An Episodic Change of rDNA Nucleotide Substitution Rate has Occurred During the Emergence of the Insect Order Diptera

Markus Friedrich\(^1\) and Diethard Tautz
Zoologisches Institut der Universität München, Germany

We have studied the potential reasons for a conspicuous deviation of substitution rates in Dipteron ribosomal genes. Systematic pairwise relative-rate tests reveal that a significant increase in substitution rate is characteristic for Diptera, but not for other insects analyzed. Estimation of sequence change in specific lineages reveals that most of these substitutions took place during the evolution of the Dipteran stem lineage. When related to the paleontologically documented periods of absolute time, the substitution rate in the stem lineage of the Diptera underwent an at least 20-fold increase compared to other insect groups and subsequently dropped by a factor of 10 before the diversification of the major Dipteran subgroups. Systematic comparisons of nucleotide composition show that this episodic change in substitution rate was accompanied by a significant increase in A+T content of Dipteran rDNA. Our data suggest that the episodic evolution of the Dipteran rDNA has most probably been caused by a change of directional mutation pressure which must have occurred during the evolution of the stem lineage of the Diptera.

Introduction

Ribosomal genes code for functional RNA molecules (rRNA) that participate in the protein biosynthesis machinery of all organisms (Gutell, Larsen, and Woese 1994). They must have been optimized early in evolution and would therefore be a priori not be expected to show much deviation from a clock-like substitution pattern due to changing adaptive constraints. This expectation is largely fulfilled for most comparisons between distantly related species (Gutell, Gray, and Schnare 1993), which makes rDNA genes the molecules of choice for the reconstruction of ancient diversifications in molecular phylogenetic analysis (Hillis and Dixon 1991). In a few species, however, like the fruitfly Drosophila melanogaster or the nematode worm Caenorhabditis elegans, rDNA substitution rates are strongly accelerated (Carman, Kimsey, and Berbee 1992; Philippe, Chenuil, and Adoutte 1994; Fitch, Bugaj, and Emmons 1995). Although this reduces the reliability of molecular phylogenetic reconstruction methods (Felsenstein 1978), no attempt has so far been made to understand the mechanisms underlying the evolutionary dynamics of the nuclear rDNA genes in these lineages.

In the nematodes, a more detailed analysis of the occurrence and absolute amount of change in substitution rate is hampered by the lack of a reliable paleontological record and controversial phylogenetic relationships (Fitch, Bugaj, and Emmons 1995). Insects, on the other hand, are highly suitable for such a study, since their evolution is well documented with respect to paleontology and phylogeny of the major subgroups (fig. 1) (reviewed in Kristensen 1991; Kukalova-Peck 1991). The oldest insect orders, such as the wingless Collembolans (spring tails), Archaeognathans (bristle tails), or Zygoptera (silver fishes), emerged during the Devonian about 380 MYA. Subsequently, a first major radiation which produced the various orders of primitive or hemimetabolous winged insects took place during the Carboniferous approximately 300 MYA. A second radiation in the early Permian (280 MYA) led to the almost synchronous occurrence of representatives of most holometabolous insect orders. The Holometabola represent a clearly monophyletic taxon, but the phylogenetic relationships among them are poorly resolved (Kristensen 1991; Carman, Kimsey, and Berbee 1992). The monophyly of the Diptera, however, is well supported by a number of clear characters, such as the modification of the hind wings to halteres. The stem lineage of the Diptera is believed to have already existed 250 MYA (Hennig 1981) and a fossil specimen (Permutipula patricia) which is considered to be a candidate representative of the Dipteron stem lineage is 240 Myr old (Willmann 1989).

The stem lineage of the Diptera has split into a crown group which comprises at least five major subgroups: (I) the Brachycera, (II) the Tipulomorpha (crane-fly-related families), (III) the Bibionomorpha (fungus-gnat-related families), (IV) the Psychodomorpha (moth-midge-related families) and (V) the Culicomorpha (mosquito-related families) (see Friedrich [1995] and Oosterbroek and Courtney [1995] for review). The diversification of the Dipteron stem lineage well before the end of the Triassic about 210 MYA (Hennig 1981) is well documented in the paleontological record (Fraser et al. 1996). Accordingly, Wootton and Ennos (1989) have suggested that the time period for the evolution of the Dipteron stem lineage might be as short as 25 Myr, but to be on the conservative side, we shall assume a period of 40 Myr in the following.

