Adaptive Evolution of Digestive RNASE1 Genes in Leaf-Eating Monkeys Revisited: New Insights from Ten Additional Colobines

Li Yu,*†,[1] Xiao-yan Wang,[1] Wei Jin,[1] Peng-tao Luan,[1] Nelson Ting,[2] and Ya-ping Zhang†,[3]

[1]Laboratory for Conservation and Utilization of Bio-resource and Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming, China

[2]Anthropology Program, City University of New York Graduate Center

[3]State Key Laboratory of Genetic Resources and Evolution and Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Kunming, China

***Corresponding author:** E-mail: yuli1220@yahoo.com.cn.

†These authors contributed equally to this work.

Associate editor: Prof. Takashi Gojobori

Research article

Abstract

Pancreatic RNase genes implicated in the adaptation of the colobine monkeys to leaf eating have long intrigued evolutionary biologists since the identification of a duplicated *RNASE1* gene with enhanced digestive efficiencies in *Pygathrix nemaeus*. The recent emergence of two contrasting hypotheses, that is, independent duplication and one-duplication event hypotheses, make it into focus again. Current understanding of Colobine *RNASE1* gene evolution of colobine monkeys largely depends on the analyses of few colobine species. The present study with more intensive taxonomic and character sampling not only provides a clearer picture of Colobine *RNASE1* gene evolution but also allows to have a more thorough understanding about the molecular basis underlying the adaptation of Colobinae to the unique leaf-feeding lifestyle. The present broader and detailed phylogenetic analyses yielded two important findings: 1) All trees based on the analyses of coding, noncoding, and both regions provided consistent evidence, indicating *RNASE1* duplication occurred after Asian and African colobines speciation, that is, independent duplication hypothesis; 2) No obvious evidence of gene conversion in *RNASE1* gene was found, favoring independent evolution of Colobine *RNASE1* gene duplicates. The conclusion drawn from previous studies that gene conversion has played a significant role in the evolution of Colobine *RNASE1* was not supported. Our selective constraint analyses also provided interesting insights, with significant evidence of positive selection detected on ancestor lineages leading to duplicated gene copies. The identification of a handful of new adaptive sites and amino acid changes that have not been characterized previously also provide a necessary foundation for further experimental investigations of *RNASE1* functional evolution in Colobinae.

Key words: RNASE1, colobines, adaptive evolution, gene conversion, independent duplication.

Introduction

Gene duplication has long been thought as the primary source for evolutionary innovations and functional adaptations specific to certain groups of organisms (Ohno 1970; Clegg et al. 1997; Force et al. 1999; Zhang 2003; Rispe et al. 2008). One of the most compelling examples comes from the duplication of pancreatic ribonucleases (*RNASE1*) gene in leaf-eating colobine monkeys (Zhang et al. 2002; Zhang 2006). Using both molecular analyses and functional assays, Zhang et al. have revealed that the duplicated *RNASE1* genes in two colobines, Asian *Pygathrix nemaeus* and African *Colobus guereza*, evolved rapidly under positive selection for enhanced digestive efficiencies, as an adaptive response to the increased demands for the enzyme for digesting bacterial RNA (Zhang et al. 2002; Zhang 2006). Furthermore, phylogenetic analysis and molecular dating of the 2-kb *RNASE1* gene sequences, including coding and flanking noncoding regions, showed that the duplication postdated the separation of Asian and African colobines, that is, independent duplication hypothesis. Duplicated *RNASE1* gene in Asian and African colobines are the result of independent duplications but have been subject to the same selective pressures and underwent similar functional changes (Zhang et al. 2002; Zhang 2006; hypothesis A in fig. 1).

Recently, however, Schienman et al. (2006) and Xu et al. (2009) suggested that the coding and noncoding regions portray entirely different evolutionary scenario of RNases in the leaf-eating monkeys. Their analyses based on five colobines, including three additional species, Asian *Semnopithecus entellus*, *Nasalis larvatus*, and African *Piliocolobus badius*, showed that noncoding phylogeny divided *RNASE1* genes into Asian- and African-specific clades, in which the duplicates within each of the five species are more closely related to each other than to their orthologues, a pattern consistent with the independent duplications in Asian and African colobines; whereas coding phylogeny clustered the duplicated *RNASE1* genes of all leaf monkeys together, suggesting an ancient duplication event preceding the

