AT rich in Drosophila melanogaster and GC rich in vertebrate species. Dot-matrix comparisons of sets of expansion segments reveal high similarities between members of a set within any 28S rRNA gene of a species, in contrast to the little or spurious similarity that exists between sets of expansion segments from distantly related species. Similarities among members of a set of expansion segments within any 28S rRNA gene cannot be accounted for by their base-compositional bias alone. In contrast, no significant similarity exists within a set of “core” segments (regions between expansion segments) of any 28S rRNA gene, although core segments are conserved between species. The set of expansion segments of a 26S/28S gene is coevolving as a unit in each species, at the same time as the family of 28S rRNA genes, as a whole, is undergoing continual homogenization, making all sets of expansion segments from all ribosomal DNA (rDNA) arrays in a species similar in sequence. Analysis of DNA simplicity of 26S/28S rRNA genes shows a direct correlation between significantly high relative simplicity factors (RSFs) and sequence similarity among a set of expansion segments. A similar correlation exists between RSF values, overall rDNA lengths, and the lengths of individual expansion segments. Such correlations suggest that most length fluctuations reflect the gain and loss of simple sequence motifs by slippage-like mechanisms. We discuss the molecular coevolution of expansion segments, which takes place against a background of slippage-like and unequal crossing-over mechanisms of turnover that are responsible for the accumulation of interspecific differences in rDNA sequences.

Introduction

The ribosomal DNA (rDNA) of eukaryotes is a large multigene family that undergoes continual turnover by rounds of unequal crossing-over (for reviews, see Dover 1982; Arnheim 1983; Flavell 1986). As a consequence of this, new variant repeats gradually spread through the family and through the population (molecular drive). The rDNA unit consists of regions with varying degrees of function. As such it serves as a useful model of the different degrees of interaction between the processes of molecular drive and natural selection. For example, the accumulation of differences in the region of transcription initiation has led to a molecular coevolutionary change

1. Key words: rDNA, 28S rRNA expansion segments, molecular coevolution, cryptic simplicity, microconversion.

Address for correspondence and reprints: Gabriel A. Dover, Genetics Department, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom.
in the genes and proteins of the RNA polymerase I complex, observed as an incompatibility between the rDNA of one species and the polymerase complex of another (for reviews, see Arnheim 1983; Reeder 1984; Gerbi 1985; Moss et al. 1985). It is probable that selection has been involved in such coevolutionary changes, as a response to the changing spectrum of promoter sequences (Dover and Flavell 1984). Both major and minor changes are observed in the promoters between species of Drosophila (Tautz et al. 1987).

To examine further examples of coevolutionary changes, we have compared the sequences of the "expansion" and "core" segments of the D. melanogaster 28S ribosomal RNA (rRNA) gene with those of several other species. We have shown in an accompanying paper that these two types of segment differ both in AT richness and in the degree to which slippage-like mechanisms have generated sequence simplicity (Tautz et al. 1988). These differences might be of functional significance, leading either to the selective conservation of sequences between species or to selective coevolutionary changes at the molecular level that are required to maintain cellular functions. The first outcome would be through the action of negative selection against any new non-neutral mutations that were being homogenized and fixed in the population, and the second outcome would be through the action of positive selection for compensatory changes in response to the homogenization/fixation of new, biologically effective mutations. We examine these possibilities below. In an accompanying paper we consider (1) the effects of sequence evolution of the D. melanogaster 28S rRNA genes on the RNA secondary structure and (2) the coevolutionary changes of compensatory mutations (Hancock et al. 1988).

Expansion segments account for almost all of the differences in length of 26S/28S rRNA genes among eukaryotic species (Clark et al. 1984). Although it has been suggested that expansion segments are functionless structures which are tolerated only because they do not interfere with ribosome function (Gerbi 1985), they nevertheless show interspecific conservation of gross secondary structure (Michot et al. 1984; Michot and Bachellerie 1987; Hancock et al. 1988), indicating that they are subject to some sequence constraint. Our analysis of 26S/28S rRNA genes shows (1) that expansion segments are an example of molecular coevolution between different parts of a single unit that is itself part of a multigene family and (2) that slippage-like mechanisms—and possibly a biased occurrence or fixation of mutations—have played an important part in their evolution.