Here we make use of the detailed information on insect evolution and the analytical tools available from molecular phylogenetics to pinpoint the evolutionary origin of the acceleration in Dipteran rDNA substitution rate and to analyze correlated changes in molecular substitution patterns.
FIG. 1.—Time scale of insect phylogeny. The paleontological record of major insect subclades according to Kukalová-Peck (1991) is superimposed onto the canonical view of insect phylogeny (Kristensen 1991). Numbers below branches refer to conspicuous derived character states: (1) double articulation of mandible; (2) possession of wings; (3) complete metamorphosis; (4) modification of hind wings to halteres, acquisition of a labellum.

Materials and Methods

Samples

28S rDNA D1, D3–5, and D7 expansion segments were sequenced from various hexapod orders. In addition, 18S and 28S rDNA sequences were retrieved from the data bank (table 1). Taxon choice was governed by the aim to cover a range of species representing the oldest insect lineages, such as Zygentoma (silverfishes) and Orthoptera (grasshoppers), as well as the most important orders of holometabolous insects, such as Hymenoptera (bees), Siphonaptera (fleas), Mecoptera (scorpion flies), and Lepidoptera (butterflies). In addition, at least one representative of each of the five major Dipteran subgroups was sequenced. Millipede and crustacean sequences were chosen as outgroups, as both represent potential sister groups to hexapods (Hennig 1981; Friedrich and Tautz 1995).

Sequence Determination

Extraction of genomic DNA from fresh or alcohol-preserved samples was carried out using either SDS or DTAB as detergent and purification was done via phenol/chloroform extraction or spin dialysis. Ten to 100 ng of genomic DNA was added to a PCR reaction mix (Saiki et al. 1988). The first 20 cycles of the PCR reaction were done with a 1-min annealing step at the primer-specific temperatures specified below, a 2-min elongation step at 72°C, and a 1-min denaturation step at 93°C. This was followed by a second round of 15–20 cycles with a prolonged elongation step of 4 min. The 28S rDNA fragments amplified correspond to the regions D1 (3338–3650), D3–5 (4066–4729), and D7 (5000–5464) of the *Drosophila melanogaster* 28S rRNA sequence (Tautz et al. 1988). Primer combinations used were (annealing temperatures are shown in parentheses): 687 (5' CGG TGG ATC ACT (C/T)GG 3') and 427 (5' CCC(C/G)CGTAA(T/C)TTAAGCATAT 3') (60°C) or...
646 Friedrich and Tautz

...to amplify D3-5-spanning fragments. Depending on their expected size, the PCR products were separated on 0.8% (>600 bp) or 1.2% agarose gels (<600 bp), then cut out from the gel and eluted using the Qiaex kit (Qiagen). Approximately 100 ng of DNA was subjected to a direct sequencing reaction optimized according to Carlson et al. (1990). The sequence of each fragment was determined in both directions using one of the amplification primers or one of the internal primers 486 (5’ TCGGAAGGAACCAGCTACTA 3’), 706 (5’CGCCAGGGTTTCGTGTGAACAG 3’), 703 (5’ TTCCAAACC(A/C)TATCTC 3’), 743 (5’ CGATTGTTCAGGTTCC 3’) for the D3 fragments.

**Multiple Alignment**

The 28S rDNA sequences of 14 taxa altogether were first aligned with the CLUSTAL V program (Higgins 1994) and subsequently manually adjusted on the basis of previously published secondary-structure models (detailed in Friedrich 1995). These models, together with the criterion of complementarity of putative helical regions in the new sequences, served also as matrices to subdivide the sequences into stem and loop regions. Eight hundred unequivocally alignable positions that contained no gaps in any of the insect taxa were used for the phylogenetic analysis. This subset of sequence sites to which we will refer as slowly evolving sites consists of 450 stem and 350 loop positions.

