Ribosomal RNA Secondary Structure: Compensatory Mutations and Implications for Phylogenetic Analysis[1]

Michael T. Dixon and David M. Hillis
Department of Zoology, The University of Texas at Austin

Using sequence data from the 28S ribosomal RNA (rRNA) genes of selected vertebrates, we investigated the effects that constraints imposed by secondary structure have on the phylogenetic analysis of rRNA sequence data. Our analysis indicates that characters from both base-pairing regions (stems) and non-base-pairing regions (loops) contain phylogenetic information, as judged by the level of support of the phylogenetic results compared with a well-established tree based on both morphological and molecular data. The best results (the greatest level of support of well-accepted nodes) were obtained when the complete data set was used. However, some previously supported nodes were resolved using either the stem or loop bases alone. Stem bases sustain a greater number of compensatory mutations than would be expected at random, but the number is <40% of that expected under a hypothesis of perfect compensation to maintain secondary structure. Therefore, we suggest that in phylogenetic analyses, the weighting of stem characters be reduced by no more than 20%, relative to that of loop characters. In contrast to previous suggestions, we do not recommend weighting of stem positions by one-half, compared with that of loop positions, because this overcompensates for the constraints that selection imposes on the secondary structure of rRNA.

Introduction

Ribosomal RNA (rRNA) genes have become firmly established as a useful systematic data base across the entire breadth of life. Their utility is founded in the ubiquitous presence and relative conservation of many rDNA nucleotide sequences throughout life (reviewed in Hillis and Dixon 1991). However, the functional importance of rRNA also places constraints on phylogenetic analyses. rRNA forms a distinct secondary structure as part of the formation and functioning of ribosomes (Noller 1984). These structures are dependent on Watson-Crick and wobble base-pairing interactions between rRNA bases. Most phylogenetic analyses assume that the characters analyzed are evolutionarily independent; that is, for nucleic acid sequences, change in one nucleotide does not affect the probability of change in another. If stem bases evolve as pairs, then the assumption of base-pair independence is violated. Selection for maintenance of complementary bases in base-pairing regions may interfere with the pattern of independent mutations that provide useful systematic characters.

Wheeler and Honeycutt (1988) examined 5S and 5.8S rRNA molecules from a diversity of organisms. Their data suggested that an increased number of substitutions occurred among paired bases, presumably as a result of selection for compensatory mutations that maintain secondary structure. They concluded that bases from single-

1. Key words: ribosomal RNA, secondary structure, phylogeny.

Address for correspondence and reprints: David M. Hillis, Department of Zoology, The University of Texas at Austin, Austin, Texas 78712.

Mol. Biol. Evol. 10(1):256–267. 1993.

0737-4038/93/1001-0015$02.00

stranded regions (loops) rendered phylogenetic relationships that were consistent with those that resulted from traditional morphological studies. In contrast, double-stranded regions (stems) supported different phylogenetic hypotheses. In addition, when the two classes of regions were analyzed together, the abundance of stem bases (two to three times as many stem bases as loop bases) obscured the relationships determined by loop bases alone. To eliminate the problems associated with the phylogenetic analysis of stem bases, Wheeler and Honeycutt (1988) recommended eliminating these nucleotide positions or weighting them by one-half. In contrast, an analysis of echinoderm 18S rRNA sequences (Smith 1989) reported that paired nucleotides were phylogenetically informative and were more likely than were unpaired bases to include the "expected tree" (= topology from morphology) as one of the equally most parsimonious trees.

The dilemma facing molecular systematists interested in analyzing rRNA data sets is whether loop bases and stem bases should both be used in phylogenetic analyses and, if so, whether bases from each class (i.e., stems and loops) should be considered equally informative and independent. We examined the phylogenetic contents of stem regions and loop regions, to quantify the extent of compensatory mutations that occur within rRNA and to determine the effect these compensatory mutations have on phylogenetic analysis.

Material and Methods

The data used (fig. 1) were taken from a recent phylogenetic analysis (Hillis et al. 1991) of vertebrate relationships, in which 1,989 28S rDNA bases from each of the following taxa were aligned: *Cyprinella lutrensis* (Actinopterygii), *Latimeria chalumnae* (Actinistia), *Xenopus laevis* (Amphibia), *Rhineura floridana* (Squamata), and *Mus musculus, Rattus norvegicus,* and *Homo sapiens* (Mammalia). They used a composite of *Drosophila melanogaster* (Insecta) and *Lampetra aepyptera* (Petromyzontiformes) sequences as an outgroup. Their DNA sequences were aligned against the *Mus* sequence by the alignment subroutines of Pustell and Kafatos (1986), with adjustments made manually to increase similarity. Nucleotides that could not be unambiguously aligned were excluded from the analysis. A parsimony analysis of these data generated relationships consistent with traditional hypotheses of vertebrate phylogeny (tree I of fig. 2).

