

How Slippage-Derived Sequences Are Incorporated into rRNA Variable-Region Secondary Structure: Implications for Phylogeny Reconstruction

John M. Hancock and Alfried P. Vogler^{*,1}

Comparative Sequence Analysis Group, MRC Clinical Sciences Centre, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0NN, United Kingdom; ^{*}Department of Entomology, The Natural History Museum, London SW7 5BD, United Kingdom; and Department of Biology, Imperial College at Silwood Park, Ascot SL5 7PY, United Kingdom

Received April 28, 1999; revised July 16, 1999

We analyzed the type and frequency of mutational changes in hypervariable rRNA regions, using the highly length-variable region V4 of the small subunit rRNA locus of tiger beetles (Cicindelidae) as an example. Phylogenetic analysis of indels in closely related species showed that (1) most indels are single nucleotides (usually A or T and sometimes G) or dinucleotides of A and T. These occur at numerous foci, and they exhibit a strong bias for duplication of 5' single and di-nucleotide motifs but not 3' motifs. (2) Insertions/deletions in stem-forming regions affected paired and unpaired bases with about equal frequency but they did not disrupt the secondary structure. (3) Recurring mutations involving short repeats of the same bases caused parallel evolution of similar sequence motifs in the rRNA of different lineages. The observed types of change are consistent with the proposition that slippage is the main mutational mechanism. Slippage-derived sequences tend to be self-complementary, and therefore the stem-loop structure could be self-organizing as a consequence of the underlying mutational mechanism. Thus, the secondary structure in the cicindelid V4 region may be conserved due to the dynamics of the mutational mechanism rather than to functional constraints. These processes may also have a tendency to produce similar primary sequences irrespective of phylogenetic associations. The findings have implications for sequence alignment in phylogenetic analysis and should caution against the use of secondary structure to improve the determination of positional homology in hypervariable regions. © 2000 Academic Press

Key Words: compensatory slippage; slipped-strand mispairing; ribosomal RNA; RNA secondary structure evolution; DNA sequence alignment; self-organization; Cicindelidae.

INTRODUCTION

Extreme length variation and differences in rates of nucleotide variation frequently compromise the utility of ribosomal DNA (rDNA) in phylogeny reconstruction (e.g., Aguinaldo *et al.*, 1997; RuizTrillo *et al.*, 1999; Whiting *et al.*, 1997). The molecular processes that lead to the idiosyncratic variation of rRNA molecules are poorly understood, but their investigation could potentially be of great help in phylogenetic reconstruction. Core regions of rRNA appear to evolve primarily by point mutations [in most eukaryotes, in the context of homogenization of the rDNA multigene family by concerted evolution (Dover, 1982)]. In contrast, much of the sequence variation in the variable regions is considered to be the result of "slippage" (Hancock and Dover, 1988, 1990; Nunn *et al.*, 1996; Vogler *et al.*, 1997), a mutational process involving the mispairing of the DNA duplex during replication (Levinson and Gutman, 1987). Whereas slippage presumably leads to the abundance of simple sequences in eukaryotic genomes (Hancock, 1995; Levinson and Gutman, 1987), the incorporation of slippage-derived DNA sequences into rRNA genes appears to be constrained by the need for maintaining a secondary (and tertiary) superstructure required for rRNA function in the ribosome. The putatively slippage-derived indels could disrupt secondary structure foldings of rRNA molecules unless insertions and deletions in one strand of a paired structure were compensated by similar length variation in the second strand. Such compensation ("compensatory slippage" of Hancock and Dover, 1990) has been attributed to selection for the maintenance of Watson–Crick pairings between participating secondary structural elements.

The apparent need for the preservation of secondary structures is also reflected in a tendency for covariation of intramolecularly paired bases. The analysis of covariation made an important contribution to current structural models (Gutell *et al.*, 1994). These models have been used in phylogenetic reconstruction to aid

¹ To whom correspondence should be addressed. Fax: +44-171-938-8937. E-mail: a.vogler@nhm.ac.uk.

DNA sequence alignment and have been claimed to improve the assignment of nucleotide homology over those based on primary sequence alone (Hickson *et al.*, 1996; Kjer, 1995; Morrison and Ellis, 1997; Titus and Frost, 1996). The rationale for this claim is that the conservation of a functionally constrained secondary structure permits recognition of homologous regions of the molecule even if they have undergone nucleotide changes to an extent that similarity at the level of the primary sequence is erased. Therefore, secondary structures are considered a better indicator of common ancestry than the (presumed more variable) primary sequence.

