

Molecular Timescale and Gene Tree Incongruence in the Guenons 3

TODD R. DISOTELL and RYAN L. RAAUM

Introduction

Guenons are well known for substantial variation in chromosome number and morphology (Ledbetter, 1981; Muleris *et al.*, 1986). While extensive multi-locus DNA sequence studies examining hominoid (Ruvolo, 1997; Satta *et al.*, 2000) and papionin (Harris and Disotell, 1998) phylogenetic relationships have been carried out, only a few sequences from multiple cercopithecine taxa have been collected. These limited studies have not examined a sufficient number of genetic systems nor enough taxa to elucidate the overall patterns of guenon phylogeny (Schätzl *et al.*, 1995; Page and Goodman, 2001). Results from an ongoing study of mitochondrial sequence variation in guenons are presented here along with estimates of the date of origin of the guenon radiation. The discussion focuses on variation in these inferred dates and recurring issues in guenon phylogeny.

Few studies including substantial numbers of guenon species have been conducted on genes or gene products, with the notable exceptions of Ruvolo's protein electrophoretic study (Ruvolo, 1988, reviewed in Disotell, 1996) and

TODD R. DISOTELL and RYAN L. RAAUM • Department of Anthropology, New York University, 25 Waverly Place, New York, NY 10003, USA.

The Guenons: Diversity and Adaptation in African Monkeys, edited by Glenn and Cords. Kluwer Academic/Plenum Publishers, New York, 2002.

the work of van der Kuyl and her colleagues on mitochondrial 12s rRNA (van der Kuyl *et al.*, 1995a,b, 2000a,b). The latter work utilized a short sequence of a slowly evolving mitochondrial gene that is well suited to identify the specific and sometimes subspecific identity of an unknown sample (van der Kuyl *et al.*, 2000b). However, these sequences are not very phylogenetically informative within the guenons due to a lack of sufficient variation (Disotell, 2000).

Current estimates of the date of divergence of the extant cercopithecine and papionin lineages range from 9.5 to 10 mya (Szalay and Delson, 1979; Kingdon, 1997). Estimates of the date of divergence of the extant cercopithecines include 6 mya (Szalay and Delson, 1979), 6.5–7 mya (Kingdon, 1997), and less than 3.5 mya (Leakey, 1988). Unfortunately, fossil evidence for cercopithecine evolution is rare, and does not exist before about 3.5 mya (Leakey, 1988). No significant new fossils have been reported since the topic was reviewed by Leakey (1988).

Methods

For phylogenetic analyses, we sequenced an 897 base pair fragment of the mitochondrial genome encompassing the 3' end of the NADH dehydrogenase subunit (ND) 4, three tRNA genes, and the 5' end of ND 5 (hereafter referred to as the Brown *et al.* (1982) region after the first author to use this locus for phylogenetic purposes) in 11 species of cercopithecines. This region is bounded by two *Hind*III restriction endonuclease sites in most primates (Brown *et al.*, 1982). We included two representatives each of the *mona*, *cephus*, and *nictitans* species groups as well as a single member each of the *lhoesti*, *neglectus*, and *aethiops* species groups. We sequenced one individual each of *Allenopithecus*, *Miopithecus*, and *Erythrocebus* as well (Table I). We also sequenced the entire 684 base pair cytochrome oxidase subunit II (COII) gene in a subset of the taxa sequenced for the Brown region (Table I) to strengthen molecular dating estimates for key nodes in the phylogeny.

We used flanking primers previously developed in our lab to amplify the Brown and COII regions (Disotell *et al.*, 1992; Wildman, 2000). We purified PCR products using exonuclease I and shrimp alkaline phosphatase (Hanke and Wink, 1994) and cycle sequenced using ABI BigDye chemistry (Applied Biosystems). We cleaned cycle sequencing products using Centri-Sep spin columns (Princeton Separations) and then electrophoresed them on either an ABI Prism 310 or 377 automated DNA analysis system (Applied Biosystems). We sequenced both strands with an array of internal sequencing primers. We assembled contigs using Sequencher 4.1.2 (Gene Codes).

