Parallel Functional Changes in the Digestive RNases of Ruminants and Colobines by Divergent Amino Acid Substitutions

Jianzhi Zhang

Department of Ecology and Evolutionary Biology, University of Michigan

A morphological or physiological trait may appear multiple times in evolution. At the molecular level, similar protein functions may emerge independently in different lineages. Whether these parallel functional changes are due to parallel amino acid substitutions has been a subject of debate. Here, I address this question using digestive ribonucleases (RNases) of two groups of foregut-fermenting mammals: ruminant artiodactyls and colobine monkeys. The RNase1 gene was duplicated twice in ancestral ruminants at least 40 MYA, and it was also duplicated in the douc langur, an Asian colobine, approximately 4 MYA. After duplication, similar functional changes occurred in the ruminant and monkey enzymes. Interestingly, five amino acid substitutions in ruminant RNases that are known to affect its catalytic activity against double-stranded (ds) RNA did not occur in the monkey enzyme. Rather, a similar functional change in the monkey was caused by a different set of nine substitutions. Site-directed mutagenesis was used to make three of the five ruminant-specific substitutions in the monkey enzyme. Functional assays of these mutants showed that one of the three substitutions has a similar effect in monkeys, the second has a stronger effect, and the third has an opposite effect. These results suggest that (1) an evolutionary problem can have multiple solutions, (2) the same amino acid substitution may have opposite functional effects in homologous proteins, (3) the stochastic processes of mutation and drift play an important role even at functionally important sites, and (4) protein sequences may diverge even when their functions converge.

Introduction

Parallel and convergent evolution refers to independent acquisitions of the same character state on more than one occasion during evolution. The distinction between parallelism and convergence is that the former refers to the situation in which the ancestral states were identical among independent lineages, whereas the latter requires different ancestral states (Zhang and Kumar 1997). For simplicity, the word ''parallelism'' will be used here for both situations unless otherwise noted. Parallel evolution occurs frequently at the morphological and physiological levels. The best known examples include body shapes of marine mammals and fish and the ability to fly in bats and birds. At the molecular level, parallel evolution has been a confusing concept because it is often unstated what kinds of parallelism are meant (Doolittle 1994). Doolittle (1994) listed many examples in which proteins independently evolved the same type of function, structural fold, or biochemical mechanism. At the protein sequence level, although the presence of parallel evolution has been questioned (Doolittle 1994), about a dozen cases have been reported (e.g., Stewart, Schilling, and Wilson 1987; Yokoyama and Yokoyama 1990; Holmes et al. 1992; Krakauer et al. 1996; Yeager, Kumar, and Hughes 1997; Briscoe 2002). Theses cases are interesting to molecular evolutionists because parallel substitutions are often thought to be results of adaptive evolution. Unfortunately, many of the reported sequence parallelisms were simply observations of unexpected sequence resemblance, and only a few cases tested the functional resemblance as well. Even in the latter cases, it is rarely demonstrated that the functional resemblance is caused by sequence resemblance, a necessity for claims of adaptive parallel evolution at the protein sequence level. Here I address this question by comparative evolutionary analyses and site-directed mutagenesis studies of digestive RNases that independently emerged in the evolution of two groups of leaf-eating mammals.

Ruminant artiodactyls and colobine monkeys mainly eat leaves. A foregut-fermenting alimentary system independently originated in these organisms to facilitate the digestion of the cellulose-rich food (Kay and Davies 1994). That is, symbiotic bacteria in the foregut ferment the leaves, and the mammals then recover nutrients by lysing and digesting the bacteria with various enzymes. One important digestive enzyme is pancreatic RNase, which is secreted from the pancreas and transported into the small intestine to degrade RNAs. Earlier studies revealed a substantially greater amount of RNase in the pancreas of ruminants and colobines than in other mammals (Barnard 1969; Beintema 1990). This is related to the fact that rapidly growing bacteria have the highest ratio of RNA-nitrogen to total nitrogen of all cells, and high concentrations of RNase are needed to break down bacterial RNAs so that nitrogen can be recycled efficiently (Barnard 1969). Pancreatic RNases of many mammals are also expressed outside the pancreas (Futami et al. 1997) and have an activity (EA_{dsRNA}) in degrading double-stranded (ds) RNA (Libonati and Floridi 1969). This activity is unrelated to digestion and is thought to be involved in the host-defense against pathogenic viruses (Sorrentino and Libonati 1997; Libonati and Sorrentino 2001).

While most mammals have only one copy of the pancreatic RNase gene that is also known as RNase1, all ruminants examined have three gene copies that are results of two consecutive gene duplications in an ancestral ruminant (Beintema and Kleineidam 1998; Breukelman et al. 2001). The three genes have different expression patterns and they are known as the pancreatic, seminal, and

Key words: RNase, ruminant, colobine, adaptive evolution, digestion.

E-mail: jianzhi@umich.edu.

The sequences reported in this paper have been deposited in the GenBank database (accession numbers AY261461 to AY261466).