By causing length variation and increased levels of repetitiveness, slipped-strand mispairing would affect the secondary structure and functional properties of rRNA molecules, but the way in which this mutational mechanism affects rRNA evolution has not been investigated in detail. For example, the sequences investigated in an early study (Hancock and Dover, 1990) were too distant evolutionarily to allow detailed comparisons of the gain and loss of repetitive sequences and their contribution to the evolution of secondary structure. We recently investigated the variable region V4 of the SSUrRNA of tiger beetles (Coleoptera: Cicindelidae) with respect to its phylogenetic information content (Vogler *et al.*, 1997) and secondary structure (Hancock and Vogler, 1998). This region is highly expanded compared to closely related coleopteran groups, possibly due to the action of slippage, and is therefore suited to this type of analysis. We analyzed the kind and frequency of mutational changes in this variable region of rRNA, as a first step toward characterizing the contribution of slippage to rRNA evolution.

MATERIAL AND METHODS

DNA Sequence Alignment and Definition of Indel Regions

From an existing data base that included 27 ingroup taxa and representatives of all major groups in the Cicindelidae (Vogler *et al.*, 1997) (Fig. 1), we studied individual insertion/deletion events in four small clades of DNA sequences that could be aligned unambiguously over most of their length. These were (A) *Cicindela repanda*/*C. dorsalis*/*C. equestre*, (B) *Cenothyla consobrina*/*Odontocheila confusa*/*O. cayennensis*/*Pentacoma rugipennis*/*Pometon singularis*, (C) *Physodeutera alluaudi*/*Peridexia fulvia*, and (D) *Pseudoxycheila bipustulata*/*Oxycheila nigroaenea*. Character changes in this region can be assessed based on a tree that was derived independently from the less alignment-variable V2 region of the same molecule and mitochondrial 16S rRNA (Vogler and Pearson, 1996); this tree was also the basis for the establishment of groupings A through D. None of the more distantly related sequences of the V4 region could be aligned unambigu-

ously, given the criteria set out below, and therefore had to be excluded from the analysis.

Starting with the published global alignment (Vogler *et al.*, 1997), the focal sequences were realigned to reflect homology (common ancestry) between close relatives (Fig. 2). Homology assignment of nucleotide positions was based on sequence similarity. Equating similarity with homology is a reasonable assumption in this case, given that the sequences in each of the four clades had been shown to be closely related, based on independent sequence data: following De Pinna (1991) and Brower and Schawaroch (1996), similar sequence motifs represent synapomorphies and thus homologies of the respective clade. As the criterion for the identification of homologous insertion/deletion motifs based on similarity, we (arbitrarily) required that insertions/deletions (present in at least one species of the focal clade) had to be bound by motifs on both 5' and 3' sides that were identical between all species at two out of three positions in the multiple alignment. The motif bounded by these regions was termed the "indel motif." By applying this criterion we could be confident that indels were defined consistently, but we do not exclude the possibility that other alignments and indel definitions could be derived for the same set of sequences. Where the indel motifs at a particular position differed between species (labeled # in Fig. 2), no attempt was made to align them or to attribute character state transformations between them, because of the difficulty of inferring multiple insertion/deletion events.

Secondary Structure Predictions

Mapping of indels onto the secondary structure was carried out based on a secondary structure model proposed previously (Hancock and Vogler, 1998), established with the use of the MFOLD algorithm (Jaeger *et al.*, 1990; Zuker, 1989, 1994; Zuker *et al.*, 1991) and taxonomic comparisons. Because of the large differences in the primary sequence in different cicindelid species, in particular in the region of stem II (Hancock and Vogler, 1998), secondary structures were modeled for each of the taxa. All secondary structure analyses using the MFOLD algorithm were carried out using the MFOLD server at <http://mfold1.wustl.edu/~mfold/rna/form1.cgi>. All structures falling within the 10% window of stability predicted by the program were retained. As well as the predicted structures, values of P-num (the number of bases with which each base in the sequence paired in the optimal and all suboptimal structures with estimated stabilities within 10% of the optimal structure) and S-num (the equivalent measure of the number of occasions on which each base was unpaired in the same set of structures) were recorded for all bases. P-num values were used to calculate H-num for the optimal helical structure, where for any predicted base pair