Hillis et al. (1991) reported the nucleotide sequence of the DNA strand synonomous with the rRNA, from which we deduced the RNA sequence. This sequence was visually compared with the model of *Xenopus* (Clark et al. 1984) and *Rattus* (Hadjiolov et al. 1984) 28S rRNA secondary structure. Each nucleotide was inferred to be a stem base or a loop base. Stem bases were defined as those that participate in base-pairing interactions. Loop bases were those that are not hypothesized to engage in base pairing in the mature rRNA.

The nucleotide sequence data were condensed to include only the positions that are phylogenetically informative in parsimony (111 characters; see fig. 1). These sites were then divided into those from stem regions and those from loop regions and were further subdivided into those that resulted from substitution events and those that resulted from insertion/deletion events. A deletion of nucleotides following base 3752 creates two characters that are presumably the result of a single event and that therefore could be considered a single phylogenetic character. Hillis et al. (1991) analyzed these bases as two characters, and we treated them similarly. Several analyses were repeated with one of these two characters deleted, and no significant changes were detected.

```
Xen model   LLSSSSSSSS SSSLLSSSLS SSSSLSLLLL SSSSSSSLSS
Rat model   SSSSLLLSSS SSSSLSLSLL SLLLLSSLLL SSSSSSSLSS
            1111111111 1111111111 1111111111 1111111111
            0001111111 1111111112 2222222233 3333333445
            9990111566 7777789991 1446689911 1222225220
            2450456801 1367970231 4462474723 9034573572
Mus         TCCCGATCCA TCGAACCGCT CCCGTCCGTG ATG-ATA-GG
Rattus      TCCCGATCCA TCGAACCGCT CCCGTCCGTG ATG-ATA-AG
Homo        TCCCGATTCA CCGAGCCGCT GCCGTCCGTG GTA-ATG-AG
Rhineura    GCCGGGCTCG TTGATC-GCG GCCCTC-ACA GCAG-TG-GG
Xenopus     GTCGACACCG CTCTGG-ACC GCGGGCGACA GCGTATG-GG
Latimeria   TCCAAAACCA TTCGTC-GTC CGGCCTCATA ACAT-TGTGA
Cyprinella  ACTGGGTGTG TTCAACGGTC CGGCCTGG-- GCGT-CG-GG
Outgroup    ATTCTATTTA TACTAGTATN NNNNNNNNNN NNNNNCATAA

Xen model   LLLLLLLLLL SLLLSSLLLL LSLSLLLLLS SSSLSLLSSL
Rat model   SSSSSSLLSL LLLLSLLLSL LSLLLLLLSS SSLLSSSSLS
            1111111111 1111122222 2222222222 2233333333
            5555555666 6678801122 2222333333 9900000222
            0001345024 4588893301 2459033346 7701333357
            7890385726 6084550826 4962817960 6789158478
Mus         AGCGGAACCA -AGGGT--CA GACAAACGGC CGGACAGGGG
Rattus      AGCGGAACCA -AGGGT--CA GATAAACGGC CGGGCAGGGA
Homo        AGCGGAACCA -AGGGT--CG GATAAACGGC CGGGCAGGGA
Rhineura    CAGTAAAGCG GAAGGT--TG GACAGACGGC CGAGCGCCAG
Xenopus     CAGGGAACAG GAAGGC--TG CGCGGGCGGC GGAGGGGCAN
Latimeria   CTCGAGACAA -AGGCC--TG GGTACGTGGC GCAAGGCGAG
Cyprinella  CGGTGGGCAT CGGTTCCCTG CGCGGATTAT ACGCGACGGG
Outgroup    AAATAAGGAN NGGTTACCNN NNNNNGCTAT NNNNNNNNNA

Xen model   LSLSLSLSSL LLLSLLLSSS LLLLSSSSLL S
Rat model   LSLSLSSSSL LLSSLLLSLS LLSSLLSSLS L
            3333333333 3333333333 3333333333 4
            3555677777 7777777778 8899999999 0
            2149112445 5556677892 6905667779 0
            5033705142 2451223097 7719262669 1
Mus         GGCCTGCCC- -TCGAGCCCG C-CGTTA--T G
Rattus      GGCCTGCCC- -TCGAGACCG C-CGTTT--T G
Homo        GGCCTCCCC- -TCGAGACGG C-CGTTT--T G
Rhineura    GATATCACC- -TCGAGACGA G-CATTA--T A
Xenopus     TACACAACC- -TCGATGCGC GGGATTAG-T A
Latimeria   GATACAACC- -TCGAACCGC G-CACTAGCT A
Cyprinella  GATATAATGG CCGTCACGGC A--ACCA--C A
Outgroup    TGGCCCATGG CCGTCGAGGA TG-GTCGGCC A
```