The 18S rDNA sequences were aligned by the same criteria and yielded 833 unambiguously alignable positions, with 523 located in stem regions and 310 located in loop positions. For the analysis of directional mutation pressure, 116 to 238 positions per species were extracted as fast-evolving sites from the variable regions in the 28S rDNA sequences which were not unambiguously alignable in the present data set due to multiple insertion and deletion events. The multiple alignments of the 18S and 28S rDNA sequences with indications of the helical sites, the slowly evolving sites, the fast-evolving sites are deposited on the EBI ftp server under the accession numbers DS25411 (18S), DS25025 (28S D1), DS25409 (28S D3–5) and DS25410 (28S D7).

**Molecular Phylogenetic Analysis**

Three different phylogenetic methods were employed: maximum likelihood (ML) (Felsenstein 1981), maximum parsimony (MP) (Fitch 1971), and neighbor joining (NJ) (Saitou and Nei 1987). MP reconstructions and subsequent bootstrap tests (Felsenstein 1985) were carried out with the heuristic TBR tree reconstruction method implemented in the PAUP package (Swofford 1993). For the estimation of ML trees according to the Hasegawa, Kishino, and Yano (1985) model of sequence evolution we used the DNAML program of the PHYLIP package (Felsenstein 1989). The same package was used to reconstruct NJ trees with distances corrected for multiple hits either according to Kimura (1980) (NJK trees) or according to Jin and Nei (1990) (NJG trees) and to analyze their support with the bootstrap method (Felsenstein 1985). The evolutionary distance estimation method by Jin and Nei (1990) corrects for substitution heterogeneity across sites. This requires estimation of the alpha parameter, which is a measure thereof. The alpha parameter was deduced according to Johnson and Kotz (1973). The number of substitutions per site was inferred from the consensus tree in figure 2a using the CHART CHARACTER STEPS option of the MacClade package (Maddison and Maddison 1992). The resulting data were imported into Statview (Feldman et al. 1988) to calculate the variance (σ²) of mean substitutions inferred per site (Mₑ). The alpha parameter was then calculated using the formula Mₑ/σ² = Mₑ.

The average ratio of transitions (Ts) to transversions (Tv) in the actual data set was also inferred from the consensus tree in figure 2a using the CHART STATE CHANGES AND STASIS option in the MacClade package (Maddison and Maddison 1992) counting unambiguous changes only.

**Comparative Sequence Analysis**

The calculations required to perform a test on the significance of relative-rate differences according to Wu and Li (1985) were implemented on a computer spreadsheet program (EXCEL, Microsoft). The pairwise Ts and Tv sequence differences were extracted from the phylogenetic analysis output of the CLUSTAL V package (Higgins 1994) and typed in manually per pairwise comparison. To test for taxon-specific differences of DNA composition, the program STATIO, developed and distributed by Rzhetsky and Nei (1995), was used. To analyze the substitution patterns of unequivocal character state changes in the different lineages, the CHART STATE CHANGES AND STASIS option of the MacClade package (Maddison and Maddison 1992) and the DESCRIBE TREE option of PAUP (Swofford 1993) were used.

**Results**

**Consistent Resolution of the Hexapod Phylogeny by 28S rDNA Sequences**

The slowly evolving sites of the 28S sequences were used to reconstruct a hexapod phylogeny. Two sets of taxa were used, one without and one with the Diptera. The Diptera were omitted in the first set to avoid possible artifacts associated with the accelerated rate known from Drosophila melanogaster (Carmean, Kimsey, and Berbee 1992; Philippe, Chenuil, and Adoutte 1994; Friedrich and Tautz 1995). Thus, initially a total of 11 taxa were included, with 5 of them being representatives of different holometabolous insect orders. The tree to-
polymorphs obtained with unweighted MP, NJ using Kimura corrected distances (NJK), or ML were congruent except for the branching order among holometabolous insect orders, i.e., the relative placement of the Hymenoptera and the Lepidoptera (fig. 2a). In order to employ more adequate sequence evolution assumptions in the tree reconstruction, we inferred the ratio of transitions to transversions as well as the alpha value (measure for rate heterogeneity across sites) from this initial consensus topology. A 1s/1v ratio of 0.8 was found in the loop regions, and a ratio of 2.02 was found in the stem regions. The alpha value for all sites was found to be 0.95. The MP bootstrap consensus tree was obtained by weighting stem versus loop sites 0.8 to compensate for compensatory substitutions as proposed by Dixon and Hillis (1993), as well as weighting Tv two times over Ts in stem regions. For the calculation of a bootstrap consensus tree using NJ with distances corrected for gamma-distributed rate heterogeneity across sites (NJG), we assumed Ts/Tv to be 1.5 (averaged over all sites) and alpha = 1. The resulting bootstrap values are mostly consistent between the two methods (fig. 2). They indicate a high support for most branching events but do not provide resolution for most holometabolan orders. Only a sister group relationship between Mecoptera and Siphonaptera is highly supported by NJG. Similar results have been reported in a previous study of holometabolan phylogeny based on 18S rDNA sequences (Carmean, Kimsey, and Berbee 1992).