$$\text{H-num (i, j)} = \text{P-num (i)} + \text{P-num (j)} - 1$$

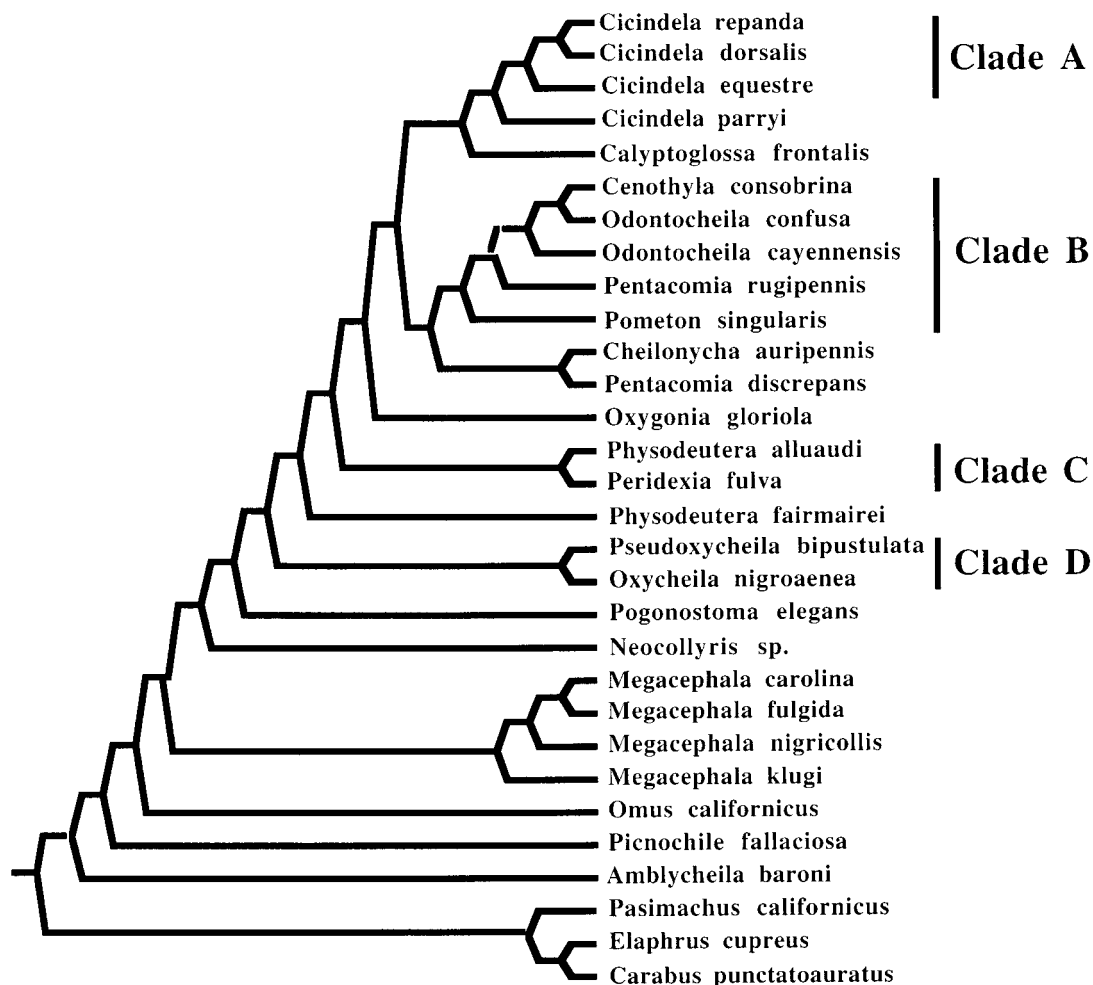


FIG. 1. Phylogenetic relationships in 27 taxa of Cicindelidae (redrawn from Vogler and Pearson, 1996). Four subgroups used for the analysis of sequence variation in this study, Clades A–D, are marked by vertical bars.

for a pair of bases *i* and *j* (Zuker and Jacobson, 1998). For a predicted helix, H-num is then the mean H-num for all predicted base pairs.

RESULTS

A total of 44 indels that conformed to the criteria outlined above were included in the analysis (Fig. 2). Of those, the most common classes of indel involved singlets or doublets of A or T (27 of 44 changes, 61%) (Table 1). Other types observed more than once were insertion/deletion of G (4 changes, 11%) and AT or multiples thereof (3 changes, 8%), whereas only 5 indels involved more complex motifs, some of which also included repetitive tracts of mono- and dinucleotide motifs.

Slippage should result in duplication of motifs immediately adjacent to inferred indel motifs. We therefore scored indels for whether they were duplications of motifs lying immediately 5' or 3' to them; 23 of the 44 indel motifs (52%) were in fact identical to the motif

immediately 5' to them. In addition, 2 indels, indel 12 of clade A (G inserted/deleted) and indel 2 of clade B (T inserted/deleted), gave rise to an elongation of a neighboring tandem repeat. Two of the longer and more complex indels [(TA)₃ in indel motif 7 of clade C and (TA)₂G in indel motif 3 of clade B] were also essentially duplications of the 5' adjacent motifs. Because the probability of flanking nucleotides being identical to indel nucleotides is high, especially for A or T indels in an AT-rich sequence, the frequencies with which A and T indels were found associated with different bases 5' and 3' to them were subjected to χ^2 analysis. Expected frequencies were calculated based on the base composition of the total data set, by calculating first the frequencies for each of the four clades of Fig. 1 and then using the summed base frequency over all clades (clades A through D of Fig. 1) as expected values in the overall χ^2 test. Observed frequencies of A and T indels being adjacent to identical bases (either 5' or 3') were 8/13 and 12/14, respectively, compared to expected values of 4.16 and 5.87. This gave a χ^2 of 15.77 and $P =$