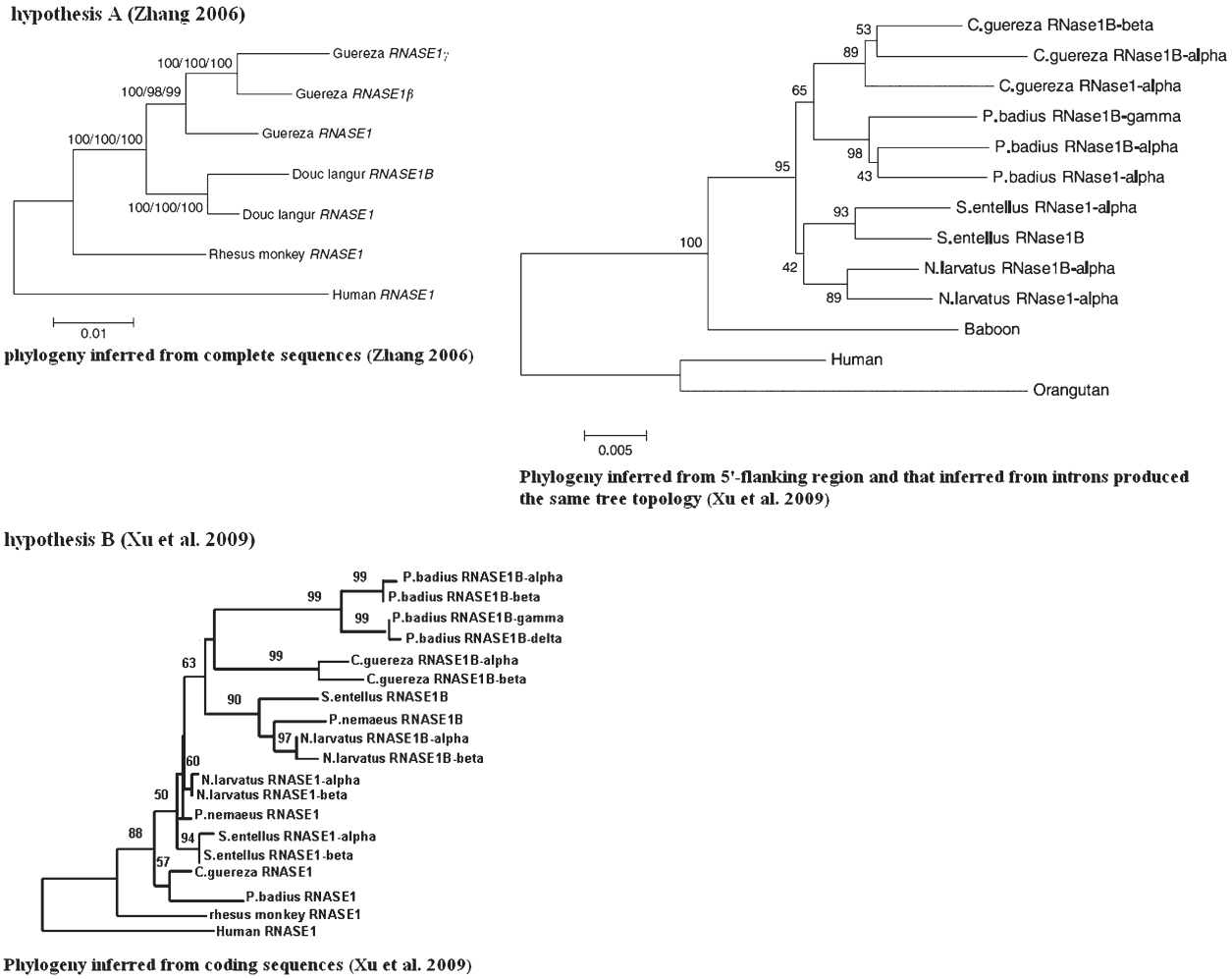

FIG. 1. Current evolutionary hypotheses for the Colobine *RNASE1* gene sequences. Hypothesis A (Zhang 2006), that is, independent duplication events in Asian and African colobines, was supported by the phylogenies inferred from complete and noncoding sequences (Schienman et al. 2006; Zhang 2006; Xu et al. 2009). Hypothesis B (Xu et al. 2009), that is, one duplication events, was supported by the phylogeny inferred from coding sequences (Schienman et al. 2006; Xu et al. 2009).

divergence between Asian and African colobines (hypothesis B in fig. 1). This result promoted Xu et al. (2009) to claim that the duplicated *RNASE1* genes in leaf monkeys were actually produced by a single-gene duplication event prior to leaf monkey speciation, that is, one-gene duplication hypothesis, and attribute the results inferred from their noncoding trees and that of Zhang's (2006) study to gene conversion. Hence, the evolutionary pattern of *RNASE1* genes in leaf-eating monkeys seems more complex and confusing than has previously been recognized.

Though it has long been believed that phylogenetic analyses have contributed much to trace the scenario of gene evolution, however, the tree topology of phylogenetic trees can sometimes be biased by sampling errors, leading to an unreliable estimation. More complete sampling of the animal group in question and of the nucleotide sites in use would provide a clearer picture of gene evolution (Pollock et al. 2002; Zwickl and Hillis 2002; Kuraku et al. 2009). Currently, our understanding of Colobine *RNASE1* gene evolution largely depends on the analyses of limited taxonomic sampling. So, it would be of considerable interest and importance to study *RNASE1* sequences of more leaf monkeys to shed new lights on the clarification of the two contrasting hypotheses, which is critical for a deeper understanding about the true evolutionary history of Colobines RNases. In addition, analysis of additional *RNASE1* gene sequences allows us to have a more thorough understanding about the molecular basis underlying the adaptation of Colobinae to the unique leaf-feeding lifestyle.

In the present study, we newly determined *RNASE1* genes of two African colobines, *Colobus polykomos* and *Colobus angolensis*, and eight Asian colobines, *Trachypithecus johnii*, *Trachypithecus vetulus*, *Trachypithecus francoisi leucocephalus*, *Trachypithecus francoisi*, *Trachypithecus phayrei*, *Rhinopithecus avunculus*, *Rhinopithecus bieti*, and *Rhinopithecus roxellanae*. A total of 21 sequences were obtained, in which three *RNASE1* genes were found in *T. vetulus*, whereas two genes in all the other colobines. Together with previously reported five leaf monkeys, the most comprehensive investigation to date of *RNASE1* gene evolution in Colobinae was performed in this study.

Materials and Methods

Data Sets

The colobines (Colobinae subfamily) are a diverse clade of Old World primates and their generic classifications have long been in debate, from four to ten genera (Groves 1970, 2001; Szalay and Delson 1979; Zhang and Ryder 1998; Brandon-Jones et al. 2004; Sterner et al. 2006). Here the genus name followed the classification scheme of Sterner et al. (2006). In total, 15 species belonging to seven genera of Colobinae subfamily were examined in this study. The seven genera are African *Colobus* and *Piliocolobus* and Asian *Nasalis*, *Pygathrix*, *Semnopithecus*, *Trachypithecus*, and *Rhinopithecus*. For each sample, total genomic DNA was isolated from blood or frozen tissues using standard proteinase K, phenol–chloroform extraction (Sambrook et al. 1989).

Pancreatic ribonuclease gene (*RNASE1*) with an approximate length of 2 kb, spanning the entire coding region and an upstream noncoding exon separated by an intron, were amplified by polymerase chain reaction (PCR) using primer pair reported by Zhang et al. (2002). The amplified PCR products were cloned into PMD18-T vector (Takara, China) and transformed into an ultracompetent *Escherichia coli* cell (Takara). Plasmids containing the *RNASE1* inserts were extracted using GenElute Plasmid Miniprep Kit (Sigma–Aldrich Co.). About 30 clones per ligation reaction were sequenced in both directions with an ABI PRISM 3700 DNA sequencer (PE Biosystems). Only those sequences with more than three mutations in protein sequence and corroborated by at least two times of independent amplification and sequencing were used in the analysis. These sequences were supported by most clones. The newly determined *RNASE1* gene sequences are in GenBank under accession numbers GQ334693–GQ334713. In our analysis, *RNASE1* sequences of *P. nemaeus* and *C. guereza* were those from Zhang et al. (2002) and Zhang (2006). For those previously reported *RNASE1* sequences of *N. larvatus*, *S. entellus*, and *P. badius* (Schienman et al. 2006), two very similar sequences were found from a species in some cases, which could be attributed to putative alleles. We just used one sequence of alleles as an independent gene copy in the analyses. In addition, *RNASE1* sequences of six noncolobine primates available from GenBank were also downloaded. A New World Monkey, Marmoset *Callithrix jacchus*, was used as outgroup.

Phylogenetic Reconstructions

Separate and combined alignments of *RNASE1* coding (468 bp) and noncoding DNA sequences (1,575 bp) for 33 ingroup and two outgroup sequences were carried out with ClustalX program (Thompson et al. 1997) and when necessary edited by eye. The protein alignment is shown in supplementary figure 1 (Supplementary Material online). Phylogenetic trees were reconstructed using molecular evolutionary genetics analysis (Kumar et al. 2008) for Neighbor-Joining (NJ) analyses based on Kimura's two-parameter model with complete deletion option for gaps, as well as PAUP*4.0b8 (Swofford 2001) for maximum parsimony (MP) analyses. For MP analyses, a heuristic search strategy was employed with the tree-bisection-reconnection branch-swapping algorithm, random addition of taxa and 1,000 replicates per search. Only one of the best trees found during branch swapping was saved (MULTREES = NO in PAUP*), and zero length branches were collapsed. The reliability of the tree topologies was evaluated using bootstrap support (BS, Felsenstein 1985) with 1,000 replicates for NJ and MP analyses.

