Molecular Phylogenetics and Evolution 50 (2009) 397–400

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Evolution of RNases in leaf monkeys: Being parallel gene duplications or parallel gene conversions is a problem of molecular phylogeny

Lin Xu [a,1], Zhixi Su [b,1], Zhenglong Gu [c,*], Xun Gu [b,d,*]

[a] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[b] School of Life Sciences, Institutes of Biomedical Sciences, and Center for Evolutionary Biology, Fudan University, Shanghai, People's Republic of China
[c] Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA
[d] Department of Genetics, Development and Cell Biology, Iowa State University, 322 Science II, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 27 May 2008
Revised 3 November 2008
Accepted 5 November 2008
Available online 13 November 2008

1. Introduction

Whether parallel changes in DNA sequences after gene duplication are caused by positive selection is an intriguing issue in evolutionary biology (Stewart et al., 1987; Yokoyama and Yokoyama, 1990; Rodriguez-Trelles et al., 2003). With the help of rapid high-throughput sequencing technologies, it is now technically easy to obtain enough sequence data to conduct meaningful phylogenetic analysis. However, debates related to parallel evolution at molecular level cannot be resolved by simply increasing number of sequences. To show parallel adaptive evolution, we have to at first establish a solid foundation to demonstrate parallel nucleotide changes during evolution, and then we can test whether these parallel changes are adaptive or merely caused by mutational hotspots. Obviously, identification of parallel DNA substitutions is crucial; otherwise the problem itself is problematic. When the phylogeny is known, whether or not nucleotide sites have experienced parallel evolution can be reliably inferred, by either parsimonious or likelihood methods.

It has been well known for a large gene family that has undergone multiple gene duplications and gene losses, that accurate reconstruction of phylogenetic relationship becomes a difficult task, particularly when the coding sequence is short. An intuitive solution to overcome this problem is to combine surrounding noncoding sequences into the analysis, assuming that noncoding and coding sequences have the same evolutionary history. Since this approach requires that noncoding regions be reliably aligned, it is applicable in practice only for gene families that have been recently duplicated, as well as among closely-related species. However, it was well documented that chance of gene conversion or unequal crossover between recently duplicated genes increases, due to their high similarity in genomic sequences (Chen et al., 2007). Consequently, the inferred phylogeny based on coding sequence may differ significantly from that based on noncoding sequence. In this short communication, we reanalyzed a recent study (Zhang, 2006) and demonstrated that the uncertainty of inferred phylogeny may lead to different conclusions about parallel evolution of DNA substitutions.

2. Results and discussion

Zhang (2006) reported three parallel changes (G95A (R4Q), A100G (K6E) and C199T (R39W)) in RNase1B, a duplicate gene of pancreatic ribonuclease (RNase1), in two colobines (leaf monkeys), *Pygathrix nemaeus* and *Colobus guereza*. This statement was mainly drawn from the phylogenetic analysis of a combined nucleotide sequence alignment including both coding and noncoding regions. The inferred phylogeny implied two independent duplication events in each species after the split of *P. nemaeus* and *C. guereza* (Fig. 1A) The author further concluded that the observed parallel changes were driven by positive selection, based on the following two pieces of evidence: (i) Accelerated evolution of RNase1B genes was observed for these two species, and (ii) experimental data further demonstrated that these three changes could contribute to function of the novel ruminant-like alimentary system in colobines.

Now that more RNase1 and RNase1B sequences are available in five leaf monkeys (Schienman et al., 2006), we revisited this issue, focusing on whether parallel evolution did occur at these three sites between RNase1 and RNase1B genes. Not surprisingly, noncoding regions (5′-flanking or introns) lead to a phylogeny that indicates five independent duplication events in five leaf monkeys (Figs. 1B and 3; also Schienman et al., 2006), supporting the notion

* Corresponding authors. Address: Department of Genetics, Development and Cell Biology, Iowa State University, 322 Science II, Ames, IA 50011, USA.
E-mail addresses: zg27@cornell.edu (Z. Gu), xgu@iastate.edu (X. Gu).
[1] These authors contribute equally to this work.