TABLE 1
Types of Indel Motif Observed

Motif	Clade A	Clade B	Clade C	Clade D	Total
A	8	1	2	1	12
AA		1			1
T	2	3		3	8
TT	2	1	1	2	6
G	1	1	1		3
GG		1			1
AT			1		1
TA	1				1
(TA) ₃			1		1
CA	1				1
TG		1			1
ATT			1		1
TTA		1			1
CTC			1		1
C ₂ T ₄		1			1
(AT) ₂ G		1			1
T ₂ (AT) ₂			1		1
TA(TG) ₃ TT(GT) ₂		1			1
T ₂ ACAT(AC) ₃ A ₂ T ₂ (AT) ₂ A		1			1
Total	15	14	9	6	44

3.8×10^{-4} ($2 df$). Thus, there was a clear overrepresentation of duplicated motifs in the indel regions. This is expected if slippage mechanisms are involved in generating this sequence variation.

We also investigated the effect of base changes with respect to their secondary structure interactions. A prerequisite for the analysis of intramolecular pairings is a good understanding of the secondary structures of the molecules under comparison. According to previous MFOLD analysis, the predicted secondary structure for the part of the V4 region under investigation here in each case corresponds to a single long stem-loop structure (Hancock and Vogler, 1998). However, because of the inherent uncertainties of rRNA secondary predictions, we further investigated the properties of these MFOLD structures by characterizing all suboptimal structures falling within the 10% window of stability predicted by the program. All sequences except that for *P. singularis* (Clade B) formed only a single structure within this window, indicating that the structures of these regions are mostly "well-determined" (i.e., they are of substantially lower energy than suboptimal foldings obtained from the same sequences) (Zuker and Jacobson, 1995, 1998). We further determined the mean value of the parameter H-num (Zuker and Jacobson, 1998) for these structures, which represents the potential for ambiguity of formation of base pairs. A value of 1 for H-num indicates that each base in a structure, on average, pairs with only one other base within the set of most stable structures defined by MFOLD, a value of 2 that they pair on average with two bases, and so on. Low values of mean H-num indicate a high probability of the structure being correct and of being observed in covariation analysis (Zuker and

Jacobson, 1998). For our set of sequences, mean H-num values varied between 1.145 (for *C. consobrina*) to 2.431 (for *P. singularis*) with an overall mean of 1.394 (SD 0.358). Based on these analyses, it was concluded that the structures were predicted with sufficient certainty that they could be used to investigate the effects of indels on the RNA secondary structure of this region. The *P. singularis* sequence was an exception and was excluded from further analysis.

The degree of pairing for the indel bases was scored based on two parameters calculated by MFOLD: P-num, which indicates the degree of ambiguity of base pair formation for a particular base (where P-num = 1 if a base forms only one base pair in the set of most stable structures), and S-num, which represents the frequency with which a base is paired (S-num = 0) or single stranded (S-num > 0). The mean value of P-num for all indel bases was 1.333, while that for all bases in all sequences except that for *P. singularis* was 1.075, indicating that pairing of indels tended to be more ambiguous than that of conserved bases. The mean S-num score for indel bases was 0.379 compared to a mean for all bases of 0.226, indicating a higher than average propensity for indel bases to be unpaired. Scoring indel bases for pairedness (S-num = 0) or single-strandedness (S-num > 0) on the predicted structures showed that 36 of 58 indel bases (i.e., the total number of bases found in the 44 indels scored) were paired (62%).

In addition to these summary statistics, we were also interested in the state of pairedness of indel motifs in sister taxa comparisons. Each insertion/deletion event was classified according to whether the site at which it occurred (i.e., the bases flanking the indel) was paired or unpaired and whether the indel motif was paired or unpaired in the species in which it was present. In cases in which indels were more than a single base long and different bases in the inserted sequence showed different states of pairedness, the score for that site was divided among the different segments of a contingency table in proportion to the numbers of bases in the different categories. Thus, each indel made an equal contribution to the final analysis, although they may in fact have resulted from more than one event. These calculations resulted in the following findings: in approximately 50% of cases (21.88/44) the pairing state of the region was the same whether or not the inserted/deleted sequence was present. Of these cases, 73% (15.87) occurred in paired regions. The numbers of sites paired in the absence of the indel motif but unpaired in its presence (i.e., sites which would be disrupted by an insertion event; 10.08/44 = 23%) and unpaired in the absence of the motif but paired in its presence (i.e., sites stabilized by an insertion event; 11.29/44 = 26%) were approximately equal. Thus, about half of the insertion/deletions had no effect on the secondary structure of the region in which they occurred, and of those that did

affect structure, approximately equal numbers tended to create and destroy base pairing.