Mol. Biol. Evol. 20(8):1310–1317. 2003
DOI: 10.1093/molbev/msg143
Molecular Biology and Evolution, Vol. 20, No. 8,

brain RNases, respectively (Beintema and Kleineidam 1998). They also differ in RNase activity and protein structure. For example, seminal RNase forms homodimers, whereas the other two RNases are monomers. Pancreatic RNase has a decreased activity in degrading dsRNA, suggesting that its function is specialized in digesting RNAs released from foregut bacteria (Jermann et al. 1995; Libonati and Sorrentino 2001). On the other hand, seminal RNase has an increased activity against dsRNA (Libonati and Sorrentino 2001). Recently, an independent duplication of the RNase1 gene was found in the douc langur (*Pygathrix nemaeus*), an Asian colobine monkey (Zhang, Zhang, and Rosenberg 2002). Interestingly, while the sequence and function of one daughter gene (RNase1A) has remained unchanged, the other gene (RNase1B) has a dramatically reduced activity against dsRNA (Zhang, Zhang, and Rosenberg 2002). However, the digestive function of RNase1B has been enhanced by a reduction in its optimal catalytic pH from 7.4 to 6.3, an adaptive response to the lower pH in the colobine small intestine (6 to 7), in comparison to those in other primates (7.4 to 8) (Zhang, Zhang, and Rosenberg 2002). RNases purified from the pancreas of a related colobine species appear to be RNase1B (Beintema 1990; Zhang, Zhang, and Rosenberg 2002), suggesting that RNase1B has taken the digestive role and has left the putative antiviral function (by EA_{dsRNA}) to its paralog, RNase1A. Therefore, specialization in digestion and reduction in the activity of degrading dsRNA occurred independently in ruminant pancreatic RNase and colobine RNase1B after gene duplication, which constitutes functional parallelism. Molecular mechanisms behind the reductions in EA_{dsRNA} have been investigated in ruminants and colobines separately and the molecular mechanism underlying the shift of optimal pH in colobines is also known to some extent. Here, I perform a comparative evolutionary analysis and conduct experiments derived from this analysis to test the hypothesis of sequence parallelism underlying the functional parallelism of the digestive RNases.

Materials and Methods

Sequencing of the RNase1 Gene from the Dolphin and Whale

To clarify the directions of the amino acid substitutions and functional changes in RNase evolution, it is necessary to know the phylogenetic relationships of the three duplicated RNase1 genes in ruminants. For this purpose, the coding region, a 5′ noncoding region, and a 3′ noncoding region of the RNase1 gene were amplified by polymerase chain reaction (PCR) from the genomic DNAs of the bottlenose dolphin (*Tursiops truncatus*) and bowhead whale (*Balaena mysticetus*), with primers derived from the published sequences of the bovine pancreatic (GenBank accession number X07283), brain (GenBank accession number X59767), and seminal (GenBank accession number AJ000518) RNase genes. The primers used are number 467 (5′-AGTGCGGTCATCATGGCTCTGAAG-3′) and number 468 (5′-TGCTCTGGCCTTAGGTAGAGACCTA-3′) for the coding region, number 469 (5′-CAGTCCACTTTGATGCTTCAGTGTAG-3′) and number 470 (5′-GAGAGGAATTATCTTAGGGCACTGAA-3′) for the 3′ noncoding region, and number 471 (5′-GCTGCTTCTGCTTGAGGGTTTG-3′) and number 472 (5′-GCATCGCCTTCGAAAAGAACC-3′) for the 5′ noncoding region. For the coding region, the N-terminus primer encodes the first four amino acids of the signal peptide, whereas the C-terminus primer is positioned downstream of the coding region. PCRs were performed under conditions recommended by the manufacturer (Invitrogen, San Diego, Calif.). PCR products were cloned into pCR4TOPO cloning vector (Invitrogen) and sequenced from both directions by the dideoxy chain termination method on an automated sequencer.

Evolutionary Analyses

Previously reported DNA sequences of RNase1 genes of artiodactyls and primates were obtained from GenBank. The protein and DNA sequences were aligned using ClustalX (Thompson et al. 1997). Gene trees were reconstructed using the neighbor-joining method (Saitou and Nei 1987) implemented in MEGA2 (Kumar et al. 2001), as well as the likelihood method (Felsenstein 1981) implemented in PAUP* (Swofford 1998). Several different distance measures or substitution models (Jukes-Cantor, Kimura's two-parameter, Tajima-Nei, Tamura-Nei, and general reversible models, with or without the gamma distribution for rate variation among sites [Nei and Kumar 2000]) were used. The interior branch test (Rzhetsky and Nei 1992; Nei and Kumar 2000) and the bootstrap test (Felsenstein 1985) with 1,000 replications were used to examine the reliability of the reconstructed trees.

Site-Directed Mutagenesis, Isolation of Recombinant Protein, and RNase Assay

The douc langur RNase1A gene was subcloned into the pFLAG CTS bacterial expression vector (Kodak, New Haven, Conn.) with a bacterial signal peptide sequence at the N-terminus and an octapeptide DYKDDDDK (FLAG) sequence at the C-terminus (Zhang, Zhang, and Rosenberg 2002). Previous studies showed that the FLAG octapeptide does not interfere with the folding or the catalytic activity of recombinant RNases (Rosenberg and Dyer 1995). The QuikChange site-directed mutagenesis kit of Stratagene (La Jolla, Calif.) was used to generate designed mutations in the gene construct following manufacturer's instructions. The mutations were confirmed by DNA sequencing. Recombinant proteins were isolated from 6 L of bacterial cultures after 4-h induction with isopropyl-1-thio-β-galactoside (IPTG, 10 μM). After harvest and cell lysis by freeze-thaw and sonication, recombinant proteins were concentrated and isolated by M2 anti-FLAG monoclonal antibody affinity chromatography (Sigma). The concentration of the recombinant protein was determined by quantitative Western analysis with a FLAG-conjugated BAP protein at known concentration (Sigma). The enzyme activity (EA_{dsRNA}) of the recombinant RNase against dsRNA [poly(U)•poly(A) combined from poly(U) and poly(A) of Pharmacia (Piscataway, NJ)] was measured by