1055-7903/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.ympev.2008.11.003

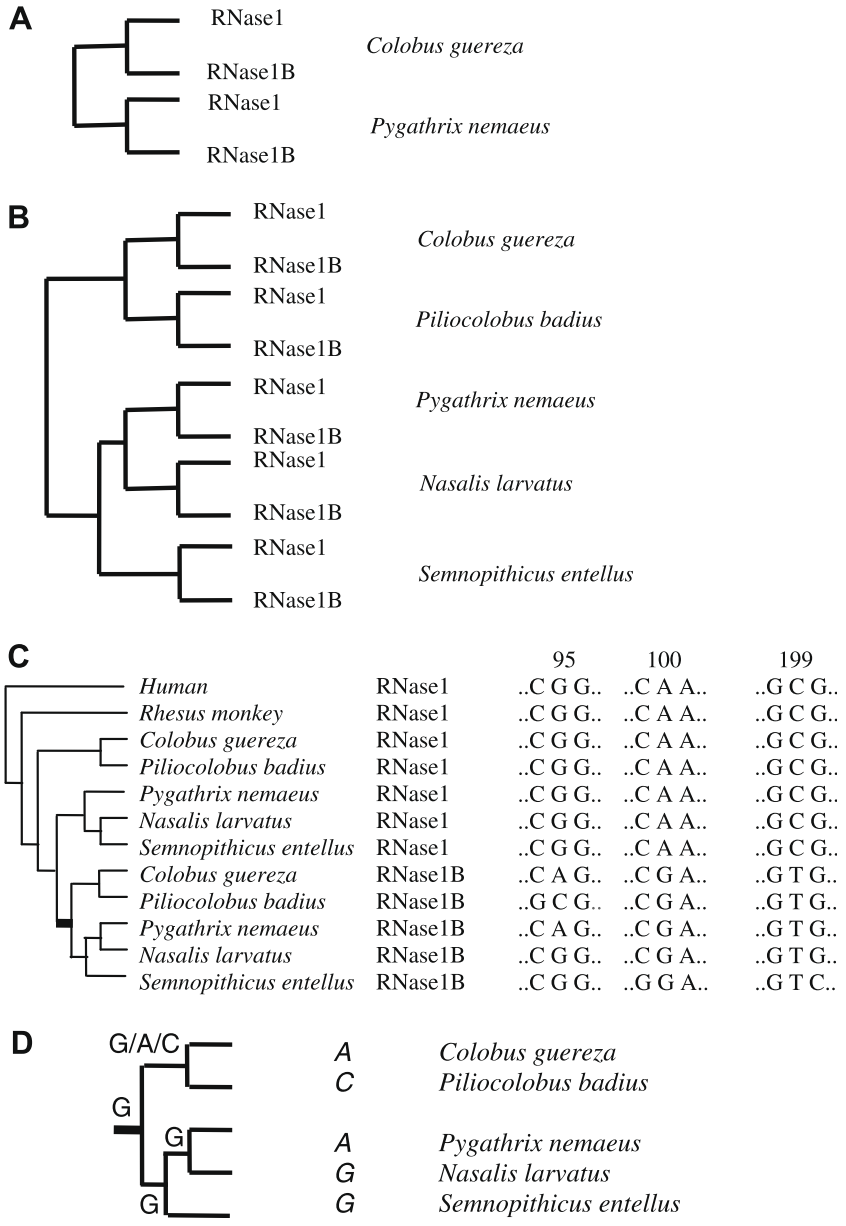

Fig. 1. (**A**) The original phylogeny used in Zhang (2006), based on combined coding and noncoding sequences. (**B**) A simplified phylogeny for colobine RNases from noncoding regions. Various tree-making methods result in virtually the same result. (**C**) A simplified phylogeny for colobine RNases and the nucleotide substitutions at the 95th, 100th and 199th sites in the RNase gene. Only one copy of RNase1 or RNase1B is shown even when there are more than one RNase1 or RNase1B gene in a colobine; for example, both RNase1β and RNase1γ of *C. guereza* are denoted by RNase1B. The DNA sequences at the four nucleotide sites in the RNase1 and RNase1B genes in five colobines (*Pygathrix nemaeus*, *Nasalis larvatus*, *Semnopithicus entellus*, *Piliocolobus badius* and *Colobus guereza*) are shown on the right. The gene tree is a simplified version of the detailed phylogeny in Fig. 2A. In this tree the RNase1s in *C. guereza* and *P. badius* are separated from the RNase1s in *P. nemaeus*, *N. larvatus* and *S. entellus*. We propose that they form sister clusters as in the case of RNase1Bs. A similar tree was proposed in Schienman et al. (2006). The bolded lineage indicates the location of single nucleotide substitutions occurred at 100th and 199th sites in the RNaseB gene. (**D**) Parsimonious inference of nucleotide changes at site 95.

that parallel duplication did occur frequently. By contrast, reconstructing a neighbor-joining or maximum parsimony tree of RNase protein sequences of five leaf monkeys plus rhesus and human, we found a totally different evolutionary scenario (Figs. 1C and 2). Apparently, RNase1B genes of leaf monkeys are clustered in the phylogeny, whereas RNases1 genes including the rhesus and human are clustered. Thus, applying the parsimonious principle for the number of duplication events, RNase1 and RNase1B in leaf monkeys can be explained by a single ancient duplication event. If one assumes rhesus monkey and human RNase1 genes as outgroups, it may require two duplication events because rhesus monkey and human RNase1 genes split leaf monkey RNase1 genes into two groups, but the bootstrapping value is too low (~50%) to draw any conclusion (Fig. 2).