DISCUSSION

The Effects of Slippage on rRNA Evolution

Phylogenetic analysis of the V4 region indicated that most of the observed mutational changes affect mono- and di-nucleotide repeats, which can be understood as duplications of the bases 5' to them. This observation is consistent with slippage as the underlying mechanism for the generation of indels in these sequences. A degradation of strictly duplicated motifs is to be expected as a result of point mutations in and adjacent to indels but we cannot exclude the possibility that other (unknown) processes that do not depend on identity between inserted and flanking bases also play a role in generating these indels.

Most indel bases appeared to be involved in base pairing within stems, but their pairing partners were identified by MFOLD as being more ambiguous than the pairing partners of non-indel bases, and a higher proportion of them were unpaired than was the case for bases in the structure overall. Thus, although our data indicate that indels are incorporated into stable secondary structures, it is also clear that many indels disrupt structures. This observation is of general importance, as this suggests that pairedness is not a strongly conserved feature; studies that compare the rate of changes in stems and loops or bulges (e.g., Vawter and Brown, 1993) need to take into account the fluid nature of the state of pairedness when changes in the primary sequence occur and the general ambiguity of predicting pairedness in indel-prone regions.

However, the incorporation of unpaired indels, despite upsetting the state of pairedness locally, does not seem to disrupt the overall secondary structures. This is due to the incorporation of indels at multiple foci, which have a tendency to compensate for the local disruption of pairings. The patterns of sequence variation therefore suggest a two-phase model of incorporation of indels into variable-region secondary structure, whereby indels first affect single-stranded regions and then become incorporated into double-stranded regions after the generation of base complementarity by subsequent slippage-derived or other point mutations (Fig. 3). The overall effect of this would be a pattern described previously as compensatory slippage (Hancock and Dover, 1990). Whereas most cases in our data set conform to the examples in Fig. 3, in which small mono- or di-nucleotide motifs are incorporated, in the *P. singularis* sequences two long indels pair with one another (Fig. 4). It is conceivable that these more complex indels also derive from very few mutational events.

If, as our findings suggest, a significant amount of slippage-derived variation is not compensatory and indels frequently disrupt the stem-loop structure, this

calls into question the assumption that selection is responsible for the conservation of secondary structure in this region, at least at the level of individual base pairs. Nonetheless, the secondary structure of the V4 region appears to be well maintained across cicindelids. Gerbi (1986) suggested that the conservation of secondary structural elements within variable rRNA regions is not necessarily an indication of a functional role. Our findings raise the possibility that this conservation could be a direct consequence of the mutational bias due to slippage and that the stem-loop structure undergoes neutral or near-neutral variation. The bases most often found in indels in these sequences (A and T) are also those most often found in noncoding regions in eukaryotic genomes and are those most likely to undergo slippage (Hancock, 1995). Repetitive motifs are also similar throughout the slippage-prone region, and the incorporation of bases into these sequences obviously occurs at multiple foci. Because of this pattern of variation, an overall self-complementarity of the sequence can result, which could lead to a high level of intrahelical pairing of bases regardless of the exact primary structure (Fig. 3). Secondary structure could therefore form as a by-product of the underlying mutational processes and not as the result of selection (for functional stem-loop elements or otherwise). Thus, secondary structure formation becomes an intrinsic, self-organizing property of single-stranded RNA molecules.

Mutational biases of this kind should not only narrowly affect a particular region of a small part of the gene but would have implications for the evolution of the entire molecule. One of the striking features of sequence variation in rRNA genes is the difference in the level of sequence divergence in the conserved core regions (the regions also present in prokaryotic rDNA) and the more variable expansion segments. In addition to the V4 region, which is the most variable, there are only two other regions in the molecule that exhibit high levels of sequence variation in Coleoptera, the variable regions V2 and V7. Both form a stem-loop structure according to minimum-energy predictions, and a preliminary assessment of the sequence variation indicates patterns of variation in primary sequence and secondary pairing structures generally similar to those described for V4. In contrast, the core regions exhibit very little variation (in length and nucleotide sequence) within the Cicindelidae and therefore no analysis of the mutational variation is possible. However, the evolutionary dynamics of core regions and variable regions may not be entirely independent. The rate of sequence evolution observed in the variable regions in the Cicindelidae is much elevated in comparison to the closely related Carabidae. This increased rate is not limited to the variable regions, however; when these highly variable regions are eliminated from phylogenetic analysis that includes a wide range of related families in the coleopteran suborder Adephaga, the Cicindelidae are