Material and Methods

Sequence analysis was carried out using the DIAGON dot-matrix sequence comparison program (Staden 1982) and the SIMPLE DNA sequence simplicity program (Tautz et al. 1986; for further details, see the accompanying paper [Tautz et al. 1988]).

Artificial rDNA Sequences

To test whether base composition generated by processes of biased mutation (as opposed to direct repeats generated by slippage) is responsible for the sequence similarities between expansion segments revealed by the dot-matrix analysis (figs. 1, 3), we used two types of artificially randomized sequence for similar dot-matrix analysis. Random sequences were generated with a desired base composition by using a program (RANDSEQ) that selects bases randomly for each position within the sequence; the selection of bases was biased so that the final base composition was close to that
Molecular Coevolution of rRNA Expansion Segments

desired. The final base composition was then adjusted to the desired value by mutating the sequence at random positions, adding A+T or G+C as desired.

The first test consisted of a fully randomized sequence of the same overall base composition as a given 28S rRNA gene, in this case *Mus musculus* (fig. 2a). The second test used random blocks at 45% and 25% AT that were spliced together using the EMBL utility ZSEQ to give precisely the same patterns of core and expansion segments as are present in the *M. musculus* 28S rRNA gene.

Nomenclature

We have used the nomenclature originated by Michot et al. (1984) for the expansion segments of eukaryotic 26S/28S rRNAs.

Results

**DIAGON Analysis**

DIAGON plots were performed on the nuclear 26S/28S rRNAs genes of *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces carlsbergensis*, *Physarum polycephalum*, *Oryza sativa*, and *Escherichia coli* (fig. 1). The *E. coli* 23S rRNA gene (Brosius et al. 1980) was taken as a prokaryotic gene with no expansion segments (Clark et al. 1984). This served as one level of control over the significance of expansion-segment similarities.

Mouse, rat, frog, and human show very strong patterns of expansion-segment sequence similarity, whereas slime mold, yeast, nematode, and *E. coli* show none. Detailed mapping of the blocks of sequence similarity show their boundaries to correspond to the boundaries of the major expansion segments.

The 28S rRNA gene of *D. melanogaster* and the 25S rRNA gene of *O. sativa* are in an intermediate position in that similarities between expansion segments are not as strong as in the first four species; nor are the regions of localized high simplicity as clear-cut. A high-stringency (100%) DIAGON plot of the *D. melanogaster* 28S rRNA gene against itself, by using a 5-base span length, produced a pattern of inter- and intra-expansion-segment sequence similarity in a noisy background (not shown). To decrease the background relative to the expansion-segment pattern, the span length was increased systematically. An improved signal:noise ratio was obtained at increased span lengths, but only if the stringency of the match was decreased. At high stringency and long span length no pattern was observed. A 35-base span length at 54% stringency (19 of 35 bases) reveals a clear pattern of expansion segments in a low background (fig. 1).

**16S/18S rRNA Genes**

Neither *D. melanogaster* nor mouse 18S rRNA genes (Raynal et al. 1984; Tautz et al. 1988) show patterns of intragenic sequence similarities (data not shown).

**Base Composition**

Overall base compositions of the core and expansion segments of the eight available 26S/28S rRNA coding sequences are presented in table 1. Base compositions of individual expansion segments are presented in table 2. The four vertebrate sequences (human, rat, mouse, and frog) are GC rich (<35% AT), and the *D. melanogaster* sequence is AT rich (60.9% AT). The sequences of the other species deviate from 50% AT by <5%. Individual expansion segments generally show greater deviation from 50% AT than does the overall sequence. Expansion segments show the same base-
Molecular Coevolution of rRNA Expansion Segments