The second set of taxa included six Dipterans but excluded some of the other hexapod taxa to cut down computation time for the bootstrap analysis. We found that NJK fails to place the Dipteran taxa within the Holometabola (not shown), while weighted MP, ML, and NJG consistently place the Diptera as sister group to the Lepidoptera within the Holometabola (fig. 2b) and the monophyly of the Diptera is strongly supported in all cases. On the other hand, no robust resolution is found among the major Dipteran subgroups, but a sister group relationship between the Brachycera and the Bibionomorpha is consistently supported.

Two major conclusions can be drawn from these results. First, the main bifurcations of hexapod phylogeny are strongly supported by both morphology and 28S rDNA sequences. Second, the tempo and mode of 28S rDNA sequence evolution provide the necessary phylogenetic information which reflects early and late Paleozoic diversification events.

Comparison of Relative Substitution Rates

The ML phylogram in figure 3a shows very long branches for the Dipterans, indicating that each representative of the major Dipteran subgroups must have accumulated many more substitutions than the representatives of the other hexapod orders under consideration. Within the Diptera, the two representatives of the sub-group Culicomorpha (Culex and Chironomus) show terminal branch lengths that are even longer than those of the other Dipterans, indicating a second taxon-specific difference of evolutionary rate. It is also notable that the internal branches which connect the major Dipteran subgroups are relatively short. This suggests that the major Dipteran sublineages must have diverged from each other within a short period of time, which is consistent with the synchronous occurrence of these subgroups in the paleontological record (Evenhuis 1994; Fraser et al. 1996).

The 18S data for a smaller, but comparable, set of taxa show a very similar phylogram topology, both with respect to the difference between Diptera and non-Diptera representatives and with respect to the difference between the culicomorphous representative (Aedes) and the nonculicomorphous representative (Drosophila) (fig. 3b). This suggests that both ribosomal genes evolve in a similar way with respect to fluctuations of substitution rates.

In order to examine whether the observed differences in branch lengths are statistically significant, we have carried out relative-rate tests (Sarich and Wilson 1973; Wu and Li 1985). These rate tests need an uncontroversial outgroup as reference taxon, a condition
The pattern of taxon-specific rate differences is visualized in phylogenograms where the length of a branch is proportional to the corresponding number of substitutions as estimated by the ML method assuming the parameters used for the reconstruction in figure 2. The bar corresponds to 0.1 substitutions per position. Node 1 represents the first diversification of holometabolous taxa. Node 2 represents the split of the Dipteran stem lineage, and node 3 represents its first diversification into the major Dipteran sublineages. a, Phylogram derived from partial 28s rRNA sequences presented in this paper (see table 1). b, Phylogram derived from previously published 18s rRNA sequences (see table 1).

which is fulfilled in the present data set at various levels of the systematic hierarchy. The results of the relative-rate tests can be summarized by plotting the test values against the evolutionary distances estimated for the pairs of taxa under consideration (fig. 4). Two major areas of dots are evident in the resulting scattergram. First, there is the area below the level of significance (stippled line) which includes the results for the comparisons among only Dipteran taxa or among only non-Dipteran taxa. This shows that, under this criterion, there are almost no significant rate heterogeneities within each of these groups, indicating that the 28s rDNA evolves approximately clock-like within the respective taxonomic groups (albeit with different rates—see below). In contrast, the area above the level of significance contains all test values that are derived from comparisons between a Dipteran species and a non-Dipteran holometabolous insect. This shows that each representative of the Dipteran subgroups evolved significantly faster than other hexapods.

Thus, a significantly enhanced rate of molecular substitution in the 28s rDNA gene is characteristic for the Diptera in general. The same result can be obtained when 18s sequence divergence is analyzed in this way (data not shown, but compare Carmean, Kimsey, and Berbee 1992).