the initial reaction rate at 25°C in 1 ml buffer of 0.15 M sodium chloride and 0.015 M sodium citrate (pH 7.0) with 40 μg substrate and 10 to 100 pmol RNase and was determined from ultraviolet absorbance at 260 nm (Libonati and Floridi 1969). At least three replications of experiments were conducted, with the means and their standard errors computed. This experimental procedure was identical to that used in Zhang, Zhang, and Rosenberg (2002) and the obtained EA_{dsRNA} values are directly comparable. We further measured the EA_{dsRNA} of the recombinant human pancreatic RNase prepared in our lab and that of the cow pancreatic RNase (Sigma). We found that our results are virtually identical to those reported in the literature (Opitz et al. 1998; Libonati and Sorrentino 2001), suggesting direct comparability between our measures of EA_{dsRNA} and those in the literature. Our previous study showed that EA_{dsRNA} is highest at pH 7.0 (Zhang, Zhang, and Rosenberg 2002). Since EA_{dsRNA} is unrelated to the digestion of bacterial RNA and is probably biologically relevant in tissues other than the small intestine, we here measured and reported EA_{dsRNA} at pH 7.0 for all subjects.

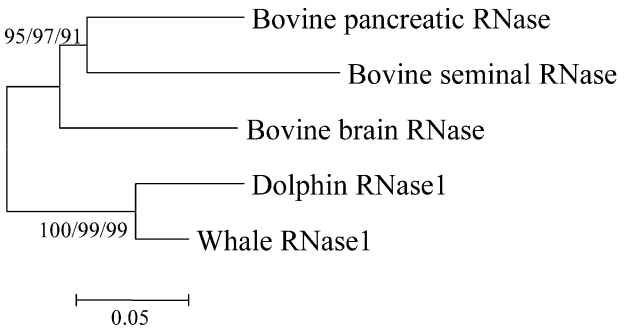

FIG. 1.—Phylogenetic relationships of ruminant pancreatic, seminal, and brain RNases. The whale and dolphin RNases are used as outgroups. The neighbor-joining (NJ) method with Kimura's (1980) two-parameter distances is used. Use of other distance measures or use of the maximum-likelihood (ML) method results in the same tree. The percentage reliability values by the NJ-bootstrap, NJ–interior-branch, and ML-bootstrap tests are shown on tree branches. For ML analysis, the general reversible model with discrete gamma distribution of among-site rate variation is assumed. A total of 1,376 nucleotide sites are used, including the 5′ noncoding (591 sites), coding (438 sites), and 3′ noncoding (347 sites) regions. An identical topology is obtained when each of the three regions is considered separately.

Results

Phylogenetic Relationships of Pancreatic, Seminal, and Brain RNases of Ruminants

As mentioned, the RNase1 gene was duplicated twice in an ancestral ruminant so that all ruminants have three copies of the gene. To understand how amino acid substitutions and functional changes occurred in these duplicated genes, it is necessary to know the phylogenetic relationships of the three genes. The closest living relatives of ruminants are hippopotamus and cetaceans (Nikaido, Rooney, and Okada 1999; Murphy et al. 2001), which appear to have only one RNase1 gene (Kleineidam et al. 1999). A phylogenetic analysis using 348 nucleotides of the RNase coding region suggested that the pancreatic and seminal RNase genes form a sister group to the exclusion of the brain RNase gene (Breukelman et al. 2001). However, this grouping only has approximately 50% bootstrap support (Breukelman et al. 2001) and thus is ambiguous. To resolve the trifurcation of the three RNase1 genes of ruminants, I sequenced the coding as well as 5′ and 3′ noncoding regions of the RNase1 genes from the bottlenose dolphin and bowhead whale. The corresponding regions of the three RNase1 genes from the cow are available in GenBank. A neighbor-joining tree was reconstructed using the concatenated sequences of the three regions totaling 1,376 nucleotide sites after the removal of alignment gaps. The tree shows that the three RNase genes of the cow form a monophyletic clade with a bootstrap value of 100% (fig. 1). Within this clade, the pancreatic and seminal RNases form a sister group (fig. 1). The bootstrap value for this grouping is now 95%, and the interior branch test gives 97% support. The same grouping is also supported by a 91% bootstrap value in a maximum-likelihood tree analysis (fig. 1). Several different distance measures or substitution models were used, and all gave the same tree topology. I also used the 5′ noncoding, coding, and 3′ noncoding regions separately in tree building, and all three regions support the same topology as shown in figure 1. Altogether, these results strongly suggest that the first duplication of the ruminant RNase1 gene generated the brain RNase gene and the common ancestor of the pancreatic and seminal genes, and the second duplication produced the latter two genes.