FIG. 1.—Intraspecific DIAGON plots and simplicity profiles (Tautz et al. 1986) of 10 23-28S rRNA genes, from Homo sapiens, Rattus norvegicus, Mus musculus, Xenopus laevis, Drosophila melanogaster, Caenorhabditis elegans, Physarum polycephalum, Saccharomyces carlsbergensis, Oryza sativa, and Escherichia coli. Simplicity profiles are aligned with the sequence in both dimensions. For all matrices, the stringency is 19 of 35 bp. The horizontal dotted line superimposed upon the simplicity profiles represents the highest number generated by any 10 adjacent positions during the 10 randomization runs. Thick and thin lines above the dot matrices represent the positions of expansion segments (thick) and core segments (thin) within the 28S rRNA sequence.

compositional bias as the core segments in all cases except in X. laevis 28S rRNA gene (which has an AT-rich core [64.6% AT] and GC-rich expansion segments [22.8% AT]) and in C. elegans (in which the difference between expansion and core segments is small). Additionally, differences in mean base composition between the expansion segments and the core segments are significant by Student's t-test at the $P < 0.01$ level in all cases except those of C. elegans and S. carlsbergensis.
Table 1
Lengths, Base Compositions, and RSFs of 23S-28S rRNA Genes

<table>
<thead>
<tr>
<th>Species</th>
<th>Length (bp)</th>
<th>rRNA (% AT)</th>
<th>Core (% AT)</th>
<th>Expansion Segments a (% AT)</th>
<th>RSF</th>
<th>RSF &gt; 1 b</th>
<th>Internal Homology c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>5,025</td>
<td>30.9</td>
<td>45.4</td>
<td>17.7</td>
<td>1.739</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>4,718</td>
<td>32.8</td>
<td>46.6</td>
<td>18.7</td>
<td>1.582</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>4,712</td>
<td>33.4</td>
<td>46.0</td>
<td>20.5</td>
<td>1.471</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Xenopus laevis</td>
<td>4,110</td>
<td>45.6</td>
<td>64.6</td>
<td>17.6</td>
<td>1.388</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>3,377</td>
<td>40.6</td>
<td>46.9</td>
<td>25.5</td>
<td>1.198</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>3,945</td>
<td>60.9</td>
<td>54.7</td>
<td>70.3</td>
<td>1.138</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>3,519</td>
<td>51.0</td>
<td>53.3</td>
<td>46.5</td>
<td>1.084</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Saccharomyces carlsbergensis</td>
<td>3,393</td>
<td>52.6</td>
<td>53.6</td>
<td>50.0</td>
<td>1.062</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Physarum polycephalum</td>
<td>3,788</td>
<td>46.6</td>
<td>49.7</td>
<td>40.6</td>
<td>1.037</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>2,904</td>
<td>46.6</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

SOURCES.—Sequences are taken from Gonzalez et al. (1985) (H. sapiens); Chan et al. (1983) (R. norvegicus); Hassouna et al. (1984) (M. musculus); Ware et al. (1983) (X. laevis); Takaïwa et al. (1985) (O. sativa); Tautz et al. (1988) (D. melanogaster); Ellis et al. (1986) (C. elegans); Veldman et al. (1981) (S. carlsbergensis); Otsuka et al. (1983) (P. polycephalum); Brosius et al. (1980) (E. coli).

a Individual values are approximately normally distributed about the mean.

b RSFs are taken to be significantly greater than 1.00 (+) if the simplicity factor of the natural sequence is greater than the mean simplicity factor of the 10 randomized runs by more than 3 SDS (see Tautz et al. 1986 for definitions and statistical significance of RSFs).

Internal homology is judged to be present (+) or absent (−) by inspection of DIAGON plots.

DNA Simplicity Analysis

Relative simplicity factors (RSFs) of the 26S/28S rRNA are presented in table 1, and graphic displays of simplicity profiles are given in figure 1.

RSFs correlate with the appearance or nonappearance of patterns of expansion-segment sequence similarity. All eukaryotic 26S/28S rRNA genes except P. polycephalum have RSFs that are significantly > 1.00 (P < 0.003) when compared with 10 randomized versions of the same sequence. Only sequences with RSF > 1.1 show a visible pattern of expansion-segment sequence similarity.