Comparison of Absolute Substitution Rates

The phylogenograms in figure 3 indicate that by far the most sequence change is concentrated in the branch which represents the evolution of the Dipteran stem lineage (between points 2 and 3 in fig. 3). On the other hand, the paleontological evidence suggests that this branch—stem lineage should have existed for a comparatively short time only (see Introduction). To assess the magnitude of relative change that must have occurred in this specific lineage, we estimated the absolute rates of substitution. This requires one to relate the amount of molecular change to periods of absolute time, which can be done by taking three reference points from the insect fossil record: the assumed primary divergence of holometabolous insect orders (approximately 280 MYA) (point 1 in fig. 3), the assumed earliest appearance of the Dipteran stem lineage (approximately 250 MYA) (point 2 in fig. 3), and the assumed primary divergence of the major Dipteran subgroups, which is the emergence of the Dipteran crown group (approximately 210 MYA) (point 3 in fig. 3). The respective values were inferred from the branch length estimates in the ML tree. Since the phylogenograms indicated that, among the Diptera, the major subgroup Culicomorpha (Culex and Chironomus) experienced a second acceleration of substitution rate which might not be typical for the average substitution rate in the Dipteran crown group, we calculated the estimates for the Dipteran crown group both with and without the representatives of the Culicomorpha. Division of the group-specific ML branch length averages by the respective periods of absolute time results in the absolute substitution rates given as substitutions per site per 100 Myr in table 2.

The values found here for the 18s gene can be compared with those found in other studies. Spears,

![Figure 3](image-url)
Abele, and Kim (1993) estimated a value of 0.02 for the
Absolute Substitution Rates of Insect rDNA (average
substitutions per site per 100 Myr)

More importantly, we find that the values for the
Dipteran stem lineage are more than 20-fold higher than
those for the other holometabolan taxa, while the Dip-
teran crown group shows only 2–4-fold higher rates (ta-
ble 2). Thus, these comparisons provide evidence for a
strong but episodic acceleration of substitution rate at
the time of the emergence of the Dipteran stem lineage.

Correlated Changes in DNA Composition

*Drosophila melanogaster* rDNA is exceptional not
only with respect to the substitution rate, but also with
respect to DNA composition. The relative A+T content
in this species was found to be considerably higher than
in other distantly related metazoans (Tautz et al. 1988).
We have therefore analyzed compositional differences
in the 18S and 28S sequences of the more closely related
insect taxa. Because there have been reports on com-
positional differences between stem and loop regions in
the rDNA (Vawter and Brown 1993), these two classes
of sites were analyzed separately.

The most obvious difference between Dipteran and
other hexapod groups is found in the 28S rDNA stem
regions (fig. 5). While non-Dipteran insects show on av-
erage an A+T content of around 37%, this value is shift-
ed to about 48% within the Diptera. A similar, although
less pronounced, difference can be be seen in the 18S
stem regions, where 48% A+T in the hexapods com-
pares with 51% in the Diptera. The loop region A+T con-
tent is also slightly elevated in the Diptera for both the
18S and the 28S sequences. It should be noted, how-
ever, that loop regions are generally more A+T-rich,
mainly due to an elevated content of A nucleotidics.

To test whether these differences are significant, we
employed the test on departure from DNA composition
stationarity developed by Rzhetsky and Nei (1995) (ta-
ble 3). We find the null hypothesis—DNA composition
is stationary—to be rejected for the 28S rRNA gene if
all taxa are included. If Dipteran or other hexapods are
tested separately, only the Dipteran 28S rDNA stem
regions show a significant departure from compositional
homogeneity. However, this is entirely due to the taxon
Culex, for which we suspect that there was another re-
cent change in substitution equilibria (see below). Over-
all, this suggests that within the Diptera and within most
other hexapod groups the evolution of 28S rRNA com-
position is generally stationary. Departure is found only
when they are tested against each other, which is due to
the differences in A+T content observed between them.
The respective tests on the 18S rDNA data set reveal
no significant results if stem or loop region sites are
tested separately. However, if all sites are included, sta-
tionarity is rejected if Dipteran and non-Dipteran taxa
are tested together but not if each group is tested sepa-
ately. In summary, we conclude that the Dipteran rDNA
in general shows a significant bias toward A+T residues
when compared to that of non-Dipteran insects.