Important Sites for Functional Changes in Ruminant and Colobine RNases

With the phylogeny of the ruminant pancreatic, seminal, and brain RNases, one can infer the directions of amino acid substitutions and the directions of functional changes in the evolution of the RNases (fig. 2*A*). The available data on EA_{dsRNA} of extant species suggest a reduction of EA_{dsRNA} in pancreatic RNase (Libonati and Sorrentino 2001). By functional assays of reconstructed ancestral proteins, Jermann et al. (1995) confirmed this prediction and they further showed an increase in EA_{dsRNA} in seminal RNase. Opitz et al. (1998) subsequently showed that five amino acid differences can explain the 25-fold difference in EA_{dsRNA} between the pancreatic and seminal RNases. These five substitutions are now mapped on the tree of figure 2*A*. Among the five substitutions, K31C (K changes to C at position 31 of the mature peptide [see fig. 3]) and R32C are responsible for the formation of homodimers in seminal RNase (Ciglic et al. 1998), which subsequently doubled EA_{dsRNA} (Ciglic et al. 1998). The other three substitutions (Q28L, G38D, and E111G) individually produced a threefold to sixfold difference in EA_{dsRNA} (Opitz et al. 1998). Thus, substitutions L28Q, C31K, C32R, G111E, and G38D will reduce EA_{dsRNA} in bovine RNases.

In the douc langur, there are nine amino acid differences between RNase1A and RNase1B in the mature peptide region (Zhang, Zhang, and Rosenberg 2002). These substitutions are now mapped in the tree of figure 2*B*. It is seen that all substitutions occurred in the

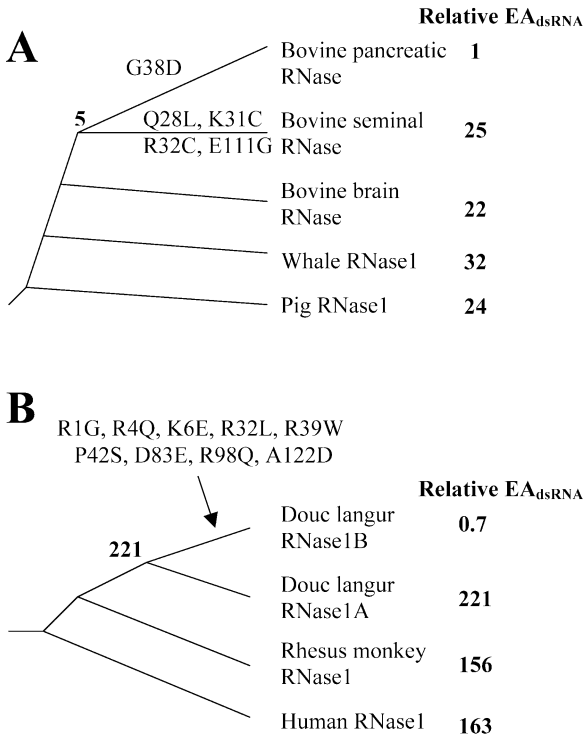

FIG. 2.—Functionally important amino acid substitutions in the changes of the enzyme activity against dsRNA (EA_{dsRNA}) during the evolution of (*A*) artiodactyl and (*B*) primate RNase1 proteins. These sites were previously identified (Jermann et al. 1995; Opitz et al. 1998; Zhang, Zhang, and Rosenberg 2002), and the substitutions are now mapped on the tree. The relative EA_{dsRNA} of bovine seminal RNase was from Opitz et al. (1998), and those of other artiodactyl RNases were from Libonati and Sorrentino (2001). The EA_{dsRNA} values of primate RNases (Zhang, Zhang, and Rosenberg 2002) were scaled according to the result in table 1. The relative EA_{dsRNA} of the common ancestor of bovine pancreatic and seminal RNases was from Jermann et al. (1995) and that of the common ancestor of RNase 1A and RNase 11B was from Zhang, Zhang, and Rosenberg (2002), and they are shown on ancestral nodes. Substitutions XyZ along a branch indicates that amino acid X is replaced by Z at position y of the mature peptide of RNase.

RNase1B branch. In fact, there were no amino acid substitutions in the RNase1A branch, and the protein sequence of the progenitor of RNase1A and RNase1B is identical to that of RNase1A (Zhang, Zhang, and Rosenberg 2002). This suggests that the EA_{dsRNA} of the progenitor was the same as that of RNase1A (fig. 2*B*). Site-directed mutagenesis analyses showed that all nine substitutions in RNase1B reduce EA_{dsRNA} (Zhang, Zhang, and Rosenberg 2002).

I aligned the primate and artiodactyl RNase1 sequences to find whether the five important sites in ruminant proteins that affect EA_{dsRNA} and the nine important sites in colobine proteins overlap. The alignment (fig. 3) showed that there is only one overlapping site between the two sets of sites. However, this overlapping site showed a different substitution in monkeys (R32L) than that (R32C) in ruminants. If I consider only those amino acid changes that reduced EA_{dsRNA}, G38D that occurred in bovine pancreatic RNase did not occur in langur RNase1B; G38 is invariant in primates (Zhang, Zhang, and Rosenberg 2002). The nine amino acid changes in langur RNase1B did not occur in bovine pancreatic RNase either. Furthermore, even when I consider all the amino acid substitutions that occurred in the exterior branches leading to bovine pancreatic RNase and langur RNase1B, none are parallel. Therefore, the parallel functional changes in EA_{dsRNA} observed from colobine and ruminant RNases was due to divergent, rather than parallel, amino acid substitutions. At this time, it is unknown which amino acid substitutions are responsible for the change in catalytic activity against bacterial RNAs in ruminants, and a direct comparison with the langur RNase1B cannot be made. Nevertheless, the fact that there were no parallel amino acid substitutions in bovine pancreatic RNase and langur RNase1B indicates that this functional change likewise was not due to parallel amino acid substitutions.