Alignment of displays of simplicity profiles with DIAGON plots (fig. 1) shows that those expansion segments with the highest cryptic simplicity are also those that show the strongest sequence similarity to one another.

Because both sequence similarities and elevated simplicity could simply reflect the strongly biased base compositions of the expansion segments, we carried out DNA sequence simplicity analysis of the individual expansion segments (table 3). The majority of the expansion segments in the species are significantly less simple (RSFs < 1.00) than random sequences of the same base composition. However, in man, rat, and mouse higher RSFs are associated with longer lengths of expansion segments. The largest expansion segments (D2, D6, D8, and D12) achieve RSFs that are significantly > 1.00 in certain vertebrate species. Because the simplicity algorithm is designed to detect high relative numbers of direct repeats, expansion segments with RSF values < 1.0 would reflect a preponderance of inverted repeats. However, short expansion segments of <200 bp are likely to show a depression of RSF because of the high "edge-
<table>
<thead>
<tr>
<th>Species</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7a</th>
<th>D7b</th>
<th>D8</th>
<th>D9</th>
<th>D10</th>
<th>D11</th>
<th>D12</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homo sapiens</em></td>
<td>23 (157)</td>
<td>17 (365)</td>
<td>24 (150)</td>
<td>23 (30)</td>
<td>15 (35)</td>
<td>16 (198)</td>
<td>28 (58)</td>
<td>56 (41)</td>
<td>13 (597)</td>
<td>8 (65)</td>
<td>16 (83)</td>
<td>60 (5)</td>
<td>21 (241)</td>
</tr>
<tr>
<td><em>Rattus norvegicus</em></td>
<td>23 (166)</td>
<td>18 (775)</td>
<td>15 (153)</td>
<td>36 (30)</td>
<td>21 (36)</td>
<td>13 (145)</td>
<td>32 (59)</td>
<td>27 (43)</td>
<td>16 (592)</td>
<td>11 (62)</td>
<td>16 (89)</td>
<td>60 (5)</td>
<td>28 (176)</td>
</tr>
<tr>
<td><em>Mus musculus</em></td>
<td>24 (168)</td>
<td>20 (589)</td>
<td>24 (150)</td>
<td>32 (20)</td>
<td>20 (37)</td>
<td>17 (154)</td>
<td>30 (59)</td>
<td>27 (39)</td>
<td>18 (608)</td>
<td>10 (59)</td>
<td>17 (93)</td>
<td>60 (5)</td>
<td>25 (238)</td>
</tr>
<tr>
<td><em>Xenopus laevis</em></td>
<td>20 (156)</td>
<td>15 (519)</td>
<td>19 (175)</td>
<td>27 (11)</td>
<td>23 (27)</td>
<td>20 (45)</td>
<td>25 (57)</td>
<td>12 (78)</td>
<td>14 (301)</td>
<td>18 (29)</td>
<td>17 (89)</td>
<td>60 (5)</td>
<td>25 (166)</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em></td>
<td>64 (148)</td>
<td>74 (345)</td>
<td>71 (160)</td>
<td>100 (19)</td>
<td>73 (40)</td>
<td>76 (48)</td>
<td>75 (87)</td>
<td>56 (51)</td>
<td>68 (237)</td>
<td>63 (28)</td>
<td>71 (213)</td>
<td>62 (8)</td>
<td>68 (174)</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>46 (151)</td>
<td>44 (235)</td>
<td>47 (151)</td>
<td>33 (6)</td>
<td>61 (37)</td>
<td>42 (35)</td>
<td>47 (61)</td>
<td>50 (41)</td>
<td>41 (163)</td>
<td>40 (10)</td>
<td>55 (83)</td>
<td>78 (9)</td>
<td>47 (157)</td>
</tr>
<tr>
<td><em>Physarum polycephalum</em></td>
<td>37 (156)</td>
<td>39 (247)</td>
<td>43 (119)</td>
<td>50 (10)</td>
<td>36 (46)</td>
<td>46 (57)</td>
<td>37 (91)</td>
<td>50 (27)</td>
<td>39 (99)</td>
<td>17 (11)</td>
<td>41 (234)</td>
<td>34 (29)</td>
<td>47 (166)</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>50 (151)</td>
<td>50 (207)</td>
<td>52 (110)</td>
<td>57 (7)</td>
<td>65 (40)</td>
<td>44 (27)</td>
<td>48 (50)</td>
<td>62 (37)</td>
<td>44 (142)</td>
<td>25 (22)</td>
<td>51 (75)</td>
<td>75 (4)</td>
<td>51 (142)</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>29 (130)</td>
<td>21 (214)</td>
<td>36 (107)</td>
<td>62 (8)</td>
<td>20 (39)</td>
<td>30 (27)</td>
<td>22 (45)</td>
<td>30 (37)</td>
<td>18 (139)</td>
<td>42 (24)</td>
<td>32 (78)</td>
<td>25 (4)</td>
<td>20 (130)</td>
</tr>
</tbody>
</table>
Table 3