Selective Constraint Analyses

The nonsynonymous to synonymous rate ratio ω (dn/ds) provided an indication of the change of selective pressures. A dn/ds ratio = 1, <1, and >1 will indicate neutral evolution, purifying selection, and positive selection on the protein involved, respectively. We first applied the method of Yang and Nielsen (2000) for estimating dn and ds between two sequences. The codon-substitution models implemented in the CODEML program in the PAML package (Yang 2007) were then used to analyze changes of selective pressure. All models correct the transition/transversion rate and codon usage biases (F3 × 4). Different starting ω values were also used to avoid the local optima on the likelihood surface (Suzuki and Nei 2001). To detect changes in selective pressure after gene duplication, the "two-ratios" model was used, which assumes that the branches of interest have different ratios from the background ratio ω_0 (Yang 1998, 2002; Yang and Nielsen 1998) in the branch-specific models. Model B assumes two ω ratios: one restricted to lineages predating a gene duplication event and the second restricted to lineages resulting from the gene duplications. Model C assumes three ω ratios: one restricted to lineages predating a gene duplication event, the second restricted to ancestor lineages leading to duplicated gene copies, and the third restricted to lineages corresponding to the duplicated gene copies. Model D assumes different ω ratios for each of the ancestor lineages and the background ratio for all the other lineages in the phylogeny (Hileman and Baum 2003). We also use the site-specific model, which allow for variable selection patterns among amino acid sites, M8a and M8, to test for the presence of sites under positive selection and identify them (Nielsen and Yang 1998; Yang 2000). We construct likelihood ratio tests (LRTs) to compare M8a with M8. Significant difference between the models was evaluated by calculating twice the log-likelihood difference following a χ^2 distribution, with the number of degrees of freedom equal to the difference in the numbers of free parameters between the models. M8 models allow for positively selected sites. When this positive-selection models fitted the data significantly better than the corresponding null model (M8a), the presence of sites with $\omega > 1$ is suggested. The conservative empirical Bayes approach (Yang et al. 2005) will then be used to calculate the posterior probabilities (PPs) of a specific codon site and identify those most likely to be under positive selection.

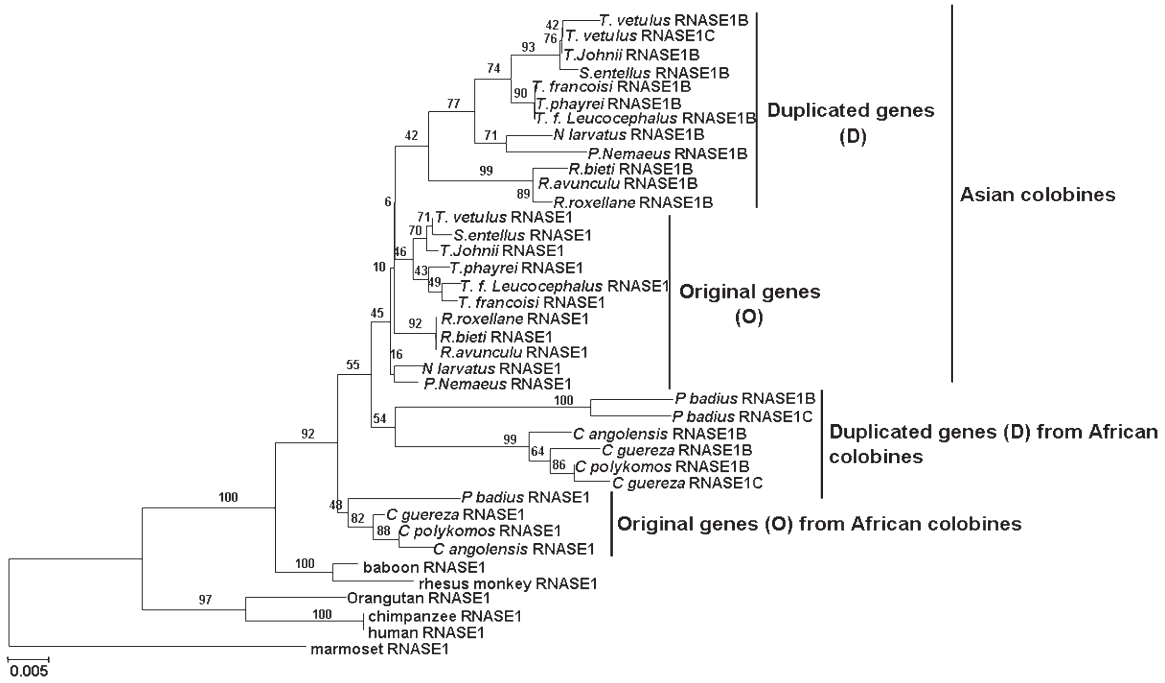

FIG. 2. NJ bootstrap consensus phylogenetic trees based on coding region (468 bp) of *RNASE1* genes. Arrows are indicating putative duplication events. MP analysis produced nearly identical tree topology to that of NJ analysis with similar bootstrapping supports.

We inferred ancestral RNASE1 sequences based on parsimony (Li and Gojobori 1983) and Bayesian (Zhang and Nei 1997) methods. Hon-new software (Zhang 2000) was used to estimate conservative and radical nonsynonymous distances between these sequences. Numbers of radical and conservative substitutions per site for each branch were calculated. In addition, GENECOVN software (Sawyer 1989) was used to search for evidence of gene conversion events among gene duplicates.

Results

Phylogenetic Inferences of Colobine *RNASE* Gene Sequences

Phylogenetic trees for coding and noncoding sequences were shown in figures 2 and 3, respectively. Noncoding MP and NJ analyses presented similar tree topologies, and both clearly split *RNASE1* sequences into African (MP BS = 82%; NJ BS = 88%) and Asian (MP BS = 80%; NJ BS = 55%) colobine-specific clusters, supporting the independent duplication of *RNASE1* in Asian and African colobines. Intriguingly, both our MP and NJ coding phylogeny did not cluster the duplicated genes of Asian and African colobines into a single clade, as indicated in previous coding analyses of fewer colobines species (Schienman et al. 2006; Xu et al. 2009), but group all Asian *RNASE* gene sequences together, albeit with weak bootstrapping supports (BS < 50%), and all African sequences were placed at the base of the phylogeny. Although the African sequences did not form a monophyletic group, presumably due to the small number of nucleotides used, our coding analyses of more colobines is notable because it appears that the hypothesis of a gene duplication event prior to the divergence of Asian and African colobines was no longer supported here, and the independent duplications in the two lineages seems more likely.

Based on our phylogenetic analyses, we can find both coding and noncoding trees indicated that *RNASE1* gene was duplicated after the speciation of Asian and African colobines. However, they gave different results about the evolutionary scenario of *RNASE1* within Asian and African colobines clusters. In coding analyses, Asian and African clusters each divided *RNASE1* sequences into two gene groups, corresponding to duplicated (D) and original (O) gene copies, implying two basal gene duplication events occurred after the Asian and African colobines speciation, with one before the Asian colobines radiation and the other before the African colobines radiation (fig. 2). However, the D and O gene groups in Asian and African clusters all received very poor supports in the analyses (BS < 50% for all four groups except NJ BS = 54% for African D groups). By contrast, in noncoding analyses, eight independent duplication events, two in African and six in Asian colobine clusters, were observed (fig. 3). Four of them received high bootstrapping supports in the analyses (BS > 90%). In addition to the basal duplication events mentioned above, the coding and noncoding phylogenies both suggested three additional recent gene duplications within *T. vetulus*, *C. guereza*, and *P. badius* lineages, resulting in three gene copies in these three colobine species (figs. 2 and 3).