Nevertheless, no matter where is the root we should locate, or whether it has one or two duplication events, according to the inferred phylogeny from protein sequences, parallel evolutions at two of the three sites (A100G and C199T) identified by Zhang (2006) do not hold. Instead, they can be easily explained by single nucleotide substitutions that occurred in the lineage indicated in Fig. 1C (bolded). At site 95, we found three independent nucleotide changes in lead monkey RNase1B genes. As shown in Fig. 1D, the proposed parallel nucleotide changes (G → A) between *C. guereza* and *P. nemaeus*) does not hold if the nucleotide at the common ancestor of *Piliocolobus badius* and *C. guereza* is nucleotide C, resulting in two successive changes (G → C) and (C → A). The parallel evolution seems to hold when the ancestral nucleotide is G or A. However, in the case of ancestral nucleotide

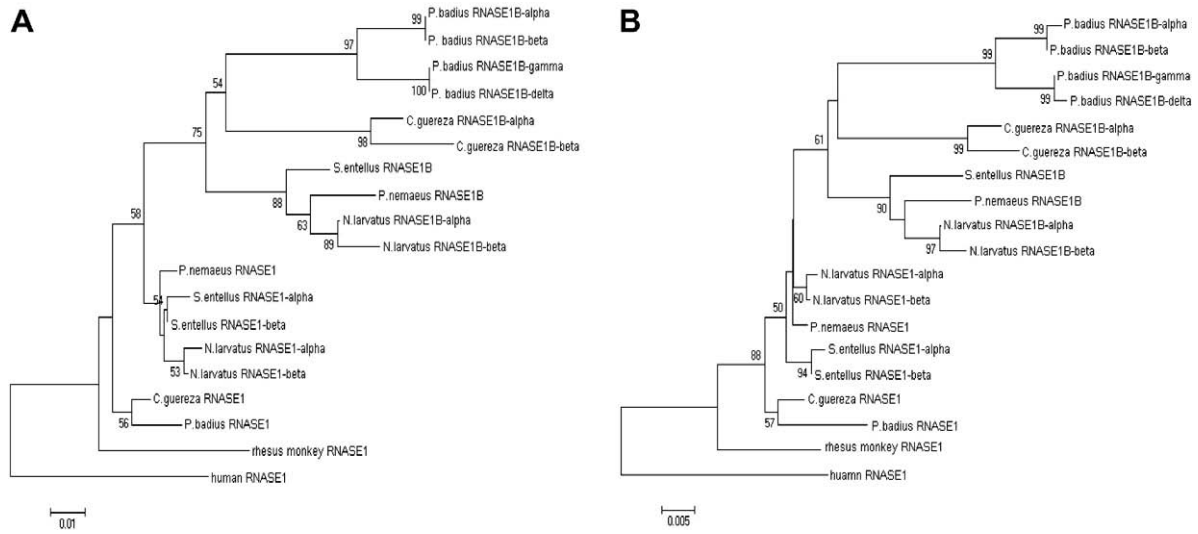

Fig. 2. Phylogenetic tree for the RNase1 and RNase1B genes in five colobine species. All RNase1 and RNase1B sequences in the five leaf monkeys (*Pygathrix nemaeus*, *Nasalis larvatus*, *Semnopithicus entellus*, *Piliocolobus badius* and *Colobus guereza*) and RNase1 sequences in the rhesus monkey and human (two outgroups) were retrieved from GenBank. The neighbor-joining (NJ) method from the MEGA3 software package was used to reconstruct the phylogenetic trees. Alpha, beta, gamma and delta were used to denote different copies if there are more than one RNase1 or RNase1B gene in a species. For *P. badius*, RNase1B-alpha = RNase1-P1a, RNase1B-beta = RNase1-P1b, RNase1B-gamma = RNase1-P2a and RNase1B-delta = RNase1-P2b in Schienman et al. (2006); for *C. guereza*, RNase1B-alpha = RNase1β and RNase1B-beta = RNase1γ in Zhang (2006); for *S. entellus*, RNase1-alpha = RNase1-Na and RNase1-beta = RNase1-Nb in Schienman et al. (2006); for *N. larvatus*, RNase1-alpha = RNase1-Na, RNase1-beta = RNase1-Nb, RNase1B-alpha = RNase1-Pa and RNase1B-beta = RNase1-Pb in Schienman et al. (2006). The RNAse1 and RNase1B genes are also clustered separately when the maximum parsimony (MP) method is used (not shown). (A) The NJ tree for the protein sequences with the Poisson model of amino acid substitution. (B) The NJ tree for the coding DNA sequences with Kimura's two-parameter model of nucleotide substitution.