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>EXPANSION SEGMENT</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D6</td>
<td>D7a</td>
<td>D7b</td>
<td>D8</td>
<td>D10</td>
<td>D12</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>0.30</td>
<td>1.32</td>
<td>0.66</td>
<td>1.08</td>
<td>...</td>
<td>...</td>
<td>1.17</td>
<td>0.09</td>
<td>0.97</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>0.36</td>
<td>1.22</td>
<td>0.74</td>
<td>0.60</td>
<td>...</td>
<td>...</td>
<td>1.11</td>
<td>0.14</td>
<td>1.21</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>0.36</td>
<td>1.08</td>
<td>0.72</td>
<td>0.65</td>
<td>...</td>
<td>...</td>
<td>1.06</td>
<td>0.14</td>
<td>1.19</td>
</tr>
<tr>
<td>Xenopus laevis</td>
<td>0.39</td>
<td>0.84</td>
<td>0.74</td>
<td>...</td>
<td>...</td>
<td>0.25</td>
<td>0.87</td>
<td>0.06</td>
<td>0.64</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>0.55</td>
<td>0.84</td>
<td>0.80</td>
<td>...</td>
<td>0.27</td>
<td>...</td>
<td>0.69</td>
<td>0.82</td>
<td>0.68</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>0.49</td>
<td>0.81</td>
<td>0.49</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.93</td>
<td>0.22</td>
<td>0.63</td>
</tr>
<tr>
<td>Physarum polycephalum</td>
<td>0.63</td>
<td>0.90</td>
<td>0.58</td>
<td>...</td>
<td>0.19</td>
<td>...</td>
<td>0.16</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td>Saccharomyces carlsbergensis</td>
<td>0.49</td>
<td>0.66</td>
<td>0.20</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.84</td>
<td>0.20</td>
<td>0.83</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>0.31</td>
<td>0.68</td>
<td>0.36</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.73</td>
<td>0.04</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Effects" (see Tautz et al. 1986 for details). Because of these operational features of the program, not much significance can be attached to low RSF values of the shorter expansion segments.

Controls

To assess the contribution of base-compositional bias to both the patterns of inter-expansion-segment sequence similarity and the RSF of 26S/28S rRNA sequences, two artificial sequences were constructed that mimicked closely the characteristics of the M. musculus 28S rRNA gene (see Material and Methods). Simplicity analysis of these sequences gave values of 1.028 and 1.101 for the artificial M. musculus sequences, compared with a value of 1.471 for the natural rRNA gene. Neither of these values was significantly > 1.0. DIAGON plots for the two types of artificial sequence (fig. 2a, 2b) show no pattern of "expansion segment" self-similarity in the completely random sequence (fig. 2a) and a pattern that is much weaker in the sequence containing artificial expansion segments than that in the natural rRNA gene.