Changes in Mutation Pressure

Sueoka (1988) has suggested that changing DNA
composition in neutrally evolving sequences should re-
fect changes in directional mutation pressure. The sites
analyzed above are among the slow-evolving ones and
are therefore certainly not neutral. However, they may
only be under negative selection and thus evolve ac-
ording to a neutral model. On the other hand, the rDNA
regions sequenced also harbor fast-evolving regions,

Table 3

<table>
<thead>
<tr>
<th>Gene</th>
<th>Taxa</th>
<th>Sites</th>
<th>P Value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28S (Diptera)</td>
<td>Total</td>
<td>6.18*</td>
<td>0.00*</td>
</tr>
<tr>
<td></td>
<td>Loop</td>
<td>3.02*</td>
<td>0.00*</td>
</tr>
<tr>
<td></td>
<td>Stem</td>
<td>5.67 *</td>
<td>0.00*</td>
</tr>
<tr>
<td>Hexapods excluding Diptera</td>
<td>Total</td>
<td>71.83</td>
<td>0.74*</td>
</tr>
<tr>
<td></td>
<td>Loop</td>
<td>7.65</td>
<td>0.74*</td>
</tr>
<tr>
<td></td>
<td>Stem</td>
<td>7.92</td>
<td>0.74*</td>
</tr>
<tr>
<td>Diptera</td>
<td>Total</td>
<td>10.77</td>
<td>0.74*</td>
</tr>
<tr>
<td></td>
<td>Loop</td>
<td>2.61</td>
<td>0.74*</td>
</tr>
<tr>
<td></td>
<td>Stem</td>
<td>2.37</td>
<td>0.74*</td>
</tr>
<tr>
<td>Diptera excluding Culex</td>
<td>Total</td>
<td>5.17</td>
<td>0.74*</td>
</tr>
<tr>
<td></td>
<td>Loop</td>
<td>28.02</td>
<td>0.74*</td>
</tr>
<tr>
<td></td>
<td>Stem</td>
<td>7.17</td>
<td>0.74*</td>
</tr>
</tbody>
</table>

Note.—Asterisks indicate significant rejection of the null hypothesis (sta-
tionarity of DNA composition).
which are yet more likely to follow a neutral evolution model. These regions were called expansion segments (Clark 1984) and might possibly be a good indicator of the directional mutation pressure acting on a genome. We have therefore compared the DNA composition in the 28S rRNA expansion segments to which we refer as the fast-evolving sites in the multiple alignment.

We find that most Dipteran lineages are considerably enriched in A+T content in this subset of sequence sites. On average, Dipters show a value around 71% (±10%) which compares with 49% (±10%) in non-Dipteran insects (fig. 5). It was previously shown that the A+T content in the expansion segments also follows the general compositional bias in the spacer regions of the rDNA clusters (Hancock and Dover 1988). This is also reflected in our data set. The A+T content of the internal transcribed spacer regions (ITS) of Drosophila (70%) (Schlotterer et al. 1994), Bradyisia (69%) (Jordan, Latil, Damotte, and Jourdan 1980), and Simulium (75%) (Tang et al. 1996) match well with those observed in the 28S rRNA gene expansion segments of the equivalent taxa in our data (Drosophila 79%, Bradyisia 64%, Chironomus 67%). Culex has the lowest A+T content of all Dipteran taxa sampled, but its value of 45% still matches the values found in the ITS regions in this species (40%–50%) (Miller, Crabtree, and Savage 1996). Similar correspondence can be found for non-Dipteran insects. The A+T content of 42% found for Apis compares with 40% in the ITS 2 of the distantly related hymenopteran species Melittobia (Campbell, Steffen-Campbell, and Werren 1993), and the A+T-content of 25% found for Acheta compares with 37% in the ITS 1 region from the orthopteran genus Melanoplus (Kuperus and Chapko 1994).

The consistency between ITS data and expansion segment data found here confirms that our expansion segment data are a representative sample for the DNA composition in the neutrally evolving parts in the genomes of the species under consideration. The obvious difference of average A+T content of approximately 20% in these regions between Dipteran (except Culex) and all other non-Dipteran hexapod taxa indicates that directional mutation pressure in the Diptera is to a considerable extent biased toward A+T.