FIG. 1.—Data matrix of 28S rRNA sequence data used for this study. The numbers above each column are from Hillis et al. (1991) and refer either to the nucleotide or to the nucleotide that immediately precedes the insertion/deletion event. "S" and "L" indicate whether this character was treated as a stem character or a loop character, respectively, on the basis of the *Xenopus* model and the *Rattus* model.

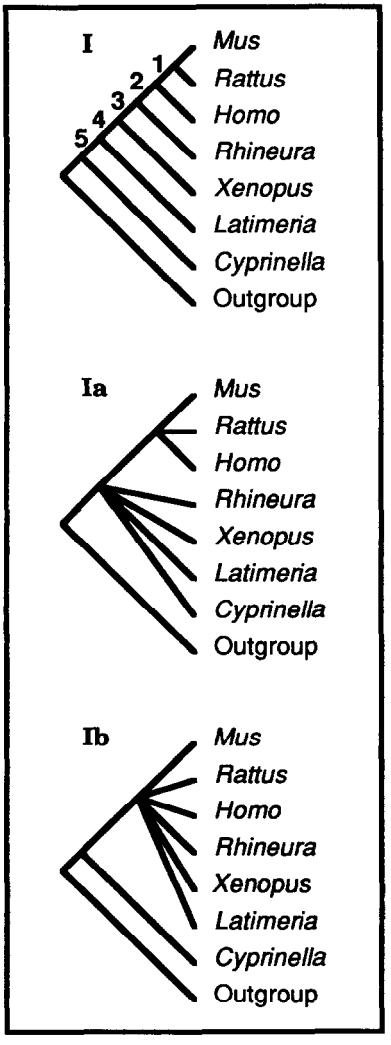

FIG. 2.—Hypotheses of vertebrate relationships obtained from the analysis of 28S rRNA sequence data. Tree I is the topology traditionally supported (numbers refer to nodes indexed in tables 2 and 3; Ia and Ib are less resolved trees that are consistent with tree I). Trees II–VI are unconventional hypotheses of these relationships, supported by various subsets of these data.

Stem bases and loop bases were analyzed separately with the exhaustive search option of PAUP version 3.0q (Swofford 1990), to determine whether they contain differing amounts of phylogenetic information. We used the same composite outgroup sequence used by Hillis et al. (1991). Agreement among the results of the various analyses was determined by visually comparing topologies, by bootstrap analysis, and by comparing the distribution of lengths of all possible trees. Skewness of tree-length distributions (as measured by the statistic g_1) is a useful indicator of phylogenetic signal in DNA sequence data (Fitch 1984; Hillis 1991; Huelsenbeck 1991). Critical values of g_1 were obtained from Hillis and Huelsenbeck (1992).

To address the question of independence of stem bases, we tallied all of the stem bases (*Xenopus* model) that had incurred substitutions. We attempted to keep this analysis independent of the phylogenetic hypothesis, by not assigning the polarity of substitution events (whenever this was possible). Base pairs that sustained more than one substitution required a phylogenetic hypothesis to determine the most likely path of transformation. In these cases, we used the phylogeny supported by these data (Hillis et al. 1991), which agrees with the relationships supported by most other authors (Romer 1966, p. 47; Rosen et al. 1981; Fritzsch 1987; Northcutt 1987; Schultze 1987). In cases in which the determination of the transformation series was ambiguous, we credited each of the shortest possible pathways equally.