A, the follow-up change (A → C) in *P. badius* is difficult to be explained by the adaptive hypothesis (Fig. 1D). Since these three possible states at the ancestor is equally parsimonious, and site 95 happens to be CpG mutational hotspot, whether there existed a parallel changes at this site remain open to question. The same conclusion is reached when the coding DNA sequences are used (Fig. 2B).

Our phylogenetic analysis is a routine practice for molecular systematic scientists. The point we attempt to make is that parallel evolution between RNase1 and RNase1B in leaf monkeys is an inference rather than a fact. It depends on the hypotheses about evolutionary history of RNase1 and RNase1B genes in leaf monkeys, which can be outlined as follows.

(i) *The recent parallel duplication hypothesis* (*Zhang, 2006*)*:* It is strongly supported by the phylogeny inferred from noncoding sequences (Fig. 3). The main assumption is that the evolutionary history of noncoding regions of these genes represents the true evolutionary history of the gene family. Moreover, it implies that the tree inferred from the protein coding regions, which differs dramatically from the noncoding tree, must have been distorted by strong adaptive evolution. Under this hypothesis, parallel changes had occurred at these three sites.

(ii) *The recent parallel gene conversion hypothesis:* It is supported by the phylogeny inferred from protein sequences or coding region nucleotides. Duplicate genes (RNase1 and RNase1B) that both appear in five leaf monkeys can be explained by a single ancient duplication event prior to leaf monkey speciation, or an additional duplication event in some leaf monkeys; we cannot distinguish between them as the bootstrapping value is low (Fig. 2). Besides, the rooting by using rhesus monkey and human as outgroups was ambiguous. This hypothesis implies that noncoding regions of two duplicate genes in the same species may have experienced extensive gene conversion (Chen et al., 2007). Under this phylogeny, there were no independent changes for two of the three originally proposed parallel sites, whereas the third one becomes spurious among leaf monkeys.

People may have different opinions about the evolutionary history of RNase1 and RNase1B genes in leaf monkeys. Functional analysis of these three sites cannot resolve this issue because it can be either interpreted as parallel adaptive evolution or functional divergence after the gene duplication. As a rule of thumb in phylogenetic analysis of character evolution, an evolutionary hypothesis (parallel evolution) cannot stand as a solved problem if two alternative inferred trees (from the coding or noncoding regions) result in contradicted results. In other words, the problem itself, i.e., whether parallel adaptations have occurred, can be nullified. However, from the Bayesian point of view, which has profoundly influenced research on molecular phylogeny and evolution, Zhang's (2006) analysis was legitimate, where the noncoding phylogeny was a 'prior' chosen by the author. Though we agree it is a valuable practice, we urge that it is necessary to consider alternative phylogenies, particularly when the main result such as parallel evolution is essentially phylogeny-sensitive.

We have made some efforts on investigating whether the currently available sequence data support the parallel gene conversion hypothesis. We examined the GC content of the noncoding regions and coding regions. As gene conversion has been shown to elevate GC content, having a higher GC content in the noncoding region is at least consistent with the possible gene conversion events. On average, the GC contents are 50% (5′-flanking, 854 bp), 52% (5′-UTR, 76 bp), 47% (introns 705 bp), 56% (coding region, 471 bp), and 55% (3′-UTR, 55 bp), which do not show a clear-cut pattern. Since the lengths of these 5′-UTR and 3′-UTR regions are too short, the resolution of inferred phylogeny has been poor (results not shown). Noting that primate RNase gene family has sophisticated synteny structures in chromosomes that include many tandem duplications, it is possible that gene conversions had occurred frequently and in a large-scale.