Besides the separate analyses of coding and noncoding regions, we also attempted to reconstruct tree based on 2,043 nucleotide sites of coding and noncoding sequences

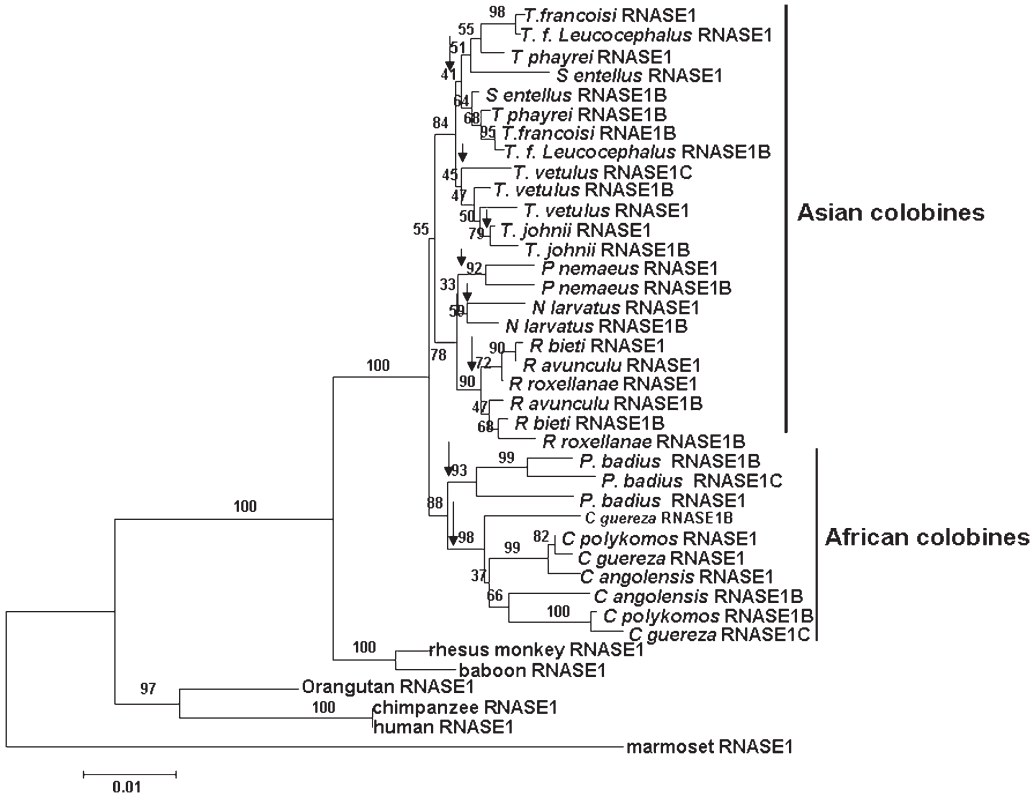

Fig. 3. NJ bootstrap consensus phylogenetic trees based on noncoding region (1,575 bp) of *RNASE1* genes. Arrows are indicating putative duplication events. MP analysis produced nearly identical tree topology to that of NJ analyses with similar bootstrapping supports.

(fig. 4). The resulting MP and NJ phylogenies divided *RNASE1* sequences into African (MP BS = 82% and NJ BS = 89%) and Asian (MP BS = 85% and NJ BS = 88%) colobine-specific clusters with high statistical supports, reinforcing the classical view that *RNASE1* gene duplications occurred after the divergence between Asian and African colobines. However, different from both separate analyses, the combined analyses indicated five parallel gene duplication events, two in African and three in Asian clusters, occurred after the Asian and African colobines speciation. All these duplications received high bootstrapping supports (BS > 80%). Although the *RNASE1* genes of *Pygathrix* and *Nasalis* did not form a monophyletic group, the duplicated gene copies of *Pygathrix* and *Nasalis* were clustered with medium to high BS (MP BS = 70%; NJ BS = 86%). Interestingly, we find that the tree topology supporting a single-gene duplication event prior to the divergence of Asian and African colobines was significantly worse than our combined tree ($P < 0.05$) under both Kishino–Hasegawa test (Kishino and Hasegawa 1989) and Shimodaira–Hasegawa test (Shimodaira and Hasegawa 1999).

The true phylogeny of Colobinae subfamily is not yet well established thus far; however, a series of recent studies based on mitochondrial DNA, nuclear DNA, and retroposonintegrations have produced a preferred relationship among the genera (Sterner et al. 2006; Osterholz et al. 2008; Ting et al. 2008). This relationship was robustly supported by present noncoding and combined analyses. African colobines diverged earliest (noncoding: MP BS = 55% and NJ BS = 80%; combined: MP BS = 85% and NJ BS = 89%). Asian colobines were split into two clades, one included *Nasalis*, *Pygathrix*, and *Rhinopithecus* genera (noncoding: MP BS = 88% and NJ BS = 78%; combined: MP BS = 82% and NJ BS = 70%), and the other included *Semnopithecus* and *Trachypithecus* genera (noncoding: MP BS = 85% and NJ BS = 84%; combined: MP BS = 84% and NJ BS = 80%). Coding analysis, however, yielded an inconsistent genus-level relationship with published Colobinae phylogeny. It did not group *Pygathrix*, *Nasalis*, and *Rhinopithecus* as a monophyletic clade. Moreover, the sister-group relationship between *Semnopithecus* and *Trachypithecus* in Asian O groups of the coding tree received weak bootstrapping supports (BS < 50%).

In sum, broader and detailed phylogenetic analyses of *RNASE1* sequences of more leaf monkeys using coding, noncoding, and combined regions provided consistent evidence of supporting the *RNASE1* gene is duplicated after the speciation of Asia and African colobines, a hypothesis originally proposed by Zhang (2006), although the statistical supports for the groups in the coding tree were relatively weak. However, different hypotheses of *RNASE1* gene evolution within Asian and African colobines were observed among present coding, noncoding, and combined analyses. Despite this, several lines of evidence indicates that coding analyses is less reliable than noncoding and combined analyses. First, the coding analyses suffer from a large sampling error and from a lack of high statistical supports, a situation having also been found in previous