We considered two classes of change within stems: substitutions that change one pair of complementary bases to another pair of complementary bases (e.g., C-G to A-U; type I) and substitutions that change one pair of complementary bases to a pair of noncomplementary bases, or vice versa (e.g., C-G to C-C; type II). The third alternative, noncomplementary bases changing to noncomplementary bases, occurs strictly in loop regions and therefore was not included in these analyses. Each of these classes is composed of changes that require only a single substitution event (e.g., only one of the two bases has been substituted) and of changes that require a pair of substitution events. In RNA, uracil can pair with guanine, making it possible to have a single change from one base-pairing couplet to another (e.g., C-G can change to U-G).

To generate expected values, we used the rate at which random substitutions would be expected to generate paired complementary bases among single substitutions and double substitutions separately (fig. 3). The probability of a double substitution converting one pair of complementary bases to another pair of complementary bases is the number of double type I changes divided by the sum of the number of double type I changes and the number of double type II changes. For RNA molecules the value is $11/(11+32) = 0.256$. The analogous value for single substitutions is $4/(4+28) = 0.125$. Expected values were calculated without regard to the base composition of the stem regions.

To quantify the evolutionary constraint of secondary structure, we tallied all the observed single and double compensatory stem substitutions and compared these with the values expected either if all stem substitutions maintained base pairing (i.e., complete dependence) or if these changes only occur at random (i.e., complete independence). To calculate the appropriate relative weighting of stem bases for phylogenetic analysis, we assumed a linear relationship between the degree of independence and the amount of weighting.

Results

The 1,989 bases aligned by Hillis et al. (1991) represent slightly less than one-half the total 28S rRNA molecule (48.4%, based on *Xenopus*). The numbers of pairing and nonpairing bases are 1,082 (54.4%) and 907 (45.6%), respectively (table 1). These proportions are not significantly different [$\chi^2 = 0.03$, degrees of freedom (df) = 1, $P > 0.8$] from the proportions predicted by the *Xenopus* model of secondary structure for the entire 28S gene [2,244 (54.6%) stem bases and 1,866 (45.4%) loop bases; Clark et al. 1984]. The proportions of variable and informative sites are not significantly different in stem regions versus loop regions (table 1).

Expansion segments (ES) are regions found in eukaryote 28S rRNA that have undergone large insertions and rapid rates of substitutions (Hassouna et al. 1984). The bases sequenced by Hillis et al. (1991) are not representative of the proportion of ES bases (based on *Xenopus;* $\chi^2 = 30.6$, df = 1, $P < 0.05$). This difference results partly from the difficulty of aligning these regions among distantly related taxa and partly because the regions sequenced were the more conservative regions of the 28S rRNA molecule. ES contain significantly more paired bases than does the rest of the gene (based on *Xenopus;* $\chi^2 = 12.8$, df = 1, $P < 0.05$).

The results of our phylogenetic analyses, based on two different models of secondary structure, are highly congruent with each other (tables 2 and 3). Stem bases supported the conventional hypothesis of vertebrate relationships (tree I), with the exception of insertion/deletion characters, which do not contain a significant amount of phylogenetic signal (based on the g_1 statistic). Loop characters support unconventional trees (trees II–VI of fig. 2) that conflict with traditional phylogenetic hypotheses, except that all of the loop character–derived trees support both a monophyletic Mammalia and Sarcopterygia. The conventional tree (tree I), however, was never far removed from the most parsimonious tree in these analyses. The best results, as judged by the level of support of the conventional tree, were obtained when the two data types were combined to generate the largest data set.

How constrained are the stem bases in this data set, as a result of complementary base-pairing? If no constraints exist, then the number of substitutions that maintain or restore complementation to bases should be no greater than would arise by chance alone. Therefore, we calculated the number of single and double changes that would be expected to create or maintain complementation at random (fig. 3, based on *Xen-*

Table 1
Synopsis of Sequence Data from Vertebrate 28S rRNA Genes

	Total No. of Aligned Bases[a]	No. (%) of Variable Nucleotides[b]	No. (%) of Informative Characters[c]
Loops	907	221 (24.4)	56 (6.2)
Stems	1,082	277 (25.6)	55 (5.1)
Total	1,989	498 (25.0)	111 (5.6)

[a] From Hillis et al. (1991).

[b] Number of aligned bases that vary among the examined taxa.

[c] Aligned bases that have derived states occurring in two or more taxa. Secondary structure is based on the *Xenopus* model.