The occurrence of functional parallelism by different sets of substitutions in two groups of organisms suggests the possibility that the number of sites (N) that can potentially affect the function (i.e., EA_{dsRNA}) is large. With the observation of one overlapping site (although with different substitutions) from two sets of independent substitutions, one may estimate N. The likelihood of observing n overlapping sites is

$$L(n\,|\,N,n_1,n_2)=\frac{\binom{n_1}{n}\binom{N-n_1}{n_2-n}}{\binom{N}{n_2}}$$
$$=\frac{\frac{n_1!}{n!(n_1-n)!}\frac{(N-n_1)!}{(n_2-n)!(N-n_1-n_2+n)!}}{\frac{N!}{n_2!(N-n_2)!}}$$
$$=\frac{n_1!n_2!(N-n_1)!(N-n_2)!}{N!n!(n_1-n)!(n_2-n)!(N-n_1-n_2+n)!} \qquad (1)$$

Here n_1 and n_2 are the numbers of functionally important sites involved in the two sets of substitutions, respectively, and $\binom{x}{y}$ represents the number of ways of partitioning x number of distinct objects into two distinct groups containing y and x-y objects, respectively. In the present case, $n = 1$, $n_1 = 5$, and $n_2 = 9$. L reaches the maximum value of 0.434 when $N = 44$ or 45 (fig. 4). That is, the maximum-likelihood estimate of N is 44 or 45. Because

```
                                  20                  40                  60                  80                 100                 120
Bovine pancreatic RNase  KETAAAKFER QHMDSSTSAA SSSNYCNQMM KSRNLTKDRC KPVNTFVHES LADVQAVCSQ KNVACKNGQT NCYQSYSTMS ITDCRETGSS KYPNCAYKTT QANKHIIVAC KGNPYVPVHF DASV
Bovine seminal RNase     ..S...... .....GN.PS .......L.. CC.KM.QGK. .......... ....K..... .K.T..... .....K...R .......... .......... .VE..... G.K.S..... ....
Bovine brain RNase       ..S.....R. .....GS.SS .NP....-.. .R.RM.HG.. .......... .D..K..... ..IT....HP .....K.... .....R...R .......... .........S .KQ.Y.T... -.......... .GA.LLPATP VPSLPPPHRL L
Bowhead whale RNase1     R.SP.M..Q. .....GN.PG NNR....-.. MR.QM.QG.. .......... .E..K..... ...L....RP ...K.N.A.R .....Q.... .......N.S .VE..... -..D..... .G..

Douc langur RNase1B      G.SQ.E..Q. .....GS.PS ...T..... .L..M.QGW. .S.......P .V...N..F. EK.T..... ..FK.N.K.H ..E..L.NG. .......Q.S PKER..... ..S..... .D..EDST
Douc langur RNase1A      R.SR.K..Q. .....GS.PS ...T..... .R..M.QG.. -.......P .V...N..F. EK.T..... ..FK.N.K.H ..-..L.NG. .......R.S PKER..... ..S..... .-..EDST
Rhesus monkey RNase1     R.SR.K..Q. .....GS.PS .N.T..... .R.SM.HG.. -.......P .V...N..F. EK.T..... ..FK.K.S.H ..-..L.NG. R......R.S PKER..... ..S.H..... .-..EDST
```

FIG. 3.—Alignment of some artiodactyl and primate RNase1 sequences in the complete mature peptide region, with sites known to be important to EA_{dsRNA} (fig. 2) underlined. The last 17 amino acids at the C-terminus of bovine brain RNase is not alignable with the last four amino acids of the primate RNase1, as evident from the sequences (Zhao et al. 2001). Dots in the sequences indicate identical amino acids to the bovine pancreatic RNase.

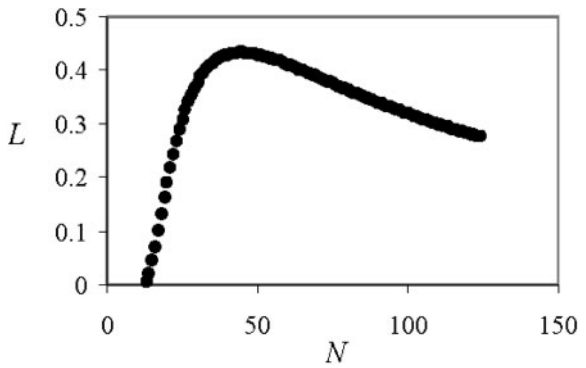

FIG. 4.—Likelihood estimation of the potential number of amino acid sites that can affect EA_{dsRNA}. Shown here is the likelihood $L(n \mid N, n_1, n_2)$ of the observation of $n = 1$ overlapping site from two independent sets of $n_1 = 5$ and $n_2 = 9$ amino acid changes, as a function of the total number of potentially important sites (N). See text and equation (1) for details.