Interspecific Comparisons

In figure 3 we present the dot matrices of selected comparisons between pairs of species, together with the corresponding simplicity profiles. The first three compare the human 28S rRNA gene with those of frog, rice, and fruit fly, respectively. Relatively high similarities are displayed across all expansion segments between human and frog, in addition to the expected core-segment similarities. Surprisingly, the expansion segments of human and rice are also similar, although less so than the inter-expansion-segment similarities detected within each species. No similarities are observed between the expansion segments of human versus fruit fly, frog versus fruit fly, and rice versus fruit fly, although the core segments are similar between species, with displacements from the true diagonal reflecting differences in lengths between the various 28S rRNA genes (fig. 3). The rice-frog comparison reveals significant but lower levels of similarity between expansion segments than are observed in each species alone (fig. 3).
The significance of these comparisons with reference to the coevolution of expansion segments is discussed below.

Discussion

Molecular Coevolution of Expansion Segments

Sequence analysis of the expansion segments of eukaryotic 26S/28S rRNA genes (table 1) shows consistent bias of base composition within any one species. For example, the expansion segments of vertebrate 28S rRNA genes are more GC rich than the core segments, while those of Drosophila melanogaster 28S rRNA are more AT rich. Furthermore, dot matrices reveal high levels of similarity among most expansion segments within a species.

It is possible to explain these data on the simple basis that (1) the rDNAs of different species show different biases in the fixation of point mutations (e.g., a 70% bias toward A+T in D. melanogaster; Hancock et al. 1988) and (2) divergence is higher in expansion segments than in the core segments (Tautz et al. 1988). This would result in any bias in point mutation being reflected more strongly in the expansion segments than in the core. This explanation does not take into account, however, the correlation between DNA simplicity in the expansion segments and the expansion-segment sequence similarity. Is there a causal relationship between slippage-generated simplicity and the coevolution of expansion segments?

Comparison of the Mus musculus 28S rRNA gene with two types of artificial sequences made up of blocks of random sequence of the same lengths and base compositions (fig. 2) shows that base composition alone cannot account for either the intensity of the patterns of internal sequence similarity detected by DIAGON plots (fig. 1) or the simplicity of the sequence as measured by its RSF and displayed in the simplicity profiles. Thus the patterns of internal sequence similarity detected by DIAGON analysis of vertebrate 26S/28S—and, to a lesser extent, of the D. melanogaster 28S
FIG. 3.—Interspecific DIAGON plots and simplicity profiles of the 28S rRNA genes between selected pairs of species. Stringency levels are as in fig. 1.
Molecular Coevolution of rRNA Expansion Segments

and Oryza sativa 25S—rRNA genes reflects true sequence relatedness between expansion segments and the molecular coevolution of these sequence regions.

Molecular coevolution is observed when the evolution of two or more functionally interacting components is constrained only by the necessity to maintain their interaction, for example, in the case of the multiple rDNA promoters and RNA polymerase I transcription complex of eukaryotes (Dover and Flavell 1984). In the case of the 26S/28S rRNA expansion segments, there is as yet no functional basis for the observed pattern of molecular coevolution. Indeed, it has been suggested that expansion segments are functionless structures that are tolerated only to the extent that they do not disrupt ribosome function (Gerbi 1985). Expansion segments might be constrained to have similar sequences, however, if, for example, they interacted with one another in the ribosome to form a structural core or to share a common interaction with another cellular component. In an accompanying paper we describe both the interspecific conservation of the rRNA secondary structures that correspond to the expansion segments and the nature of the compensatory mutations that have occurred within them (Hancock et al. 1988).