Table 2
Effect of Secondary Structure on Phylogenetic Analysis of rRNA Sequence Data

Character Type and Region	No. of Informative Characters[a]	g_1[b]	Most Parsimonious Tree[c]	Steps to Tree I[d]	Bootstrap Support of Node[e] 1	2	3	4	5
All:									
Total	111	−0.50**	I	...	71	100	65	50	100
Stems	57	−0.49**	I	...	87	93	74	45[f]	83
Loops	54	−0.59**	II	4 (36)	44[f]	97	14[f]	9[f]	96
Substitutions:									
Total	95	−0.51**	I	...	76	100	63	49[f]	96
Stems	51	−0.52**	I	...	89	95	62	52	71
Loops	44	−0.50*	II	2 (22)	34[f]	96	16[f]	11[f]	81
Insertions/deletions:									
Total	16	−0.51**	III	2 (54)	0[f]	97	17[f]	8[f]	65
Stems	6	−0.41	IV	2 (546)	0[f]	93	7[f]	2[f]	2[f]
Loops	10	−1.47**	III	2 (231)	0[f]	57	0[f]	7[f]	87

[a] Partitioned into stem and loop regions on the basis of the *Xenopus* model and by type.
[b] Amount of structure within a data set.
[c] Tree topologies are presented in fig. 2.
[d] Number of steps that separate tree I from the most parsimonious tree; the number in parentheses is the number of trees that are less than or equal in length to tree I.
[e] Support of the conventionally recognized nodes was determined by bootstrap analysis ($N = 100$). 1 = rodents; 2 = mammals; 3 = amniotes; 4 = tetrapods; and 5 = sarcopterygians (see fig. 2).
[f] Node was not present in the majority-rule consensus tree of the bootstrap pseudoreplicates.
* $P \leq 0.05$.
** $P \leq 0.01$.

opus). Table 4 presents both the calculated expected values for these changes and the observed results. There are significantly more compensatory mutations than expected by chance, among both single ($\chi^2 = 47.6$, df = 1, $P < 0.01$) and double ($\chi^2 = 16.7$, df = 1, $P < 0.01$) changes, as previously noted, by Michel and Dujon (1983) and Curtiss and Vournakis (1984), for other rRNA data sets.

Because the stem bases are not evolving independently, we wished to determine an appropriate relative weighting of these characters, for phylogenetic analysis. We generated expected values for hypotheses of complete dependence (all stem substitutions are compensatory) and complete independence (compensatory changes occur only by chance). The expected number of compensatory substitutions was calculated for a hypothesis of complete independence, by summing the predicted number of single compensatory mutations and two times the number of expected double compensatory mutations (table 4). We observed that 43 of 91 substitutions maintained base pairing; this is only 38.2% of the potential compensatory mutations but is considerably more than the 13.3 compensatory substitutions that are expected to arise without selection for compensation.

We transformed the presence of compensatory mutations into a phylogenetic weighting scheme by equating the expected value of "compensatory" mutations that arise at random with the logical end points of relative weighting (fig. 4). That is, if stem characters were independent and if compensatory changes appeared no more often than expected at random, we would weight stem characters equally with loop characters (i.e., 1.0). If stem characters were 100% dependent on their counterparts, all substitutions would maintain complementarity, and they would be weighted 0.5

Table 3
Effect of *Rattus* Model of Secondary Structure on Phylogenetic Analysis of rRNA Sequence Data

Character Type and Region	No. of Informative Characters	g_1	Most Parsimonious Tree	Steps to Tree I	Bootstrap Support of Node				
					1	2	3	4	5
All:									
Total	111	−0.50**	I	...	71	100	65	50	100
Stems	62	−0.51**	I	...	55	100	57	44	92
Loops	49	−0.44**	IV	2 (9)	70	99	29[a]	16[a]	95
Substitutions:									
Total	95	−0.51**	I	...	76	100	63	49[a]	96
Stems	55	−0.53**	I	...	55	95	51	49[a]	76
Loops	40	−0.41**	V	1 (6)	67	94	33[a]	24[a]	86
Insertions/deletions:									
Total	16	−0.51*	III	2 (54)	0[a]	97	17[a]	8[a]	65
Stems	7	−0.43	VI	2 (489)	0[a]	77	25[a]	3[a]	10[a]
Loops	9	−0.85**	Ib	2 (189)	0[a]	75	0[a]	13[a]	94

NOTE.—Definitions are as in table 2.
[a] Node was not present in the majority-rule consensus tree of the bootstrap pseudoreplicates.
* $P \leq 0.05$.
** $P \leq 0.01$.

relative to loop characters. When linear scaling is assumed, interpolating between these values suggests the appropriate relative weighting of 0.81 for these characters, on the basis of the observed level of compensatory changes.