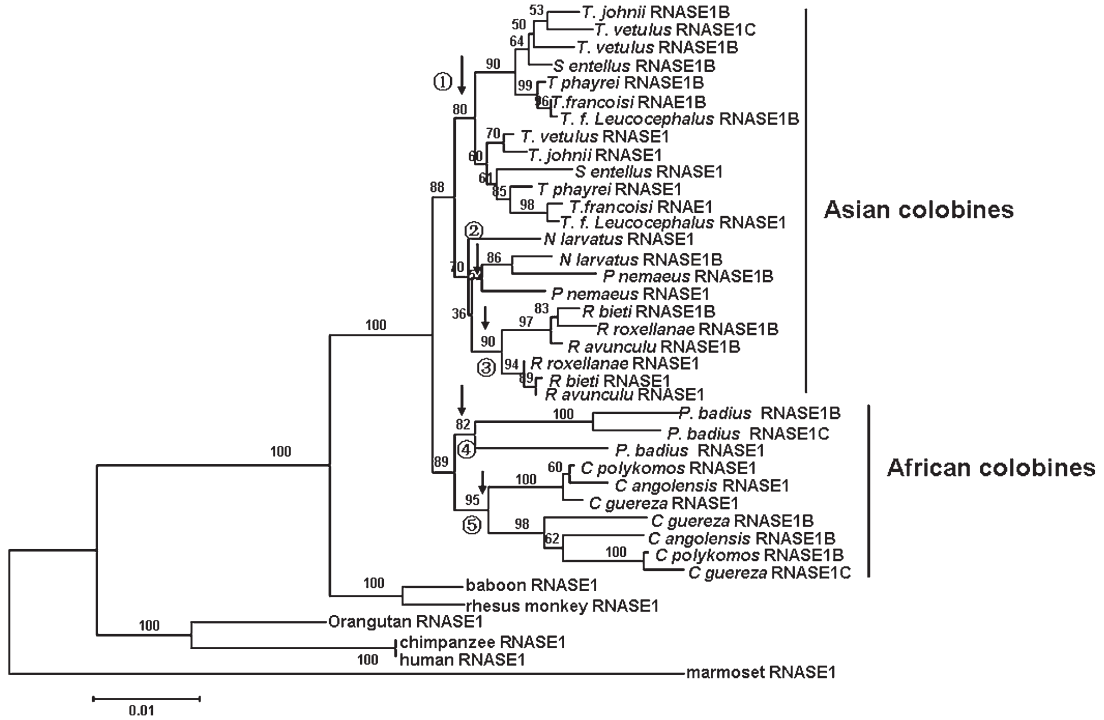

Fig. 4. NJ bootstrap consensus phylogenetic trees based on coding and noncoding regions (2,046 bp) of *RNASE1* genes. MP analysis produced nearly identical tree topology to that of NJ analyses with similar bootstrapping supports. Numbers 1–5 and arrows indicate the duplication events.

coding analyses of fewer colobines (Schienman et al. 2006; Xu et al. 2009), mostly due to the much small number of nucleotide sites used. By contrast, the noncoding and combined analyses based on four to five times of nucleotides than that of coding analyses demonstrate moderate to high bootstrapping supports for most of the nodes. Second, the inferred relationships among the genera of Colobinae from the coding analyses were incongruent with the published phylogeny, whereas those from the noncoding and combined analyses followed current knowledge of Colobinae phylogeny. Hence, noncoding or combined trees is more likely to represent "true" evolutionary history of *RNASE1* genes within Asian and African colobines. In the studies of Schienman et al. (2006) and Xu et al. (2009), the tree based on noncoding sequences was speculated to be spurious due to gene conversion. However, in present analyses of 15 species, only within five (*T. johnii*, *T. vetulus*, *P. nemaeus*, *N. larvatus*, and *P. badius*) and one species (*P. badius*), in noncoding and combined trees, respectively (figs. 3 and 4), the duplicates were more similar to each other than to their orthologues in the other species. Moreover, when gene conversion test of Sawyer's (1989) method was used to examine the noncoding and combined sequences, no significant results were given, suggesting that gene conversion is unlikely to have confounded the analyses (data not shown).

Subsequent selective constraint analyses using CODEML likelihood method on the combined tree topology, which was inferred from the maximum nucleotide sites available, were presented below. Compared with the noncoding tree, the combined tree provided higher statistical supports for major nodes (e.g., MP BS = 85% and NJ BS = 88% vs. MP BS = 55% and NJ BS = 80% for Asian colobines cluster) and more parsimonious deduction for the number of duplication events (eight vs. five). In fact, considering that CODEML analyses may be sensitive to the tree topology employed, we also conducted the analyses using the noncoding tree and obtained nearly the same results.

Selective Patterns in Colobinae Subfamily

The method of Yang and Nielsen (2000) was used to estimate synonymous (ds) and nonsynonymous (dn) substitution rates between two sequences. In total, 206 out of 528 pairwise comparisons for these Colobinae *RNASE1* genes have a ω (dn/ds) > 1, indicative of the positive selection acting during colobines *RNASE1*'s evolution. The plot of dn against ds is shown in figure 5a.

Further examination of the selective patterns in colobines was performed using the codon-based maximum likelihood analyses (table 1). The analyses were conducted using the combined tree topology but with the poorly supported nodes in NJ and MP phylogenies (BS $< 70\%$) collapsed into polytomy (fig. 6). Thirty-three colobine and two most closely related noncolobine primates, rhesus monkey and baboon, were used for the analyses. As summarized in table 1, models B, C, and D in the branch-specific models revealed a significantly better fit to the data than did the one-ratio model, M0 ($0.001 < P < 0.01$, $P < 0.001$, and $P < 0.001$, respectively). Under model B, estimates of the ω_1 ratio for the lineages resulting from gene duplication events is 1.1298, and the background ratio ω_0 ratio for those predating gene duplication is 0.2907. Interestingly, when we

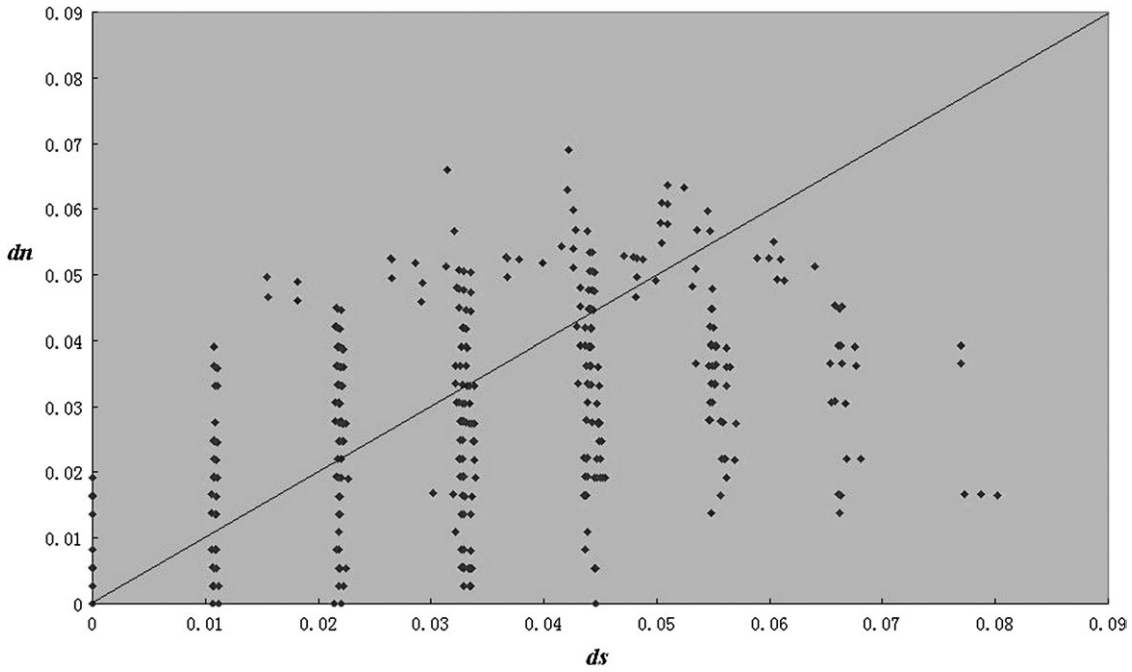

Fig. 5. The plot of nonsynonymous (dn) and synonymous (ds) substitution rates for pairwise comparisons of RNASE1 sequences.

assigned ω_1 and ω_2 to the ancestor lineages leading to duplicated gene copies and the terminal lineages corresponding to duplicated gene copies, respectively, under model C, we found that the ω_1 ratio (2.8014) was much higher than ω_2 ratio (0.5928) and ω_0 ratio (0.2904), indicating that positive selection may have operated on these ancestor lineages (a to e branches in the tree of fig. 6). Estimates of the ω ratios for each of the ancestor lineages under model D were 1.079 (0.0257/0.0238), ∞ (0.0197/0), ∞ (0.0229/0), ∞ (0.0440/0), and 1.0842 (0.0259/0.0239), respectively.