Possible Mechanisms for Maintaining Expansion-Segment Coevolution

At the sequence level, two mechanisms can be envisaged that might result in the coevolution of expansion segments despite their rapid accumulation of point mutations. First, sequence similarity of short regions might be amplified by slippage-like mechanisms to produce larger blocks of related sequence. There is evidence that slippage-like mechanisms have been active in D. melanogaster and other 26S/28S rRNA genes but not in 18S rRNA genes, which do not show inter-expansion-segment sequence similarity (Tautz et al. 1988; see below). Alternatively, sequence similarity might be maintained by intragenic gene conversion, that is, microconversion between intragenic segments. An explanation for the molecular coevolution of expansion segments based on selection only is faced with the usual problems associated with selection operating in buffered genetic systems such as multigene families (for full discussion of this, see the accompanying paper [Hancock et al. 1988]). In the absence of these mechanisms, any intragenic patterns of sequence similarity would be expected to be obliterated in time by the accumulation of point mutations, although unequal crossing-over would still maintain homogeneity between the 28S rRNA genes of the rDNA family. This may explain (1) why certain 26S/28S rRNA genes show no pronounced patterns of sequence similarity among their expansion segments and, in particular, (2) why Physozum polycephalum 26S rRNA shows no such pattern. This species has an RSF that is not significantly > 1.00 despite being little shorter than D. melanogaster 28S rRNA (3,788 nucleotides compared with 3,945) and having expansion segments that are significantly more G-C rich than its core.

Role of Slippage-like Mechanisms in the Evolution of Expansion Segments

We have shown in an accompanying paper that only the 28S gene of the D. melanogaster rRNA genes shows evidence of the products of slippage-like mechanisms, as measured by its high RSF factor (Tautz et al. 1988). The other rRNA genes in this species reveal little evidence of simplicity. High RSFs are also measurable for the 28S rRNAs of tetrapods (frog, mouse, rat, and human) but not for those of lower eukaryotes (yeast, slime mold, and nematode). Table 1 shows a correlation between the RSFs and the lengths of eukaryotic 26S/28S rRNAs. If the simplicity profiles of these rRNAs
are superimposed upon their DIAGON patterns (fig. 1), it is clear that peaks of simplicity within the molecules correlate strongly both with regions of internal sequence similarity and with expansion segments (see also Tautz et al. 1986). In addition, the appearance of patterns of self-similarity between expansion segments correlates with the overall RSF of the 28S rRNA gene, supporting the proposal that slippage-like mechanisms have played a role in the molecular coevolution of expansion segments (see above).

Analysis of RSFs of individual expansion segments, which eliminates effects of base-compositional bias, shows a correlation between length and RSF similar to that shown by the whole of the 28S rRNA gene (tables 2, 3). In many cases RSFs of individual expansion segments are significantly <1.00, possibly as a result of (1) greater end effects of short lengths of sequences on the algorithm used to measure simplicity factors and (2) the presence of inverted repeats (see Tautz et al. 1986 for details). Expansion segments D2, D6, D8, and D12 have RSFs >1.00 in one or more of the vertebrates (table 3). These expansion segments are those that both are the largest and have undergone the greatest change in length during evolution. For example, expansion segments D1 and D3, which are larger than D6 in many species, vary less in length than does D6 and do not show elevated RSFs (see table 2). This suggests that expansion segments D2, D6, D8, and D12 have undergone expansion by the acquisition of simple sequences—and, therefore, as the result of the action of slippage-like mechanisms. This suggestion is supported by the observation of considerable sequence variation between expansion segments in different copies of human 28S rRNA genes, copies that frequently reflect differences in copy number of short repetitive motifs (Gonzalez et al. 1985; Maden et al. 1987). Hassouna et al. (1984; see also Michot and Bachellerie 1987) have suggested that expansion segments enlarge primarily by the addition of short tracts of sequence at the tips of secondary-structural stems and that the added sequences are simple. This is consistent with our observations, described in an accompanying paper (Hancock et al. 1988), of a conserved pattern of expansion-segment secondary structure, even between species whose expansion segments show complete sequence divergence.

In conclusion, the simple sequence composition of expansion segments of eu-karyotic 26S/28S rRNAs reflects the activity of slippage-like mechanisms, and some expansion segments have expanded by the acquisition of simple sequences, presumably as the result of slippage-like mechanisms. Slippage-like mechanisms generating copy-number variation of short DNA motifs possibly underlie much of the hypervariability of minisatellite DNA families in diverse species (Jeffreys et al. 1985) and CpG-rich islands (Bird 1986) (for reviews, see Tautz et al. 1986; Levinson and Gutman 1987).