Discussion

The phylogenetic analysis of Hillis et al. (1991) supported the traditionally recognized taxonomic groups (i.e., rodents, mammals, amniotes, and tetrapods) and recognized *Latimeria* as a sarcopterygian. Our analyses indicate that the support for these relationships is a result of the information contained in both stem and loop characters. Stem characters alone supported the established tree, while loop characters alone strongly supported the monophyly of the mammals and sarcopterygians but yielded a shortest tree that was otherwise unconventional. The best results, on the basis of relative bootstrap support of the generally accepted hypothesis of relationships, were obtained by combining all of the data.

The two models of secondary structure that were used represent the two most diverse models available for these data: they differ substantially in some regions of the 28S rRNA molecule. For example, the pairing status of 33 of the 95 informative substitution characters is different in the two models. In spite of these differences, our conclusions are robust and appear to be independent of the specific model of secondary structure used.

Contrary to the conclusions of Wheeler and Honeycutt (1988), we conclude that stem-region nucleotides do contain phylogenetically useful information. It is surprising that loop characters, as defined by either model, were found to support relationships that were at odds with other analyzed data sets when only the shortest tree was considered. This difference may be due to the different rRNA genes that were analyzed

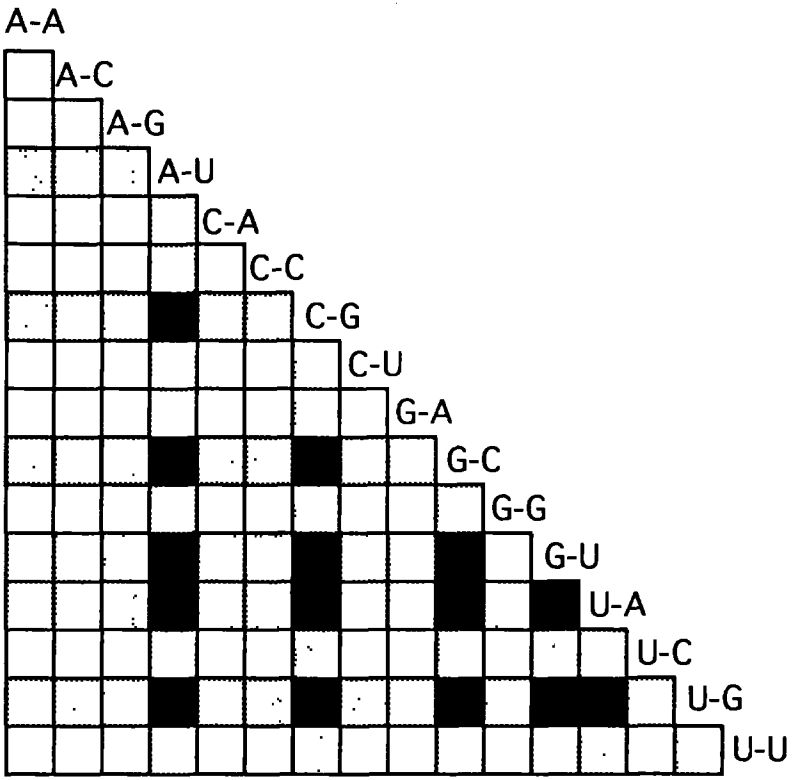

FIG. 3.—All possible changes between pairs of RNA nucleotides. Each square represents the change between the base pair represented at the top of its column and that at the end of its row. ■ = Complementary pair to complementary pair (15 classes = 4 one-step changes + 11 two-step changes); □ = complementary pair to noncomplementary pair, or vice versa (60 classes = 28 one-step changes + 32 two-step changes); and ☐ = noncomplementary pair to noncomplementary pair (45 classes, excluded from this analysis).

in the two studies, to the difference in taxonomic groups and depth of evolutionary time spanned, or to the different numbers of characters in the two data sets.

Smith (1989) arrived at conclusions similar to ours, using a different rRNA gene (18S) and different taxa (echinoderms). It appears that analysis of the larger rRNA genes, 18S (Smith 1989) and 28S (present study), generate results different from those of analyses of the secondary structure of the smaller rRNA genes (5S and 5.8S; Wheeler and Honeycutt 1988). Recent analyses have identified problems in the use of these genes for phylogenetic analyses (Halanych 1991; Steele et al. 1991). The smaller rRNA genes contain fewer nucleotides and hence usually contain fewer informative characters. In splitting their data set into two subgroups, Wheeler and Honeycutt (1988) may not have been left with enough characters to accurately reconstruct relationships. In our analyses, the smallest subsets of our data, insertion/deletion events, yielded unusual trees when analyzed alone and yet, when combined with the substitution data, rendered the expected results. Smith (1989) seems to concur on this point, at least for more recent divergences, by suggesting that using both stem and loop characters increases the chance of finding the correct tree, because of the increase in the number of characters.