Since almost all major Dipteran lineages analyzed show the same type of compositional bias, one can infer that this bias should have originated during the evolution of its stem lineage. This inference can be directly tested by analyzing the character state changes along the respective branches of the tree. For this purpose, we have counted the numbers and types of substitutions that have taken place along the branches in different parts of the ML 28S rDNA tree, using the criterion of MP and analyzing stem and loop regions separately (table 4). The relative proportion of G+C to A+T substitutions is denoted as \( u \) and that of A+T to G+C as \( v \). The ratio of \( u/v \) in the helical regions is found to be 3.1 and 3.6 in the Dipteran and non-Dipteran lineages, respectively, but two times higher, namely 7.3, in the Dipteran stem lineage. The same tendency can be observed in the loop regions, where a \( u/v \) ratio of 1.0 and 1.4 in Dipteran and non-Dipteran insects can be compared to 1.8 in the Dipteran stem lineage. This episodic change of substitution equilibria inferred for the evolution of the Dipteran stem lineage most likely reflects the onset of a directional mutation bias toward A+T residues during the evolution of this lineage.

### Discussion

We provide evidence for an episodic acceleration of substitution rate accompanied by a generally elevated A+T content in Dipteran rDNA, which most likely originated during the evolution of the Dipteran stem lineage. In this section, we discuss conclusions on the types of selection pressure acting on ribosomal genes in general, the evolutionary mechanism which may have led to the observed episodic acceleration of Dipteran rDNA evolution, and implications for molecular phylogenetic analysis.

#### Selective Constraints on rDNA Genes

The 18S and 28S rRNAs are functionally involved in different aspects of the translational process within the ribosome (Dahlberg 1989). Still, they appear to evolve under very similar constraints within an organism, which is in part explained by the close genomic and functional linkage of these genes, but also by selective forces which seem to act on the rRNA molecules in a general manner.

Most conspicuous in this respect is the homogeneity of nucleotide composition when loop regions are compared which show a generally elevated level of adenosine content. For the 18S gene, this shift was proposed to be a reflection of the fact that the interaction of the loop regions with the hydrophobic ribosomal proteins is facilitated by the excess of adenosine residues, since this base exhibits the least polarity (Gutell et al. 1985). In combination with a recent report on a similar excess of adenosine in the mitochondrial 16S rDNA (Springer, Hollar, and Burk 1995), our data on the 28S gene confirm this hypothesis.

Another well-characterized constraint lies on the stem regions of the rRNA, which have to provide a stable tertiary structure to the functional molecule (Gutell, Larsen, and Woese 1994). An apparent consequence of this is that the G+C content is higher than in the loop regions, since G-C hydrogen bonds are thermodynamically more stable than A-U or G-U hydrogen bonds. The
upward shift of A+T content observed in Dipteran 28S rDNA seems therefore counterintuitive at first sight. On the other hand, it has also been speculated that the stem regions must allow flexibility for biologically necessary conformational changes (Gutell, Larsen, and Woese 1994), indicating that there might exist a lower as well as an upper evolutionary compatible limit for the G+C content in stem regions. The taxon-specific differences observed within this range would then be the consequence of changes in mutational pressure rather than changes in selection pressure on the thermodynamic stability of the ribosomal complex.

Episodic Rate Acceleration as a Result of Changing Directional Mutation Bias

Different explanations are possible for the observation of an episodic change of substitution rate. These may be categorized into selectionist, life history related, or DNA-turnover-related scenarios. The selectionist scenario would assume that positive selection leads to an adaptation of a certain molecule and thus to changes in the sequence of the gene that encodes it. The changes would occur mainly during the phase of adaptation and would thus be episodic. This proposal was put forward as a potential general mode of protein evolution (Gil-lespie 1984). However, this model has recently been challenged, since simulation studies in a more realistic phylogenetic framework showed that the variance observed may not be significant in most cases (Goldman 1994). On the other hand, there are some clear examples of episodic acceleration of substitution rates driven by selection. These include the evolution of lysozymes in higher mammals (Stewart, Schilling, and Wilson 1987; Messier and Stewart 1997), the evolution of the hyper-variable domain of the HLA II molecule (Hughes and Nei 1988), and the evolution of the vesicular stomatitis virus in cattle (Nichol, Rowe, and Fitch 1993). Still, these are clearly examples that were caused by adaptation to environmental challenges. Such challenges are very unlikely to have occurred for the rDNA genes in view of the universally conserved function of these molecules and the relative constancy of their cellular environment, at least when related to the insect case described here.