In addition, we also attempted to divide the nonsynonymous substitutions into radical and conservative substitutions to look for the signal of positive selection of *RNASE1* genes in Colobines. A significantly higher rate of radical nonsynonymous substitution than conservative substitution has been regarded as evidence for positive selection (Hughes 1992, 1994; Hughes and Hughes 1993). Our results (table 2) showed that when amino acid charge was considered, the number of radical substitutions per site was detected to be significantly greater than that of conservative substitutions per site on three (b, c, and d branches) of five ancestor branches leading to duplicated gene copies (b: $P = 0.033671$; c: $P = 0.000989$; d: $P = 0.017417$; Fisher's exact test). Therefore, this result, combined with LRT results described above, provided evidence for the operation of positive selection on the ancestor lineages leading to duplicated gene copies. Furthermore, our finding suggested that following gene duplication there has been a change in amino acid charge, reflecting the critical role of amino acid charge during the adaptive evolution of *RNASE1* genes in leaf monkeys.

We therefore examined important adaptive amino acid replacements that are likely to be responsible for acquisition of digestive specified function of Colobinae RNase using our data sets. Results from maximum likelihood analyses were shown in table 1. The positive-selection model (M8) in site-specific models provided a significantly better fit to the data than did the neutral model (M8a) ($P < 0.001$) and suggested the presence of positively selected residues. Nine such sites were predicated (positions 1, 4, 6, 39, 42, 78, 98, 101, and 122) with high PPs (PP > 95%). In previous study of *P. nemaeus*, Zhang et al. (2002) have observed nine amino acid substitutions in the duplicated *RNASE1* gene. PAML analyses here predicated all but two (sites 32 and 83) of them as having been under positive selection and, moreover, identified two adaptive sites, 78 and 101, that have not been previously characterized in the previous studies.

In addition, we investigated the evolutionary pattern of the identified nine positive sites among the species examined. As shown in figure 6, seven parallel amino acid substitutions in seven sites occurred in Asian and African duplicated *RNASE1* genes, including R4Q, K6E, R39W, P42S, R78K, R98Q, and A122D. Among them, R4Q, K6E, and R39W have been previously observed and characterized by mutagenesis experiments in Zhang's (2006) study of *P. nemaeus* and *C. guereza*, and make up three known adaptive parallel amino acid replacements contributing to the decrease of the optimal pH in the Asian and African colobine small intestine. The present analyses, therefore, identified four previously uncharacterized parallel amino acid changes (P42 S, R78K, R98Q, and A122D), which are particularly interesting candidates for future study by site-directed mutagenesis. In addition to the parallel changes displayed between Asian and African colobine lineages, there are multiple occasions of parallel changes among different lineages within either Asian or African colobine clusters (fig. 6). For example, two parallel

Table 1. CODEML Analyses of Selective Pattern for RNASE1 Genes in Colobinae.

Models	ln *L*	Parameter Estimates	2 Δ *L*	Positively Selected Sites
Branch-specific models				
M0	−1430.376388	$\omega = 0.6591$		
Model B	−1425.11674	$\omega_1 = 1.1298$; $\omega_0 = 0.2907$ $\omega_1 = 2.8014$; $\omega_2 = 0.5289$;	(Model B vs. M0) 10.519296**	
Model C	−1421.250942	$\omega_0 = 0.2904$ ω_1 (a branch) = 1.079; ω_2 (b branch) = ∞; ω_3(c branch) = ∞; ω_4 (d branch) = ∞; ω_5 (e branch) = 1.0842;	(Model C vs. M0) 18.250892***	
Model D	−1418.532975	$\omega_0 = 0.3776$	(Model D vs. M0) 23.686826***	
Site-specific models				
M8a	−1388.997102	$P_0 = 0.65038$; $P = 0.00500$; $q = 2.65039$; $P_1 = 0.34962$; $\omega = 1$ $P_0 = 0.90761$; $P = 0.00500$; $q = 0.01177$;		Not allowed
M8	−1373.642373	$P_1 = 0.09239$; $\omega = 5.24151$	(M8 vs. M8a) 30.709458***	1 (0. 988), 4 (0. 996), 6 (0. 999), 39 (0. 999), 42 (0. 981), 78 (0. 990), 98 (0. 974), 101 (0. 981), 122 (0. 999)

**0.001 < *P* < 0.01; ***$P < 0.001$; ω values larger than 1 shown in boldface.

substitutions (K6E and R39W) have been acquired in the African d and e branches and two parallel substitutions (R1G and R39W) in the Asian b and c branches, etc.

Besides the parallel changes described above, it is interesting to find seven nonparallel amino acid substitutions at five of these sites, including R1S and R1K in the duplicated genes of *P. badius* and *Colobus* genus species, respectively; R4A and R4G in the duplicated genes of *P. badius* and *Rhinopithecus* genus species, respectively; K6Q in those of *Rhinopithecus* genus species; R39S in those of *Trachypithecus* genus and *S. entellus* species; and R98G in *T. johnii*, *T. vetulus*, and *S. entellus* (fig. 6). It is possible that these amino acid substitutions may also be important for reducing the optimal pH of colobine RNases. On the other hand, the observed diversity of amino acid substitutions also indicated that gene conversion is not likely to have acted on the coding region, consistent with the view that positive selection may have prevented the gene conversion in *RNASE1* coding sequences (Schienman et al. 2006).