We have concluded that stem characters are phylogenetically informative, and

Table 4
Substitutions Observed in Vertebrate 28S rDNA Sequence Data, on the Basis of the *Xenopus* Model of Secondary Structure

Type of Substitution (no. of ways[a])	No. Expected[b]	No. Observed
Single:		
Base pairing to base pairing (4)	9.3	29
Base pairing to non–base pairing (28)	65.6	46
Double:		
Base pairing to base pairing (11)	2	7
Base pairing to non–base pairing (32)	6	1

[a] Number of ways (of 120 possible) in which one can choose two pairs of nucleotides that have this character.

[b] Based on the frequency of complementary pairs expected at random (see fig. 3).

yet we have also empirically determined that compensatory mutations among stem bases do occur, so independence of the characters is reduced. How, then, should base-pairing characters be weighted? Any proposed scheme of weighting requires assumptions. Many practicing systematists prefer to treat all characters equally, accepting that each datum, as they have defined it, is independent and potentially equally informative. If stem bases are not a meaningful source of phylogenetic information, as suggested by Wheeler and Honeycutt (1988), then ignoring them is equivalent to a relative weighting of zero. On the other hand, if stem bases evolve strictly as pairs, then weighting by one-half is appropriate. However, our data show that complementary bases are not inextricably linked. Given that the potential end points of weighting stem bases are 1.0 among bases that are independent and 0.5 for two characters that represent a single evolutionary event, our data suggest an intermediate weighting value of ≈ 0.8.

Is a weighting of 0.8 likely to produce results different from those produced by

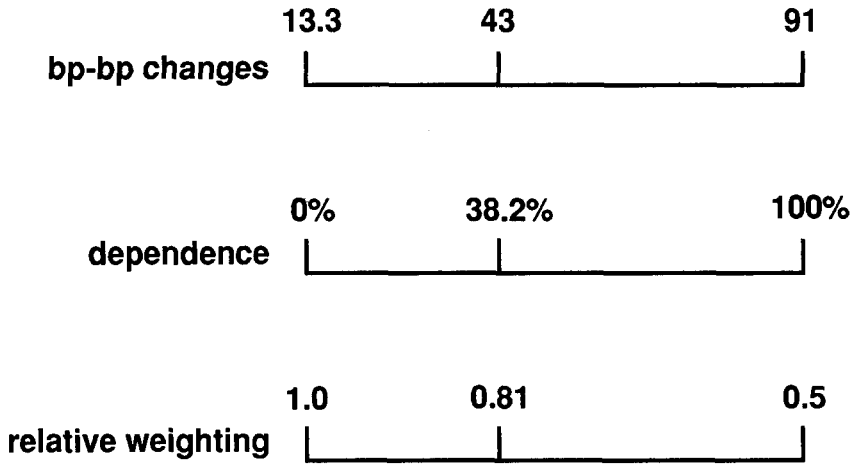

FIG. 4.—*Top,* Number of substitutions that maintain base-pair complementation between pairs of stem bases (based on the *Xenopus* model): observed (43), expected at random (13.3), and expected if all substitutions maintain complementation (91). *Middle,* Percent dependence associated with these values. *Bottom,* Relative weighting appropriate for phylogenetic analysis.

a weighting of 0.5? We reanalyzed the entire 28S data set, weighting stem characters (*Xenopus* model) first at 0.8 and then at 0.5. In the former analysis, tree I (the conventional tree; see fig. 2) was supported; in the latter analysis, tree II (an unconventional loop tree; see fig. 2) was generated. Thus, this difference can be important in phylogenetic inference. However, there was no obvious difference between the results of the analysis with all characters equally weighted and the results of the analysis with stem characters weighted 0.8.

Although the secondary structure of rRNA does reduce the evolutionary independence of the paired nucleotides, a weighting of one-half for paired bases overcompensates for this interdependence. Equal weighting of paired and unpaired nucleotides is actually closer to the weighting (i.e., 0.8) that we conclude is most appropriate for removing the effects of nonindependence due to the constraints of secondary structure. Different data sets might deserve different weightings that could be obtained by the method employed here.