We carried out phylogenetic analyses on the Brown region via bootstrap parsimony (1000 replicates) with heuristic search algorithms with uniform weighting, the stepwise addition and total branch swapping options,

Table I. Species Sequenced for the Mitochondrial Loci Brown and COII

Latin name	Brown	COII
<i>Allenopithecus nigroviridis</i>	✓	✓
<i>Miopithecus talapoin</i>	✓	✓
<i>Erythrocebus patas</i>	✓	✓
<i>Cercopithecus aethiops</i>	✓	✓
<i>Cercopithecus cephus</i>	✓	✓
<i>Cercopithecus ascanius</i>	✓	✓
<i>Cercopithecus mona</i>	✓	
<i>Cercopithecus pogonias</i>	✓	✓
<i>Cercopithecus lhoesti</i>	✓	✓
<i>Cercopithecus nictitans</i>	✓	
<i>Cercopithecus mitis</i>	✓	✓
<i>Cercopithecus neglectus</i>	✓	✓

and a transition/transversion ratio of 12:1 under PAUP* 4.0b8 (Swofford, 2001). We also performed likelihood analyses (100 replicates) with the same parameter settings and the empirical base frequency model. Varying the parameter settings of the two types of analyses did not essentially change the results. Parsimony and likelihood analyses did not have any conflicts in their resolved clades (Fig. 1).

We calculated divergence date estimates for key nodes within the inferred guenon tree as well as for the hominoid-cercopithecoid and cercopithecine-papionin splits relative to the reasonably well accepted calibration points of *Homo-Pan* at 6 mya and *Theropithecus-Papio* at 5 mya (Fig. 2). We compared sequences from the various clades of Old World monkeys and hominoids to a New World monkey sequence to look for evidence of rate heterogeneity in the SRY and TSPY loci (Tosi *et al.*, 2000), the albumin gene (Page and Goodman, 2001), the mitochondrial 12s rRNA (van der Kuyl *et al.*, 1995a), and the combined COII-Brown regions (this study). Only the 12s rRNA data showed evidence of rate heterogeneity. We applied correction factors to the 12s rRNA estimates (Fig. 2).

Since the three nuclear loci showed relatively small amounts of sequence divergence, we made direct extrapolations based on uncorrected average pairwise nucleotide differences (“*p*” values) between members of each lineage in comparison to the calibration points’ pairwise distances (Fig. 2). Earlier studies of protein coding mitochondrial genes demonstrated that third position codon bases show evidence of multiple hit saturation effects (Adkins *et al.*, 1996; Yoder *et al.*, 1996). Therefore, for the combined COII and Brown data set, we made an additional calculation in which we excluded noncoding and third codon position bases from our calculation of uncorrected pairwise distances and their corresponding divergence date estimates (Fig. 2).

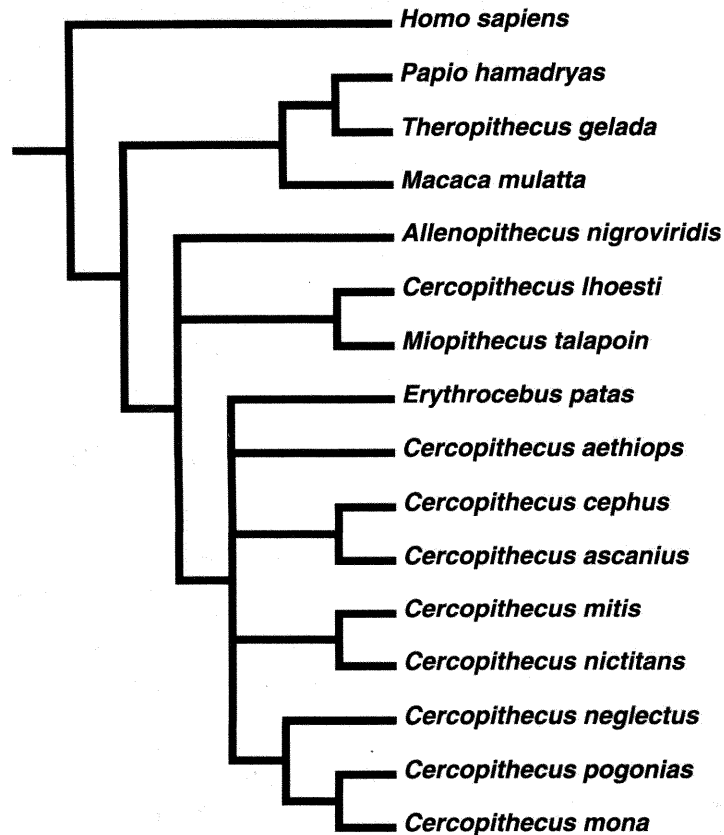


Fig. 1. Consensus tree from our study of the mitochondrial Brown and COII loci. This tree presents nodes consistent among 1000 replicate parsimony bootstrap and 100 replicate likelihood bootstrap analyses of the combined sequence dataset. All nodes have 90% or greater bootstrap support under either likelihood or parsimony analysis.