Potential life-history-related explanations would include historic differences in either generation time or number of germ cell divisions per generation (Wu and Li 1985; Martin and Palumbi 1993). In fact, the ecology and physiology of the Dipteran stem lineage representatives are largely unknown. On the other hand, if life history factors had caused the episodic change in substitution rate, one would expect that the mitochondrial and the nuclear genomes would have been similarly affected, at least within a four-fold range, due to the difference in effective population size of nuclear and mitochondrial genes. However, this is not the case. Although a recent study shows that the evolutionary rate of mitochondrial 16S rDNA sequences of insects is generally faster than that of other metazoans, it is not the Diptera, but the hymenopteran and lepidopteran taxa that show the fastest rates (De Rijk et al. 1995). This indicates that the evolutionary rates of nuclear and mitochondrial genes are not coupled, and it would largely rule out life history factors as an explanation.

It is therefore most likely that the observed episodic evolution is a result of changes in a part of the nuclear DNA turnover machinery, namely replication or repair enzymes. Sueoka (1993) has shown by using analytical derivation and simulation techniques that episodic rate changes can be the consequence of a change in directional mutation bias. If the balance between $u$ and $v$ substitutions is changed in a constantly evolving DNA sequence, a sudden increase of substitutions is caused that decreases exponentially with time until a new substitutional equilibrium has been reached. This would result in a shift of DNA composition to a value that is proportional to the $u/v$ ratio. The actual extent of the correlated episodic changes in substitution rate and nucleotide composition is modulated by the negative selection pressure on the sequence. One prediction of this model is the occurrence of asymmetric branch lengths in phylogenetic trees correlated with a change in DNA composition (Sueoka 1993), which exactly fits the observations found for the Diptera. Taking all the evidence together, the Diptera presumably represent a real example of extreme episodic substitution rate enhancement due to a change in directional mutation pressure.

Implications for Molecular Phylogenetic Analysis

A better understanding of the evolutionary parameters of molecular sequence change may help to evaluate their utility for molecular phylogenetic reconstructions. As pointed out by Sueoka (1993), episodic accelerations of substitution rates which result from changing directional mutation pressure lead to asymmetrical gene trees. This will negatively influence the ability of molecular phylogenetic reconstruction methods to recover the true tree, a problem commonly known as the unequal-rate effect (Swofford and Olsen 1990). The effects of unequal rates of internal branches in more complex topologies have been studied in a few cases and found to be a worst-case scenario for the performance of most molecular phylogenetic tree reconstruction methods (Rohlf et al. 1990; Kim, Rohlf, and Sokal 1993; Kuhner and Felsenstein 1994). This potential pitfall has to be taken into consideration with respect to the molecular reconstruction of the controversial origin of the Diptera among holometabolous insects using the nuclear encoded rDNA genes (Carmean and Crespi 1995). An additional problem may be the associated shifts of DNA composition, since convergent changes of relative A+T content have also been reported to introduce an artificial bias in molecular phylogenetic tree reconstruction (Hasegawa and Hashimoto 1993). The claim of a sister group relationship between the Diptera and the peculiar insect order Stenoptera (Whiting and Wheeler 1994) based on 18S rDNA sequences must be seen in the light of these problems, since a recent study shows that the rDNA genes of the latter group have also evolved under a strong compositional bias towards A+T (Chalwatzis et al. 1994). In view of the possibility that this might have occurred independently in both groups, these results of
the molecular phylogenetic reconstruction must be treated with caution.

Acknowledgments

We are indebted to A. v. Haeseler for suggestions on the manuscript; R. Sommer, M. Retzlaff, B. Haudorf, M. Schacht, F. Reiss, and P. Makings for providing samples; and S. Pääbo for providing some of the primers. This work was supported by a Ph.D. studentship of the University of Munich to M.E and by the Gerhard Hess prize of the DFG to D.T.

LITERATURE CITED


652 Friedrich and Tautz


PAUL M. SHARP, reviewing editor
Accepted March 3, 1997