Molecular Dating

The relative rate test of Li and Bousquet (1992) was performed to test the molecular clock hypothesis. We first divided *RNASE1* genes into five groups according to five

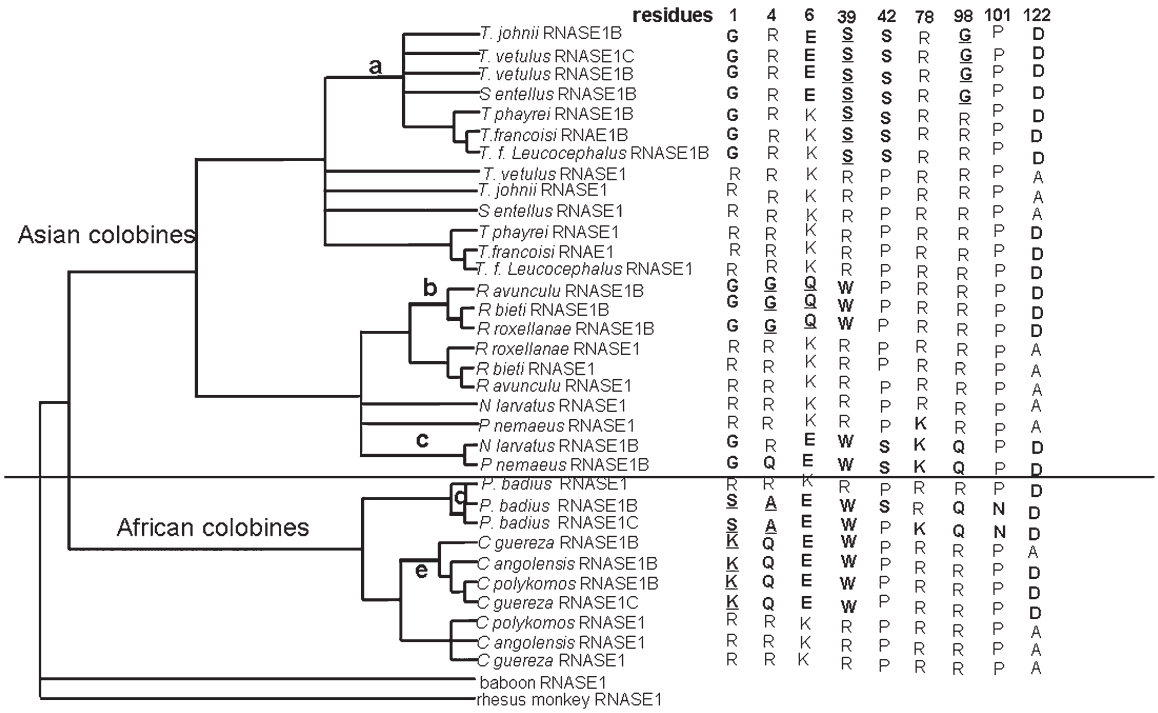

Fig. 6. Evolutionary patterns of positively selected sites among species examined. The parallel amino acid substitutions were boldface and the nonparallel amino acid substitutions were boldface and underlined.

Table 2. Evidence of Positive Selection in Ancestor Lineages from Radical/Conservative Nonsynonymous Substitution Rate Ratio Test for Amino Acid Charge.

Lineages	Number of Radical Substitutions Per Site	Number of Conservative Substitutions Per Site	Fisher's Exact Test
b	0.05	0.01	$P = 0.033671$
c	0.058	0	$P = 0.000989$
d	0.08	0.024	$P = 0.017417$

independent gene duplication events indicated in combined tree topology (fig. 4). The outgroup sequences in all comparisons were rhesus and baboon. No significant rate difference was detected between these five groups at 0.5% level (after Bonferroni correction), suggesting that molecular clock assumption holds for this region. We use three calibration points, that is, the divergence time of 15 million years (Ma) between colobines and ceropithecines (Delson 1994), 10.8 Ma between Asian and African colobines (Sterner et al. 2006), and 8.3 Ma between *Colobus* and *Piliocolobus* genus (Sterner et al. 2006), to roughly estimate the gene duplication times. The results show that five duplication events (branches 1–5 indicated in fig. 4) occurred, respectively, at the date of 4.46 (3.82–5.1) Ma, 5.1 (4.46–5.73) Ma, 2.87 (2.23–3.5) Ma, 9.55 (9.55–10.82) Ma, and 7.96 (7.01–8.92) Ma. We can find that all duplication events postdated the divergence of Asian and African colobines and also provide evidence for independent duplication hypothesis.

Discussions

Pancreatic RNase genes implicated in the adaptation of colobines to leaf eating have long intrigued evolutionary biologist. Our study with more intensive taxonomic and character sampling contribute to a clearer picture of Colobine *RNASE1* gene evolution.

The present broader and detailed phylogenetic analyses yielded two important findings: 1) All trees provided consistent evidence for independent duplication hypothesis proposed by Zhang (2006). The hypothesis of a single duplication prior to the radiation of Asian and African colobines was not supported here (Schienman et al. 2006; Xu et al. 2009). In the study by Schienman et al. (2006) and Xu et al. (2009), the coding tree clustered all duplicated *RNASE1* genes into a single clade, whereas the original genes were grouped into another clade, with weak bootstrapping supports (fig. 1). Our coding analyses, however, generated different tree topology, albeit also with low nodal supports. The duplicated genes of Asian and African colobines are no longer clustered, and duplicated genes of African colobines were moved to the base of the phylogeny. All genes of Asian colobines were grouped into a clade (fig. 2); 2) No obvious evidence of gene conversion in *RNASE1* gene was found, favoring independent evolution of Colobine *RNASE1* gene duplicates. To resolve the tree topology incongruence from their analyses of coding and noncoding regions (fig. 1), Schienman et al. (2006) and Xu et al. (2009) proposed that high levels of gene conversions among *RNASE1* gene duplicates have taken place predominantly in the noncoding region, although they cannot rule out the possibility of independent duplications after speciation. In their noncoding tree, the duplicates within each of the five species examined are more closely related to each other than to their orthologues in the other species. Our analyses of 15 species, however, showed that only within four and one species in noncoding and combined analyses, respectively, the duplicates were more similar to the each other than to their orthologues (figs. 3 and 4). With regard to the five species previously studied, the duplicates within *S. entellus* and *C. guereza* are no longer clustered in our noncoding tree, whereas the duplicates within all but *P. badius* are not clustered in the combined tree. In addition, the statistical algorithm GENECOVN did not predict any gene conversion events among the duplicates for our noncoding and combined sequences. Combined with the observed diversity of amino acid substitutions among the species examined, which is not likely to be expected if gene conversion occurred, the present study did not support the conclusion drawn by Schienman et al. (2006) and Xu et al. (2009) that gene conversion played a significant role in the evolution of Colobine *RNASE1* and significantly indicated the independent evolution.

Not only the phylogenetic pattern of *RNASE1* sequences reported here provide a solid reference for further studies investigating *RNASE1* evolution and function but also our selective constraint analyses provided interesting insights. Our CODEML branch analyses find significant evidence for adaptive evolution of ancestor lineages leading to duplicated gene copies, which is strengthened by the observation of a significantly higher rate of radical than conservative nonsynonymous substitution for amino acid charge on these lineages. The radical/conservative rate ratio results also suggested that following gene duplication there has been a change in amino acid charge, consistent with the results of earlier works of fewer colobines (Zhang et al. 2002; Zhang 2006; Schienman et al. 2006), in which the charge-altering substitutions were shown to be selectively favored for enhanced RNase activity at the relatively low pH environment of the colobine small intestine. Interestingly, our previous study of positive selection acting on *RNASE1* duplicated genes of Mustelidae species in order Carnivora also illustrated that the amino acid charge change was one of the crucial determinants of amino acid substitutions among the eight positively selected residues identified (Yu and Zhang 2006). Of course, the relatively small number of amino acid substitutions in our case may have limited the power of radical/conservative rate ratio test to detect more lineages with higher radical/conservative rate ratio. Our detection of significant positive selections in the ancestor lineages of particular species indicates functional divergences are most likely to occur after the gene duplication and before the speciation of these species.