Since Colobinae (leaf monkey) is a subfamily of Old World monkeys that include 58 species in 10 genera (Sterner et al., 2006), the final settlement of this problem might be reached if we can sequence genomic regions of RNase1 and RNase1B genes in all these leaf monkeys. Synteny relationship in these genomic regions will

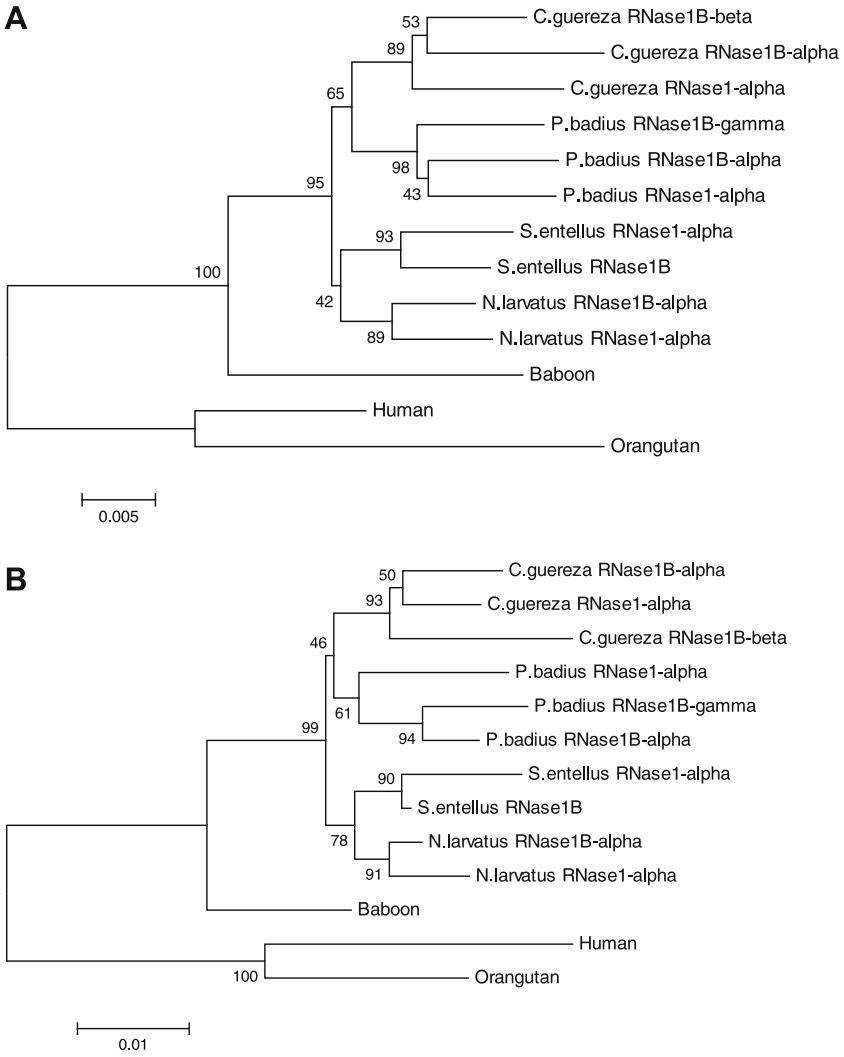

Fig. 3. Phylogenetic tree for the RNase1 and RNase1B genes in colobine species. (A) From 5′-flanking region. (B) From introns.

also be informative for resolving this issue. At this moment, however, one should be cautious to draw any general conclusion.

Acknowledgment

This work was supported by Hi-Tech Research and Development Program of China (863 project) (No. 2006AA02Z326 to Z.S. and X.G.).

References

Chen, J.M., Cooper, D.N., Chuzhanova, N., Ferec, C., Patrinos, G.P., 2007. Gene conversion: mechanisms, evolution and human disease. Nat. Rev. Genet. 8, 762–775.

Rodriguez-Trelles, F., Tarrio, R., Ayala, F.J., 2003. Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. Proc. Natl. Acad. Sci. USA 100, 13413–13417.

Schienman, J.E., Holt, R.A., Auerbach, M.R., Stewart, C., 2006. Duplication and divergence of 2 distinct pancreatic ribonuclease genes in leaf-eating African and Asian colobine monkeys. Mol. Biol. Evol. 23, 1465–1479.

Sterner, Kirstin N., Raaum, Ryan L., Zhang, Ya-Ping, Stewart, Caro-Beth, Disotell, Todd R., 2006. Mitochondrial data support an odd-nosed colobine clade. Mol. Phylogenet. Evol. 40 (1), 1–7.

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

Yokoyama, R., Yokoyama, S., 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., 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat. Genet. 38, 819–823.