Results

The mitochondrial phylogeny shows a clade composed of the *Cercopithecus* guenons—excluding *Cercopithecus lhoesti*—and *Erythrocebus patas* (Fig. 1). Basal to this are *Allenopithecus nigroviridis* and a clade containing *Cercopithecus lhoesti* and *Miopithecus talapoin*. All resolved clades are supported by at least a 90% bootstrap measure under either parsimony or likelihood conditions.

The most internally consistent DNA sequence data for date estimation is the COII and Brown combined data excluding third codon positions. Dating analyses (Fig. 2) support a date of divergence between the papionins and cercopithecins between 10.5 and 13.5 mya and an origin for the guenon

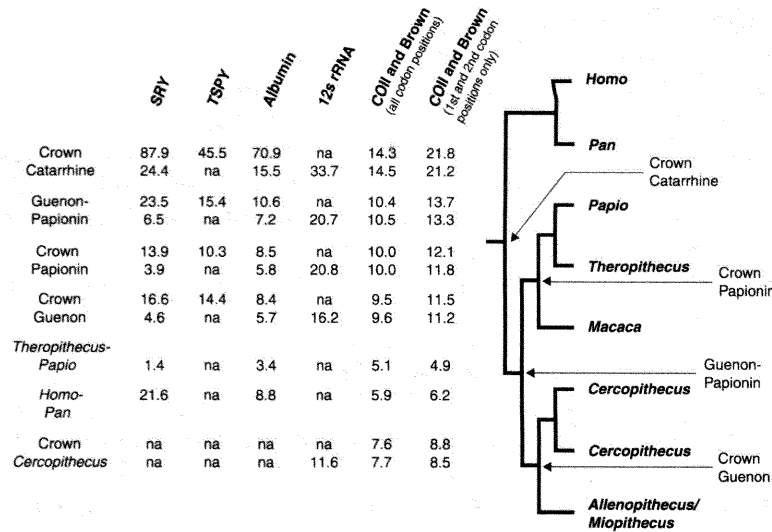


Fig. 2. Date estimates of selected Old World monkey divergences. The table on the left shows estimated times of divergence for six sequence datasets at selected nodes, which are illustrated in the tree diagram to the right. Each node has been estimated from two calibration points. The top number results from calibration on the *Theropithecus-Papio* node set at 5 mya. The bottom number results from calibration on the *Homo-Pan* node set at 6 mya; na indicates that a date could not be calculated for that node based upon the sequences available.

radiation beginning at least 9.5 mya, and more likely over 11 mya. The genus *Cercopithecus* (excluding *C. lhoesti*) most likely began to diversify between 7.5 and 8.5 mya.

Discussion

The mitochondrial phylogeny (Fig. 1) conflicts with the Y chromosome phylogeny of Tosi *et al.* (2002). The most significant conflict concerns the placement of *Miopithecus talapoin*, *Erythrocebus patas*, *Cercopithecus aethiops*, and *C. lhoesti*. While the Y chromosome gene tree supports a clade containing *Erythrocebus patas*, *Cercopithecus aethiops*, and *C. lhoesti*, the mtDNA sequences support a contradictory placement of *Miopithecus talapoin* and *C. lhoesti* as sister taxa.

Incongruence between cytoplasmic and nuclear gene trees is not a new phenomenon. It has been seen before in the primates (*Macaca*: Tosi *et al.*, 2000) as well as many other animal and plant groups (*Drosophila*: Goto and Kimura,

2001; the beetle subgenus *Ohomopterus*: Sota and Vogler, 2001; and the legume subgenus *Glycine*: Doyle *et al.*, 1999). Tosi *et al.* (2000) attributed the incongruence between mtDNA and Y chromosome gene trees in the macaques to lineage sorting of ancestral mtDNA haplotypes with subsequent fixation of different variants in different lineages. Other studies documenting such incongruence have attributed it to hybridization (Sota and Vogler, 2001), lateral gene transfer (Doyle *et al.*, 1999), or inaccurate phylogenetic inference due to homoplasy (Goto and Kimura, 2001).