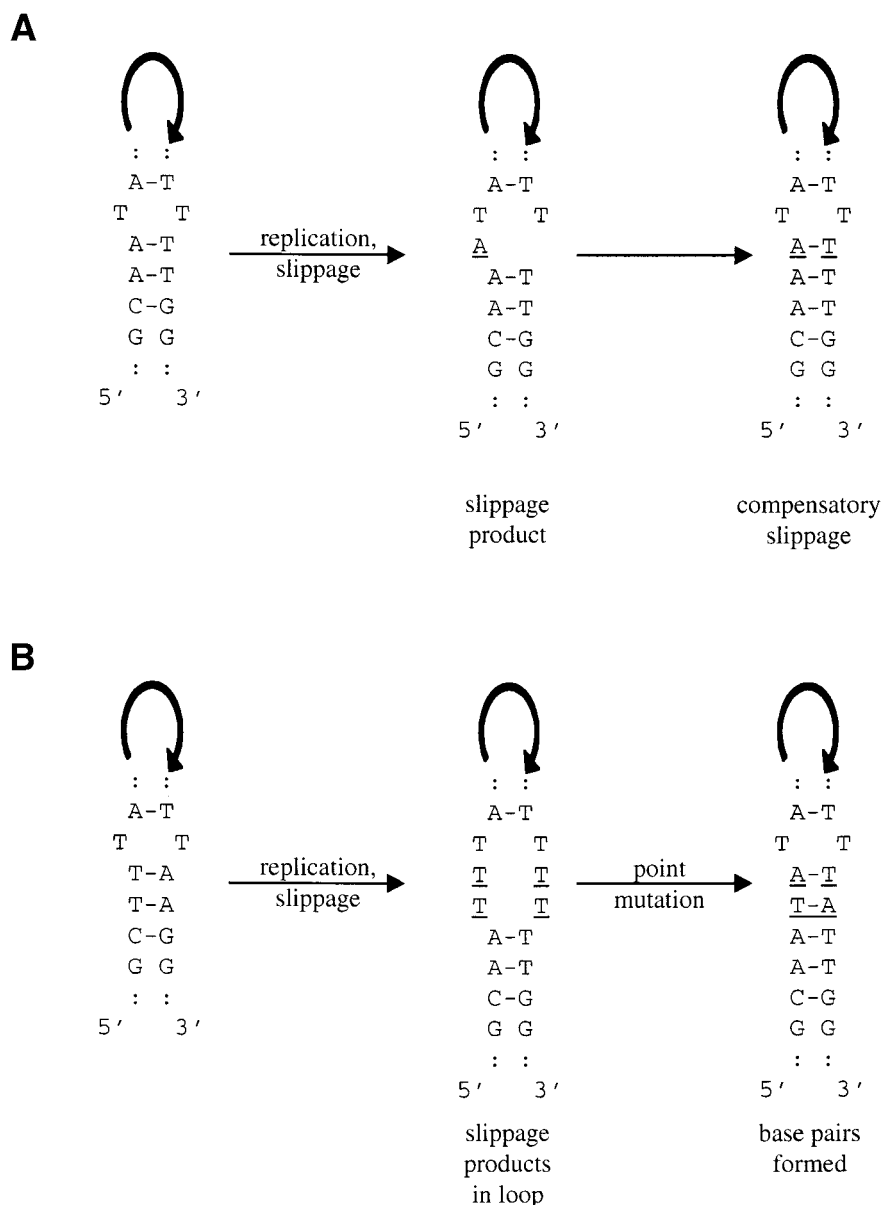


FIG. 3. Models for the effect of slippage on the secondary structure of a stem-loop structure in rRNA. (A) Compensatory slippage may result from two independent slippage events which have a high chance of being self-compensatory in sequences with strong bias in base composition. (B) Slippage may enlarge loop regions preferentially. The new sequences produced in this way may become incorporated into stem regions as a result of subsequent point mutations. Newly incorporated bases are underlined.

still very divergent (V. Shull and APV, unpublished), indicating that mutational biases can also affect sequences that are functionally highly constrained.

Implications for Phylogeny Reconstruction

Lack of covariation of indel motifs at different foci in the stem-loop regions of variable regions would reduce the problem of nonindependence in phylogenetic reconstruction (see Hillis and Dixon, 1991; Vawter and Brown, 1993; Wheeler and Honeycutt, 1988). Nonetheless, slippage-derived character variation is likely to result in potentially misleading phylogenetic signal. The prevalence of certain types of character changes,

confined to a limited number of sites in the primary structure, is likely to produce high levels of homoplastic variation, and even rather large indels which would appear unlikely to arise multiple times by chance may not be free of homoplasy (Fig. 4). If homoplastic character changes affect variation only locally and result in the recurring gain and loss of repeat motifs in particular positions, this would simply produce stochastic "noise" as seen in other types of sequence data. However, slippage-derived mutations and their tendency to repeat existing sequence motifs are likely to favor certain types of motifs throughout a larger DNA region (Hancock, 1995; Hancock *et al.*, 1999), exacerbating