Expansion segments also show molecular coevolution, which is detectable at the sequence level as well as at the level of base composition. The molecular coevolution of expansion segments may be the result purely of the action of DNA turnover mechanisms on the 28S rRNA genes irrespective of the function of the rRNA; for example, slippage-like mechanisms acting on similar motifs may generate tracts of related sequence, or micro gene-conversion domains may be shared between expansion segments. It remains to be established, however, whether this phenomenon reflects a functional role for the expansion segments. Although the available evidence is consistent with nonfunctional expansion segments (Gerbi 1985), a structural role that involves interaction between expansion segments with related sequences cannot be ruled out (see accompanying paper [Hancock et al. 1988]). For a discussion of the origins of expansion segments, see the work of Clark (1987).
Interspecific Comparisons of Expansion Segments

The interspecific comparisons of expansion segments and core segments support the above conclusion, with some interesting qualifications. The total absence of interspecific similarity between sets of expansion segments when the comparison is between fruit fly and human, fruit fly and frog, and fruit fly and rice indicates that the set of expansion segments of any given species has diverged and evolved as a unit. This is in contrast to species core segments, which on occasion are also AT or GC rich but reveal no similarity among themselves. High levels of similarities within species but not between species are reminiscent of concerted evolution (Dover 1982), with the difference here being that intraspecific similarity is now observed at two levels—among the set of expansion segments of a given 28S rRNA gene and between the multiple 28S rRNA genes of the rDNA family. The latter level of similarity can be explained as a consequence of the activities of unequal crossing-over, which operates in all examined rDNA families (for reviews, see Fedoroff 1979; Long and Dawid 1980; Coen et al. 1982; Dover 1982; Arnheim 1983; Flavell 1986; Flavell et al. 1986). As discussed above, the former level of similarity might be a consequence of slippage-generated variation, biased mutation, and, possibly, gene conversion.

The existence of interspecific similarities in expansion segments in dot-matrix comparisons of human and frog and of human and rice introduce two further considerations. The human-frog similarities might reflect the slow divergence in sequence of the ancestral rRNA 28S gene in the vertebrate lineage. However, the similarity of general sequence composition of expansion segments, as revealed in vertebrate 28S rRNAs by dot-matrix analysis at a given stringency, might obscure details of the sequence divergence between the sets of expansion segments of the different species. Given that divergence is due to gains and losses of short motifs by slippage, alignment is hazardous. However, attempts to align mammalian and frog 28S rRNAs (see Ware et al. 1983; Gorski et al. 1987) show considerably more divergence than is apparent at the cruder level of dot-matrix sequence comparison. The maintenance of apparent sequence relatedness between expansion segments over long periods of evolutionary time could reflect the propensity for a mechanism of turnover such as slippage to conserve preestablished motifs, because slippage would tend to occur on the basis of units of sequence that are already repetitive. In this sense slippage is a feedback mechanism, capable of retarding the divergence of existing repetitive motifs (Dover and Tautz 1986). In the case of vertebrate expansion segments, the assumption would be that their progenitors consisted of various permutations of short GC-rich motifs (see table 2). The longer expansion segments (e.g., D2 and D8) of the human 28S rRNA would reflect the growth of GC-rich motifs at the ends of existing motifs.

It is unlikely that the same argument can be made for the expansion-segment similarities between human and rice. This is probably a consequence of the accidental generation of similar GC-rich repetitive motifs by slippage in the separate lineages. This example of convergent evolution illustrates the dangers of constructing phylogenetic trees without some knowledge of the types, rates, and biases of turnover mechanisms and mutation in different parts of the rDNA unit. For a fuller discussion of DNA turnover and the molecular clock, see the work of Dover (1987).

Acknowledgments

We thank the Science and Engineering Research Council for grants GR/D18677 and GR/E 53538 toward the cost of this research.
LITERATURE CITED


Richard B. Flavell, reviewing editor

Received September 6, 1987; revision received January 20, 1988