Furthermore, incongruence could be the result of inaccurate gene tree inference. In either the mtDNA or Y chromosome data sets, one or more of these species could be represented by non-orthologous sequences. Non-orthologous mtDNA sequences are usually nuclear insertions of mitochondrial sequences, or “numts” (Collura and Stewart, 1995; Bensasson *et al.*, 2001). Our laboratory methods, including mitochondrial enrichment of source DNA, the sequencing of multiple clones, and the use of very long PCR fragments, were designed to avoid such problems. None of the sequences analyzed contained premature stop codons in protein coding sequences or insertion or deletion events common in “numts”. Furthermore, the substitution pattern in which third codon position substitutions are most common, followed by first position and then second position substitutions in protein coding sequences, along with the observed transition/transversion bias, all suggest these are true mtDNA sequences. Therefore, the incongruence between mtDNA and Y chromosome is probably not due to nuclear insertions of mitochondrial DNA.

The mtDNA regions sequenced have also been extensively studied in several groups of primates and other mammals (Brown *et al.*, 1982; Disotell *et al.*, 1992; Adkins and Honeycutt, 1994; Adkins *et al.*, 1996; Wildman, 2000). These new guenon sequences show no evidence of excessive homoplasy which could lead to inaccurate phylogenetic inferences (Goto and Kimura, 2001). Without evidence to the contrary, we accept that the mtDNA and Y chromosome evidence produce gene trees which accurately reflect the evolutionary history of those molecules, and we cannot explain away their incongruence by homoplasy. More sequence information from either the mitochondrial genome or the Y chromosome is unlikely to affect the phylogenetic conclusions already drawn from the loci presented here or in Tosi *et al.* (2002).

The most likely reasons for this phylogenetic incongruence are hybridization or lineage sorting. Hybridization may be viewed along a continuum ranging from full hybrid origin of a descendent species from the merger of two parental lines to the introgression of foreign genetic material through occasional cross-species mating. Both ends of this continuum are theoretically possible within the guenons. Hybridization has been observed between many guenon taxa in the wild, and in exceptional circumstances may be common (Detwiler, 2002). Whatever the extent of hybridization, the result will be that variant proportions of the descendent species' genome will derive from more than one lineage.

Lineage sorting may result when an ancestral population is polymorphic for a given locus. If the different alleles at a locus are preferentially preserved in different descendent populations, then the gene tree for the locus will differ from the true species tree. Factors conducive to lineage sorting are large ancestral effective population size and short internodes between speciation events. Both of these situations are likely to have occurred in guenon evolutionary history. In their discussion of mtDNA and Y chromosome discordance in macaques, Tosi *et al.* (2000) have argued that female philopatry and male dispersal leads to an increased effective population size for the mitochondrial genome and a decreased effective population size for the Y chromosome. Thus, they argued that lineage sorting was more likely to have occurred in mtDNA lineages than in Y chromosome lineages.

Incongruence cannot be resolved with two loci. Sampling additional individuals from throughout the species' ranges may help to identify additional alleles and to infer ancient effective population sizes. More importantly, multiple independent autosomal loci are needed. They need to be either of sufficient length or of a rapid enough rate of evolution so that they contain an adequate number of phylogenetically informative sites in order to avoid inaccurate gene tree inferences. Several laboratories, including our own, are attempting to collect such data. The conflicts between the mtDNA and Y chromosomal phylogenies do not affect the divergence date estimates inferred from the mtDNA data presented herein because none of the contested nodes are being estimated (Fig. 2).

While only the mitochondrial 12s rRNA region showed evidence of rate heterogeneity based on the relative rate test, the divergence date estimates derived from the autosomal loci (SRY, TSPY, albumin) vary widely from each other. They also vary within a locus depending on which calibration point is used. We believe this variation, both overestimating and underestimating what are realistic divergence dates, is most likely caused by the small number of substitutions observed between most lineages. For instance, there are only three, 18, and 13 substitutions observed between *Papio* and *Theropithecus* for the SRY, TSP, and albumin loci, respectively. Extrapolating from such small amounts of change is likely to yield variable results. Much longer sequences of non-coding nuclear loci will be needed to match the levels of variation found in the mitochondrial genome. For comparison, the combined COII and Brown region, which is shorter than either the albumin or TSPY regions sequenced, has 123 substitutions between *Papio* and *Theropithecus*.