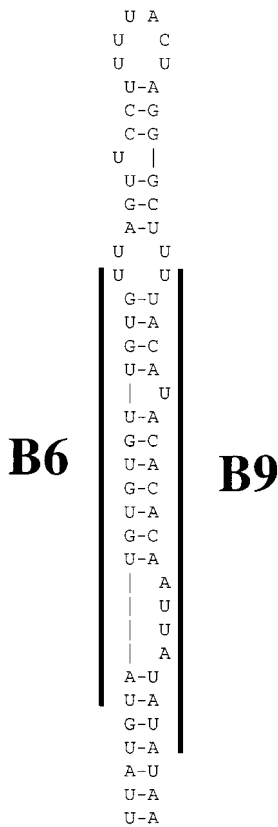


FIG. 4. Predicted minimum-energy secondary structure for the central part of the *P. singularis* V4 region. The labeled regions of sequence correspond to indels B6 and B9 in Fig. 2.

existing biases in base composition. For example, repetition of short sequence motifs downstream from existing bases at multiple foci is likely to reinforce any existing bias in the overall GC content or may result in an increased level of simplicity throughout a DNA sequence. If several regions in a gene (or even unlinked parts of the genome) are affected by these mutational processes, a similar bias in GC content, degree of repetition, and sequence composition should arise. This possibility would also explain a common AT bias in different length-variable regions of an rRNA molecule (Hancock and Dover, 1988). For example, the V2 and V4 regions (Hancock and Vogler, 1998), as well as the highly length-variable V7 region (unpublished data) of cicindelids, are all highly AT biased, whereas the AT content of the intervening (length-invariable) regions is lower. Whether or not such biased sequence composition is due to a common phylogenetic history or is the result of homoplastic variation is frequently the cause of contention over the interpretation of phylogenetic data (Whiting, 1998).

Finally, these results have implications for the alignment of rDNA sequences, which is often guided by the position of nucleotides in a secondary structure element predicted from minimal-energy base pairings (e.g., Hickson *et al.*, 1996; Kjer, 1995). The rationale for

the use of secondary structure in the homology assignment is usually that the secondary structure is largely fixed and can therefore provide a “template” (e.g., Hickson *et al.*, 1996) to which to fit novel sequences for establishing homology. This is a problematic suggestion, as this procedure does not take into account the possibility that stems transform into loops and vice versa as the primary sequence changes. As our data confirm, this is, however, a common outcome of sequence variation, but it is not usually tested with the help of explicit algorithmic procedures (in contrast to procedures for alignments based on primary structure).

The use of secondary structure “models” as guides for sequence alignment could possibly be justified if these structures were conservative as the result of evolutionary processes that maintain them. However, the finding of a mutational process (slippage) that may be agnostic with respect to secondary structure, at least in the most length-variable and alignment-ambiguous parts of the sequence, makes this proposition problematic for precisely those regions for which it would in principle be most applicable. If secondary structure in these regions is merely an incidental result of the variation in primary structure, its inclusion in the phylogenetic analysis does not add another level of confidence to evaluating which bases are homologous between taxa. In addition, if regions aligned according to secondary structure are predominantly the product of nonhomologous parallelisms, their phylogenetic information content is low (and the laborious determination of secondary structure alignment will not aid phylogeny reconstruction to any significant extent).

In conclusion, the study of indels in length-variable rRNA regions of cicindelids revealed strong bias in type and position of the mutational events that are generally consistent with a slippage-driven process. Because highly stable stem-loop structures will form as a result of these mutational processes, the existence of stable secondary structures does not necessarily constitute evidence of selection. It will be a challenge to take slippage-derived sequence variation into account when selecting alignment protocols, especially those based on secondary structure predictions. Similarly, slippage-like variation leads to strong bias in types and rates of base changes and indels; this may result in parallelisms in primary and secondary structures which mislead phylogenetic reconstruction.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council of the UK (J.M.H.) and the Leverhulme Trust (A.P.V.; Grant F/696/H). We are grateful to two anonymous reviewers for useful comments on the statistical analysis of base composition and the secondary-structure-based alignment. We also thank the reviewers and editors evaluating earlier versions of this manuscript for demonstrating to us the great diversity of thought about rRNA evolution.

Note added in proof: In a recent study, E. A. Schultes, P. T. Hrabec, and T. H. LaBean (“Estimating the contribution of selection and self-organization in RNA secondary structure,” *J. Molec. Evol.* **49**:76–

83) conclude that most of the conformational structure of functional RNAs "results not from evolutionary optimization but from constraints imposed by rules intrinsic to RNA polymer folding." Using random permutations of natural RNA sequences, their study also provides evidence for self-organization of complex secondary structures in single stranded RNAs, due to the minimization of unfavorable interactions in stable structures, and thus underscores the limited evolutionary relevance of these structures.