Acknowledgments

We thank Loren Ammerman for assistance in the preliminary phases of this work and John Huelsenbeck for assistance in the use of g_1. Jim Bull, David Cannatella, Paul Chippindale, and Tod Reeder provided useful comments on the manuscript. This work was supported by funding from National Science Foundation grants BSR 86-57640 and BSR 87-96293, both to D.M.H.

LITERATURE CITED

CLARK, C. G., B. W. TAGUE, V. C. WARE, and S. A. GERBI. 1984. *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res. **12**:6197–6220.

CURTISS, W. C., and J. N. VOURNAKIS. 1984. Quantification of base substitutions in eukaryotic 5S rRNA: selection for the maintenance of RNA secondary structure. J. Mol. Evol. **20**:351–361.

FITCH, W. M. 1984. Cladistic and other methods: problems, pitfalls, and potentials. Pp. 221–252 *in* T. DUNCAN and T. F. STUESSY, eds. Cladistics: perspectives on the reconstruction of evolutionary history. Columbia University Press, New York.

FRITZSCH, B. 1987. Inner ear of the coelacanth fish *Latimeria* has tetrapod affinities. Nature **327**:153–154.

HADJIOLOV, A. A., O. I. GEORGIEV, V. V. NOSIKOV, and L. P. YAVACHEV. 1984. Primary and secondary structure of rat 28S ribosomal RNA. Nucleic Acids Res. **12**:3677–3693.

HALANYCH, K. M. 1991. 5S ribosomal RNA sequences inappropriate for phylogenetic reconstruction. Mol. Biol. Evol. **8**:249–253.

HASSOUNA, N., B. MICHOT, and J.-P. BACHELLERIE. 1984 The complete nucleotide sequence of mouse 28S rRNA gene: implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res. **12**:3563–3583.

HILLIS, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. Pp. 278–294 *in* M. M. MIYAMOTO and J. CRACRAFT, eds. Phylogenetic analysis of DNA sequences. Oxford University Press, Oxford and New York.

HILLIS, D. M., and M. T. DIXON. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**:411–453.

HILLIS, D. M., M. T. DIXON, and L. K. AMMERMAN. 1991. The relationships of the coelacanth *Latimeria chalumnae:* evidence from sequences of vertebrate 28S ribosomal RNA genes. Environmental Biol. Fishes **32**:119–130.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. **83**:189–195.

HUELSENBECK, J. P. 1991. Tree-length distribution skewness: an indicator of phylogenetic information. Syst. Zool. **40**:257–270.

MICHEL, F., and B. DUJON. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast-, and nuclear-encoded members. EMBO J. **2**: 33–38.

NOLLER, H. F. 1984. Structure of ribosomal RNA. Annu. Rev. Biochem. **53**:119–162.

NORTHCUTT, R. G. 1987. Lungfish neural characters and their bearing on sarcopterygian phylogeny. J. Morph. Suppl. **1**:277–297.

PUSTELL, J., and F. C. KAFATOS. 1986. A convenient and adaptable microcomputer environment for DNA and protein manipulation and analysis. Nucleic Acids Res. **14**:479–488.

ROMER, A. S. 1966. Vertebrate paleontology, 3d ed. University of Chicago Press, Chicago.

ROSEN, D. E., P. L. FOREY, B. G. GARDINER, and C. PATTERSON. 1981. Lungfishes, tetrapods, paleontology, and plesiomorphy. Bull. Am. Museum. Hist. **167**:159–276.

SCHULTZE, H.-P. 1987. Dipnoans as sarcopterygians. J. Morph. Suppl. **1**:39–74.

SMITH, A. B. 1989. RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics **5**:321–344.

STEELE, K. P., K. E. HOLSINGER, R. K. JANSEN, and D. W. TAYLOR. 1991. Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Mol. Biol. Evol. **8**:240–248.

SWOFFORD, D. L. 1990. PAUP: phylogenetic analysis using parsimony, version 3.0h. Illinois Natural History Survey, Champaign.

WHEELER, W. C., and R. L. HONEYCUTT. 1988. Paired sequence difference in ribosomal RNAs: evolution and phylogenetic implications. Mol. Biol. Evol. **5**:90–96.

WESLEY M. BROWN, reviewing editor

Received July 15, 1991; revision received June 19, 1992

Accepted June 19, 1992