In addition, CODEML site analyses identified nine residues under positive selection, several of which have not

been previously characterized by mutagenesis experiments. These results do not rule out the possibility that positive selection may have occurred at more lineages and sites, however. It has been suggested that the statistic power of CODEML analyses can be reduced when small numbers of nucleotide sites were used and the foreground branches tested were short, just as our case. In fact, we also attempted to test a model accommodating ω ratios to vary both among lineages of interest and amino acid sites, that is, the branch-site model, using our data set. No significant results from the LRTs were obtained. This is not unexpected when we have the reasons described above in mind.

Besides more parallel amino acid changes identified here, we also discovered a handful of nonparallel amino acid changes presented in particular colobine lineages. We suspected that when more Colobine *RNASE1* gene sequences became available, it will be possible to identify additional amino acid changes that may also contribute to the optimal pH of colobine RNases. In sum, our identification of new adaptive sites and amino acid substitutions provide a necessary foundation for further experimental investigations of *RNASE1* functional evolution in Colobinae.

As a principle digestive enzyme for adaptation of unique leaf feeding in Colobinae, *RNASE1* gene is so important that it will remain the focus of future study. Although available information suggested *RNASE1* gene duplication appear after divergence of Asian and African colobines and multiple events of gene duplication occurred in a lineage-specific mode, the timing and numbers of gene duplication events in different colobine lineages were still uncertain. The further analyses of *RNASE1* genes from representatives of the other four genera in Colobinae will be expected to address the question.

Supplementary Materials

Supplementary figure 1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Professors Jianzhi Zhang, Prof Peng Shi, and Soochin Cho for valuable comments. This work was supported by grants from the State Key Basic Research and Development Plan (2007CB411600) and National Natural Science Foundation of China (U0836603, 30621092).

References

Brandon-Jones D, Eudey AA, Geissmann T, Groves CP, Melnick DJ, Morales JC, Shekelle M, Stewart C-B. 2004. Asian primate classification. *Int J Primatol*. 25:97–164.

Clegg MT, Cummings MP, Durbin ML. 1997. The evolution of plant nuclear genes. *Proc Natl Acad Sci USA*. 94:7791–7798.

Delson E. 1994. Evolutionary history of the colobine monkeys in paleoenvironmental perspective. In: Davies AG, Oates JF, editors. Colobine monkeys: their ecology, behaviour, and evolution. Cambridge University Press, Cambridge. p. 11–43.

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

Force A, Lynch M, Pickett F, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 151:1531–1545.

Groves CP. 1970. The forgotten leaf-eaters, and the phylogeny of the colobinae. In: Napier JR, Napier PH, editors. Old World monkeys. Academic Press, NewYork. p. 555–587.

Groves CP. 2001. Primate taxonomy. Smithsonian Institution Press, Washington (DC).

Hileman L, Baum DA. 2003. Why do paralogs persist? Molecular evolution of CYCLOIDEA and related floral symmetry genes in Antirrhineae (Veronicaceae). *Mol Biol Evol*. 20:591–600.

Hughes AL. 1992. Coevolution of the vertebrate integrin alpha- and beta-chain genes. *Mol Biol Evol*. 9:216–234.

Hughes AL. 1994. Evolution of cysteine proteinases of eukaryotes. *Mol Phylogenet Evol*. 3:310–321.

Hughes AL, Hughes MK. 1993. Adaptive evolution in the rat olfactory receptor gene family. *J Mol Evol*. 36:249–254.

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol*. 29:170–179.

Kumar S, Dudley J, Nei M, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. 9:299–306.

Kuraku S, Meyer A, Kuratani S. 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol Biol Evol*. 26:47–59.

Li P, Bousquet J. 1992. Relative-rate test for nucleotide substitutions between two lineages. *Mol Biol Evol*. 9:1185–1189.

Li WH, Gojobori T. 1983. Rapid evolution of goat and sheep globin genes following gene duplication. *Mol Biol Evol*. 1: 94–108.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*. 148:929–936.

Ohno S. 1970. Evolution by gene duplication. Springer-Verlag, Heidelberg, Germany.

Osterholz M, Walter L, Roos C. 2008. Phylogenetic position of the langur genera Semnopithecus and Trachypithecus among Asian colobines, and genus affiliations of their species groups. *BMC Evol Biol*. 8:58.

Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Syst Biol*. 51:664–671.

Rispe C, Kutsukake M, Doublet V, Hudaverdian S, Legeai F, Simon JC, Tagu D, Fukatsu T. 2008. Large gene family expansion and variable selective pressures for cathepsin B in aphids. *Mol Biol Evol*. 25:5–17 2007.

Sambrook E, Fritsch F, Maniatis T. 1989. Molecular cloning. Cold Spring Harbor Press, Cold Spring Harbor (NY).

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol*. 6:526–538.

Schienman JE, Holt RA, Auerbach MR, Stewart CB. 2006. Duplication and divergence of 2 distinct pancreatic ribonuclease genes in leaf-eating African and Asian colobine monkeys. *Mol Biol Evol*. 23:1465–1479.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol*. 16:1114–1116.

Sterner KN, Raaum RL, Zhang YP, Stewart CB, Disotell TR. 2006. Mitochondrial data support an odd-nosed colobine clade. *Mol Phylogenet Evol*. 40:1–7.

Suzuki Y, Nei M. 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. *Mol Biol Evol*. 18:2179–2185.

Swofford DL. 2001. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4.0b8. Sinauer Associates, Sunderland (MA).

Szalay FS, Delson E. 1979. Evolutionary history of the primates. Academic Press, New York.

Ting N, Tosi AJ, Li Y, Zhang YP, Disotell TR. 2008. Phylogenetic incongruence between nuclear and mitochondrial markers in the Asian colobines and the evolution of the langurs and leaf monkeys. *Mol Phylogenet Evol*. 46:466–474.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The clustalx windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 24:4876–4882.

Xu L, Su Z, Gu Z, Gu X. 2009. Evolution of RNases in leaf monkeys: being parallel gene duplications or parallel gene conversions is a problem of molecular phylogeny. *Mol Phylogenet Evol*. 50:397–400.

Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol*. 51:423–432.

Yang Z. 2002. Inference of selection from multiple species alignments. *Curr Opin Genet Devel*. 12:688–694.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46:409–418.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 17:32–43.

Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22:1107–1118.

Yang ZZ. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15:568–573.

Yu L, Zhang YP. 2006. The unusual adaptive expansion of pancreatic ribonuclease gene in carnivora. *Mol Biol Evol*. 23: 2326–2335.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol*. 50:56–68.

Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol*. 18:292–298.

Zhang J. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet*. 38:819–823.

Zhang J, Nei M. 1997. Accuracies of ancestral amino acid sequences inferred by parsimony, likelihood, and distance methods. *J Mol Evol*. 44:S139–S146.

Zhang J, Zhang YP, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet*. 30:411–415.

Zhang YP, Ryder OA. 1998. Mitochondrial cytochrome b gene sequences of Old World monkeys: with special reference on evolution of Asian colobines. *Primates*. 39:39–49.

Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol*. 51:588–598.