The combined COII and Brown sequences underestimate the actual times for the older divergences. This underestimate is caused by the well-characterized phenomenon of multiple hit saturation in rapidly evolving genes. When the most variable third codon positions are removed from the analysis, reasonable divergence estimates emerge. These dates are older than those proposed by paleontologists. Given the extreme paucity of fossil evidence, and our inability to infer rates of morphological change, the dates for the key

cercopithecine divergences based on the first and second codon positions of the COII and Brown regions may indeed be a more accurate reflection of reality.

Summary

While the phylogenetic inferences based on either the mitochondrial or Y chromosome data presented in this volume may not fully represent the species' phylogeny, the dates of key divergence points can be relatively confidently inferred. These dates are older than those suggested by extrapolation from the fossil record. Multiple independent autosomal loci from several individuals throughout a species' range will need to be sequenced in order to provide a robust phylogeny of the guenons.

ACKNOWLEDGMENTS

This work was supported by NSF grant SBR 9506892 and NIH grant R01 GM60760. We would like to thank Caro-Beth Stewart, Cliff Jolly, Derek Wildman, Paul Telfer, Colleen Noviello, Kirstin Sterner, Anthony Tosi, and Kate Detwiler for assistance, samples, discussion, and access to manuscripts.

References

- Adkins, R. M., and Honeycutt, R. L. 1994. Evolution of the primate cytochrome c oxidase subunit II gene. *J. Mol. Evol.* **38**:215–231.
- Adkins, R. M., Honeycutt, R. L., and Disotell, T. R. 1996. Evolution of eutherian cytochrome c oxidase subunit II: heterogeneous rates of protein evolution and altered interaction with cytochrome c. *Mol. Biol. Evol.* **13**:1393–1404.
- Bensasson, D., Zhang, D.-X., Hartl, D. L., and Hewitt, G. M. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* **16**:314–321.
- Brown, W. M., Prager, E. M., Wang, A., and Wilson, A. C. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* **18**:225–239.
- Collura, R. V., and Stewart, C.-B. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* **378**:485–489.
- Detwiler, K. M. 2002. Hybridization between red-tailed monkeys (*Cercopithecus ascanius*) and blue monkeys (*C. mitis*) in East African forests. In: M. E. Glenn, and M. Cords (eds.), *The Guenons: Diversity and Adaptation in African Monkeys*, pp. 79–97. Kluwer Academic Publishers, New York.
- Disotell, T. R. 1996. The phylogeny of Old World monkeys. *Evol. Anthropol.* **5**:18–24.
- Disotell, T. R. 2000. The molecular systematics of the Cercopithecidae. In: P. F. Whitehead, and C. J. Jolly (eds.), *Old World Monkeys*, pp. 29–56. Cambridge University Press, Cambridge.
- Disotell, T. R., Honeycutt, R. L., and Ruvolo, M. 1992. Mitochondrial DNA phylogeny of the Old World monkey tribe Papionini. *Mol. Biol. Evol.* **9**:1–13.