$N \geq n_1 + n_2 - n = 13$, the smallest possible N is 13. The largest possible N equals the sequence length, which is 124 aligned sites in the present case (fig. 3). The relationship between L and N is given in figure 4. Using a standard likelihood ratio test (Edwards 1972), one can show that any null hypothesis with $N < 16$ is rejected at the 5% significance level. Thus, the present data suggest that the number of sites affecting EA_{dsRNA} is most likely to be 44 or 45 and that the number of sites is unlikely to be smaller than 16. Note that the above formulation and estimation depends on the assumption that substitutions at different sites are independent to each each. This assumption probably is not entirely correct in reality (Zhang and Rosenberg 2002). Depending on the form and strength of site–site interaction, N may be overestimated or underestimated. Nevertheless, this computation illustrates that functional parallelism can occur by divergent amino acid substitutions simply because many sites can affect the same function. In such cases, the stochastic nature of mutation and drift plays an important role in determining the actual substitutions that took place in evolution. Nevertheless, it is also possible that in homologous proteins, identical substitutions show different functional effects such that parallel substitutions are not observed. To test this hypothesis, I examined the functional effects of the ruminant-specific substitutions in monkeys by site-directed mutagenesis, as described below.

Functional Effects of Parallel Substitutions in Homologous Proteins

Using reconstructed ancestral proteins, Jermann et al. (1995) showed that substitution G38D that occurred in the evolution of ruminant pancreatic RNase reduced EA_{dsRNA} by approximately fivefold. This substitution did not occur in the parallel reduction of EA_{dsRNA} in douc langur RNase1B. To examine whether G38D would have a similar functional effect if it had occurred in langur RNase, I made a G38D mutation in RNase1A, which is identical in sequence to the common ancestor of RNase1A and 1B (fig. 2*B*). Functional assay shows that this mutation reduced EA_{dsRNA} by about 17-fold (DC-1 in table 1). Thus, G38D would have had a greater functional effect if it had occurred in the evolution of langur RNase.

Table 1
Relative Ribonuclease Activities Against dsRNA (EA_{dsRNA})

RNases	Relative Activity	Effect in Langur (fold)	Effect in Cow (fold)
Bovine pancreatic RNase	1		
Douc langur RNase1A	221 ± 1.2		
DC-1 (G38D)	12.7 ± 0.8	÷ 17	÷ 5
DC-2 (Q28L)	77.0 ± 1.8	÷ 3	× 4
DC-3 (E111G)	595 ± 10	× 3	× 4

NOTE.—DC-1, DC-2, and DC-3 are mutated proteins from douc langur RNase1A, which has the same sequence as the common ancestor of RNase1A and RNase1B. Effects of the corresponding substitutions in bovine RNases were from Jerman et al. (1995) and Opitz et al. (1998).

There are four substitutions (K31C, R32C, Q28L, and E111G) in the evolution of ruminant seminal RNase that increased EA_{dsRNA} (fig. 2*A*). One may test whether these substitutions would affect EA_{dsRNA} in a similar way when they occur in langur RNase. As mentioned, substitutions at positions 31 and 32 caused the formation of homodimers of seminal RNases, which further rendered EA_{dsRNA} high. Because the homodimer structure has other functional consequences in ruminant seminal RNase (D'Alessio et al. 2001), and this structure does not exist in monkey RNases, it is pointless to examine their effects in langur. I therefore examined the effect of Q28L and E111G substitutions in langur RNase. Substitution E111G increased EA_{dsRNA} by 2.7-fold (DC-3 in table 1), similar to its effect (3.6-fold) in ruminant seminal RNase (Opitz et al. 1998). Substitution Q28L, however, showed an unexpected result. Instead of increasing EA_{dsRNA} by approximately fourfold in ruminant seminal RNase, it decreased the catalytic activity by threefold in langur RNase (DC-2 in table 1). Thus, of the three ruminant-specific substitutions, the first (E111G) would have a similar functional effect, the second (G38D) would have a stronger effect, and the third (Q28L) would have an opposite effect, should they occur in the langur RNase.

Discussion

The parallelism of foregut fermentation in ruminants and colobines is a fascinating physiological phenomenon worth detailed molecular genetic and evolutionary analyses. In this work, I compared amino acid substitutions responsible for the similar changes in EA_{dsRNA} that occurred independently in the evolution of the digestive RNases of ruminants and colobines. I found that the similar functional changes were due to different amino acid substitutions in these organisms and that parallel evolution in function was caused by divergent sequence changes. I further showed that this observation may have two reasons. First, the number of sites that can potentially affect EA_{dsRNA} is large. A maximum-likelihood estimate of this number is 44 to 45, although the estimate is associated with a large stochastic error. This implies that an evolutionary problem may have multiple solutions and

the stochastic process of mutation and drift can play an important role even at functionally important sites (Zhang and Rosenberg 2002). Second, identical substitutions can have different functional consequences in homologous proteins due to changes in the genetic background, thus they will not occur when parallel functional changes take place. The present results demonstrate that protein sequences may diverge even when their functions converge. The same has been observed in benzimidazole resistant plants, fungi, and animals (Hughes 1999).

Because it is the parallel reduction in EA_{dsRNA} that is studied here, one may think that the case simply represents functional deterioration, which of course can be due to many different amino acid substitutions. This view is too simplistic. While having reduced EA_{dsRNA}, the digestive RNases of ruminants and colobines have become more specialized in digesting bacterial RNAs. The digestive activity is shown to have increased in colobines (Zhang, Zhang, and Rosenberg 2002), which is probably true in ruminants as well (Libonati and Sorrentino 2001). Therefore, although the reduction in EA_{dsRNA} itself is unlikely an adaptation, the evolution of the RNases is under specific functional constraints and natural selection. It is interesting to see that under parallel pressures for enhanced digestive function, ruminant and colobine RNases independently lost EA_{dsRNA} by different means. More surprisingly, EA_{dsRNA} was enhanced in seminal RNase by substitutions at a different set of sites.