REFERENCES

- Aguinaldo, A. M. A., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A., and Lake, J. A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**: 489–493.
- Brower, A. V. Z., and Schawaroch, V. (1996). Three steps of homology assessment. *Cladistics* **12**: 265–272.
- De Pinna, M. C. C. (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics* **7**: 367–394.
- Dover, G. A. (1982). Molecular drive: A cohesive mode of species evolution. *Nature* **299**: 111–117.
- Gerbi, S. (1986). The evolution of eukaryotic ribosomal DNA. *Biosystems* **19**: 247–258.
- Gutell, R. R., Larson, N., and Woese, C. R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**: 10–26.
- Hancock, J. M. (1995). The contribution of slippage-like processes to genome evolution. *J. Mol. Evol.* **41**: 1038–1047.
- Hancock, J. M., and Dover, G. A. (1988). Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol. Biol. Evol.* **5**: 377–391.
- Hancock, J. M., and Dover, G. A. (1990). "Compensatory slippage" in the evolution of ribosomal RNA genes. *Nucleic Acids Res.* **18**: 5949–5954.
- Hancock, J. M., and Vogler, A. P. (1998). Modelling the secondary structures of slippage-prone hypervariable RNA regions using suboptima folding and statistical testing: The example of the tiger beetle 18S rRNA variable region V4. *Nucleic Acids Res.* **26**: 1689–1699.
- Hancock, J. M., Shaw, J., Bonneton, F., and Dover, G. A. (1999). High sequence turnover in the regulatory regions of the developmental gene hunchback in insects. *Mol. Biol. Evol.* **16**: 253–265.
- Hickson, R. E., Simon, C., Cooper, A., Spicer, G. S., Sullivan, J., and Penny, D. (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* **13**: 150–169.
- Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66**: 411–453.
- Jaeger, J. A., Turner, D. H., and Zuker, M. (1990). Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol.* **183**: 281–306.
- Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Mol. Phylogenet. Evol.* **4**: 314–330.
- Levinson, G., and Gutman, G. A. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**: 203–221.
- Morrison, D. A., and Ellis, J. T. (1997). Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. *Mol. Biol. Evol.* **14**: 428–441.
- Nunn, G. B., Theisen, B. F., Christensen, B., and Arctander, P. (1996). Simplicity-correlated size growth of the nuclear 28S ribosomal RNA D3 expansion segment in the Crustacean order Isopoda. *J. Mol. Evol.* **42**: 211–223.
- RuizTrillo, I., Riutort, M., Littlewood, D. T. J., Herniou, E. A., and Baguna, J. (1999). Acoel flatworms: Earliest extant bilaterian metazoans, not members of Platyhelminthes. *Science* **283**: 1919–1923.
- Titus, T. A., and Frost, D. R. (1996). Molecular homology assessment and phylogeny in the lizard family Opluridae (Squamata: Iguania). *Mol. Phylogenet. Evol.* **6**: 49–62.
- Vawter, L., and Brown, W. M. (1993). Rates and patterns of srDNA change. *Genetics* **134**: 597–608.
- Vogler, A. P., and Pearson, D. L. (1996). A molecular phylogeny of the tiger beetles (Cicindelidae): Congruence of mitochondrial and nuclear rDNA data sets. *Mol. Phylogenet. Evol.* **6**: 321–338.
- Vogler, A. P., Welsh, A., and Hancock, J. M. (1997). Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). *Mol. Biol. Evol.* **14**: 6–19.
- Wheeler, W. C., and Honeycutt, R. L. (1988). Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Mol. Biol. Evol.* **5**: 90–96.
- Whiting, M. F. (1998). Long branch distraction and the Strepsiptera. *Syst. Biol.* **47**: 134–137.
- Whiting, M. F., Carpenter, J. C., Wheeler, W. D., and Wheeler, W. C. (1997). The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* **46**: 1–68.
- Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* **244**: 48–52.
- Zuker, M. (1994). Prediction of RNA secondary structure by energy minimization. *Methods Mol. Biol.* **25**: 267–294.
- Zuker, M., and Jacobson, A. B. (1995). "Well-determined" regions in RNA secondary structure predictions: Analysis of small subunit ribosomal RNA. *Nucleic Acids Res.* **23**: 2791–2798.
- Zuker, M., and Jacobson, A. B. (1998). Using reliability information to annotate RNA secondary structures. *RNA* **4**: 669–679.
- Zuker, M., Jaeger, J. A., and Turner, D. H. (1991). A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucleic Acids Res.* **19**: 2707–2714.