The molecular basis of the RNase activity against dsRNA has been studied quite extensively in bovine seminal and pancreatic RNases, which show 25-fold difference in this activity. Libonati and Sorrentino (2001) proposed that positively charged amino acids (arginines and lysines) are critical to the activity, whereas Opitz et al. (1998) contested their hypothesis. It is worth noting that the difference in EA_{dsRNA} between RNase1A and RNase1B of the douc langur is much greater than that between bovine pancreatic and seminal RNases (fig. 2). Of the nine functionally important substitutions in langur RNase1B, seven involved charge changes (fig. 2*B*). This observation appears in support of Libonati's hypothesis. However, there are two substitutions (P42S and D83E) that do not alter the net charge of the protein, yet are among the most influential ones in changing EA_{dsRNA} (Zhang, Zhang, and Rosenberg 2002). Therefore, Libonati's theory on charged amino acids is valuable, but additional factors certainly exist.

It is interesting to note here that another protein, lysozyme, has been suggested to undergo parallel evolution in ruminants and colobines, from a bactericidal defense protein to a digestive stomach protein that lyses symbiotic bacteria (Stewart, Schilling, and Wilson 1987). Parallel amino acid substitutions have been identified (Stewart, Schilling, and Wilson 1987; Zhang and Kumar 1997), although they have not been shown to be important to the proposed new role of lysozyme in digestion. Furthermore, as pointed out by Kay and Davies (1994), stomach lysozyme has virtually no catalytic activity in the acidic stomach environment (pH 2 to pH 3) (Dobson, Prager, and Wilson 1984), and its presence in stomach is enigmatic.

Parallel evolution at the protein sequence level has been an interesting subject to many molecular evolutionists because it is often presumed that observed parallel substitutions are functionally important and are under adaptive selection. Rigorously speaking, four requirements appear necessary to establish a case of sequence parallelism that would indicate functional importance of and adaptive selection on the parallel amino acid changes. First, parallel amino acid substitutions are observed in independent lineages. Second, proteins under investigation have independently evolved similar functions. Third, the parallel substitutions are indeed responsible for the parallel functional changes. Fourth, the number of observed parallel substitutions is greater than expected by chance alone. Most claims of sequence parallelism satisfy the first requirement or the first and second requirement. To fulfill the third requirement, experiments involving site-directed mutagenesis and recombinant protein techniques are usually needed. The final necessity is a statistical test demonstrating that the parallel substitutions are not attributable to chance alone. Such a test has been developed by Zhang and Kumar (1997). To my knowledge, there is only one case of sequence parallelism so far discovered that satisfies all four requirements, and it is the red and green opsins of vertebrates. Opsin is the protein component of visual pigments that determine the wavelength of light that turns on photoreceptors. Three amino acid sites experienced parallel substitutions in fish and primate opsins, which have similar changes in maximum wavelength absorption (Yokoyama and Yokoyama 1990). Experimental studies confirmed the functional roles of these three substitutions (Asenjo, Rim, and Oprian 1994) and a statistical test by the method of Zhang and Kumar (1997) rejects the hypothesis that these parallel substitutions were due to chance alone (Zhang, unpublished data). In fact, these parallel substitutions were later found to have occurred on more than two occasions, all being correlated with changes in maximum wavelength absorptions (Boissinot et al. 1998; Yokoyama and Radlwimmer 2001), strengthening the claim that they were under adaptive selection. One can imagine that the number of sites (N in equation 1) that potentially affect the wavelength absorption in red/green opsins is small. Indeed, it has been found that functional variations of red/green opsins of most vertebrates can be explained by amino acid substitutions at five sites (Yokoyama and Radlwimmer 2001). In the future, it would be interesting to reexamine the previously reported cases of sequence parallelism using the above four rules and reevaluate the frequency of adaptive parallel evolution at the protein sequence level.

Acknowledgments

The whale and dolphin genomic DNAs were kindly provided by Alex Rooney. I thank Jaap Beintema, Alex Rooney, Mark Springer, David Webb, and two anonymous reviewers for their valuable comments on an earlier version of the manuscript. This work was supported by a start-up fund and a Rackham grant from the University of Michigan and an NIH grant (R01GM067030) to J.Z.

Literature Cited

Asenjo, A. B., J. Rim, and D. D. Oprian. 1994. Molecular determinants of human red/green color discrimination. Neuron **12**:1131–1138.

Barnard, E. A. 1969. Biological function of pancreatic ribonuclease. Nature **221**:340–344.

Beintema, J. J. 1990. The primary structure of langur (*Presbytis entellus*) pancreatic ribonuclease: adaptive features in digestive enzymes in mammals. Mol. Biol. Evol. **7**:470–477.

Beintema, J. J., and R. G. Kleineidam. 1998. The ribonuclease A superfamily: general discussion. Cell Mol. Life Sci. **54**:825–832.

Boissinot, S., Y. Tan, S. K. Shyue, H. Schneider, I. Sampaio, K. Neiswanger, D. Hewett-Emmett, and W.-H. Li. 1998. Origins and antiquity of X-linked triallelic color vision systems in New World monkeys. Proc. Natl. Acad. Sci. USA **95**:13749–13754.

Breukelman, H. J., P. A. Jekel, J. Y. Dubois, P. P. Mulder, H. W. Warmels, and J. J. Beintema. 2001. Secretory ribonucleases in the primitive ruminant chevrotain (*Tragulus javanicus*). Eur. J. Biochem. **268**:3890–3897.

Briscoe, A. D. 2002. Homology modeling suggests a functional role for parallel amino acid substitutions between bee and butterfly red- and green-sensitive opsins. Mol. Biol. Evol. **19**:983–986.

Ciglic, M. I., P. J. Jackson, S. A. Raillard, M. Haugg, T. M. Jermann, J. G. Opitz, N. Trabesinger-Ruf, and S. A. Benner. 1998. Origin of dimeric structure in the ribonuclease superfamily. Biochemistry **37**:4008–4022.

D'Alessio, G., A. Di Donato, R. Piccoli, and N. Russo. 2001. Seminal ribonuclease: preparation of natural and recombinant enzyme, quaternary isoforms, isoenzymes, monomeric forms; assay for selective cytotoxicity of the enzyme. Methods Enzymol. **341**:248–263.

Dobson, D. E., E. M. Prager, and A. C. Wilson. 1984. Stomach lysozymes of ruminants. I. Distribution and catalytic properties. J. Biol. Chem. **259**:11607–11616.

Doolittle, R. F. 1994. Convergent evolution: the need to be explicit. Trends Biochem. Sci. **19**:15–18.

Edwards, A. W. F. 1972. Likelihood. Cambridge University Press, New York.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

Futami, J., Y. Tsushima, Y. Murato, H. Tada, J. Sasaki, M. Seno, and H. Yamada. 1997. Tissue-specific expression of pancreatic-type RNases and RNase inhibitor in humans. DNA Cell Biol. **16**:413–419.

Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Brown. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89**:4835–4839.

Hughes, A. L. 1999. Adaptive evolution of genes and genomes. Oxford University Press, New York.

Jermann, T. M., J. G. Opitz, J. Stackhouse, and S. A. Benner. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature **374**:57–59.

Kay, R. N. B., and A. G. Davies. 1994. Digestive physiology. Pp. 229–250 *in* A. G. Davies and J. F. Oates, eds. Colobinem monkeys: their ecology, behaviour and evolution. Cambridge University Press, Cambridge, UK.

Krakauer, D. C., M. Pagel, T. R. Southwood, and P. M. Zanotto. 1996. Phylogenesis of prion protein. Nature **380**:675.

Libonati, M., and A. Floridi. 1969. Breakdown of double-stranded RNA by bull semen ribonuclease. Eur. J. Biochem. **8**:81–87.

Libonati, M., and S. Sorrentino. 2001. Degradation of double-stranded RNA by mammalian pancreatic-type ribonucleases. Methods Enzymol. **341**:234–248.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

Kleineidam, R. G., G. Pesole, H. J. Breukelman, J. J. Beintema, and R. A. Kastelein. 1999. Inclusion of cetaceans within the order Artiodactyla based on phylogenetic analysis of pancreatic ribonuclease genes. J. Mol. Evol. **48**:360–368.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

Murphy, W. J., E. Eizirik, S. J. O'Brien et al. (11 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science **294**:2348–2351.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Nikaido, M., A. P. Rooney, and N. Okada. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA **96**:10261–10266.

Opitz, J. G., M. I. Ciglic, M. Haugg, K. Trautwein-Fritz, S. A. Raillard, T. M. Jermann, and S. A. Benner. 1998. Origin of the catalytic activity of bovine seminal ribonuclease against double-stranded RNA. Biochemistry **37**:4023–4033.

Rosenberg, H. F., and K. D. Dyer. 1995. Eosinophil cationic protein and eosinophil-derived neurotoxin. Evolution of novel function in a primate ribonuclease gene family. J. Biol. Chem. **270**:21539–21544.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**:945–967.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Sorrentino, S., and M. Libonati. 1997. Structure-function relationships in human ribonucleases: main distinctive features of the major RNase types. FEBS Lett. **404**:1–5.

Stewart, C. B., J. W. Schilling, and A. C. Wilson. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature **330**:401–404.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Mass.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.

Yeager, M., S. Kumar, and A. L. Hughes. 1997. Sequence convergence in the peptide-binding region of primate and rodent MHC class Ib molecules. Mol. Biol. Evol. **14**:1035–1041.

Yokoyama, S., and F. B. Radlwimmer. 2001. The molecular genetics and evolution of red and green color vision in vertebrates. Genetics **158**:1697–1710.

Yokoyama, R., and S. Yokoyama. 1990. Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human. Proc. Natl. Acad. Sci. USA **87**:9315–9318.

Zhang, J., and S. Kumar. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. Mol. Biol. Evol. **14**:527–536.

Zhang, J., and H. F. Rosenberg. 2002. Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. Proc. Natl. Acad. Sci. USA **99**:5486–5491.

Zhang, J., Y. P. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. **30**:411–415.

Zhao, W., E. Confalone, H. J. Breukelman, M. P. Sasso, P. A. Jekel, E. Hodge, A. Furia, and J. J. Beintema. 2001. Ruminant brain ribonucleases: expression and evolution. Biochim. Biophys. Acta **1547**:95–103.

Mark Springer, Associate Editor

Accepted April 1, 2003