- Doyle, J. J., Doyle, J. L., and Brown, A. H. 1999. Incongruence in the diploid B-genome species complex of glycine (Leguminosae) revisited: histone H3-D alleles versus chloroplast haplotypes. *Mol. Biol. Evol.* **16**:354–362.
- Goto, S. G., and Kimura, M. T. 2001. Phylogenetic utility of mitochondrial *COI* and nuclear *Gpdh* genes in *Drosophila*. *Mol. Phylogenet. Evol.* **18**:404–422.
- Hanke, M., and Wink, M. 1994. Direct DNA sequencing of PCR-amplified vector inserts following enzymatic degradation of primer and dNTPs. *Biotechniques* **18**:858–860.
- Harris, E. E., and Disotell, T. R. 1998. Nuclear gene trees and the phylogenetic relationships of the mangabeys (Primates: Papionini). *Mol. Biol. Evol.* **15**:892–900.
- Kingdon, J. 1997. *The Kingdon Field Guide to African Mammals*. Academic Press, New York.
- van der Kuyl, A. C., Kuiken, C. L., Dekker, J. T., and Goudsmit, J. 1995a. Phylogeny of African monkeys based upon mitochondrial 12s rRNA sequences. *J. Mol. Evol.* **40**:173–180.
- van der Kuyl, A. C., Kuiken, C. L., Dekker, J. T., Perizonius, W. R., and Goudsmit, J. 1995b. Nuclear counterparts of the cytoplasmic mitochondrial 12s rRNA genes: a problem of ancient DNA and molecular phylogenies. *J. Mol. Evol.* **40**:652–657.
- van der Kuyl, A. C., van Gennep, D. R., Dekker, J. T., and Goudsmit, J. 2000a. Routine DNA analysis based on 12s rRNA gene sequencing as a tool in the management of captive primates. *J. Med. Primatol.* **29**:307–315.
- van der Kuyl, A. C., Dekker, J. T., and Goudsmit, J. 2000b. Primate genus *Miopithecus*: evidence for the existence of species and subspecies of dwarf guenons based on cellular and endogenous viral sequences. *Mol. Phylogenet. Evol.* **14**:403–413.
- Leakey, M. 1988. Fossil evidence for the evolution of the guenons. In: A. Gautier-Hion, F. Bourlière, J.-P. Gautier, and J. Kingdon (eds.), *A Primate Radiation: Evolutionary Biology of the African Guenons*, pp. 7–12. Cambridge University Press, Cambridge.
- Ledbetter, D. H. 1981. *Chromosomal Evolution and Speciation in the Genus Cercopithecus (Primates, Cercopithecinae)*. Ph.D. Thesis, The University of Texas at Austin, Austin, Texas.
- Muleris, M., Couturier, J., and Dutrillaux, B. 1986. Phylogénie chromosomique des Cercopithecoidea. *Mammalia* **50**:38–52.
- Page, S. L., and Goodman, M. 2001. Catarrhine phylogeny: noncoding DNA revidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. *Mol. Phylogenet. Evol.* **18**:14–25.
- Ruvolo, M. 1988. Genetic evolution in the African guenons. In: A. Gautier-Hion, F. Bourlière, J.-P. Gautier, and J. Kingdon (eds.), *A Primate Radiation: Evolutionary Biology of the African Guenons*, pp. 127–139. Cambridge University Press, Cambridge.
- Ruvolo, M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA data sets. *Mol. Biol. Evol.* **14**:248–265.
- Satta, Y., Klein, J., and Takahata, N. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. *Mol. Phylogenet. Evol.* **14**:259–275.
- Schätzl, H. M., Da Costa, M., Taylor, L., Cohen, F. E., and Prusiner, S. B. 1995. Prion protein gene variation among primates. *J. Mol. Biol.* **245**:362–374.
- Shultz, A. H. 1970. The comparative uniformity of the Cercopithecoidea. In: J. R. Napier, and P. H. Napier (eds.), *Old World Monkeys*, pp. 39–51. Academic Press, New York.
- Sota, T., and Vogler, A. P. 2001. Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. *Syst. Biol.* **50**:39–59.
- Swofford, D. L. 2001. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Szalay, F. S., and Delson, E. 1979. *Evolutionary History of the Primates*. Academic Press, New York.
- Tosi, A. J., Morales, J. C., and Melnick, D. J. 2000. Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. *Mol. Phylogenet. Evol.* **17**:133–144.
- Tosi, A. J., Buzzard, P. J., Morales, J. C., and Melnick, D. J. 2002. Y-chromosomal window onto the history of terrestrial adaptation in the Cercopithecini. In: M. E. Glenn, and M. Cords (eds.).

36 MOLECULAR DIVERGENCE AND INCONGRUENCE

- The Guenons: Diversity and Adaptation in African Monkeys*, pp. 15–26. Kluwer Academic Publishers, New York.
- Wildman, D. E. 2000. *Mammalian Zoogeography of the Arabian Peninsula and Horn of Africa with a Focus on the Cladistic Phylogeography of Hamadryas Baboons (Primates: Papio hamadryas)*. Ph.D. Thesis, New York University, New York.
- Yoder, A. D., Vilgalys, R., and Ruvolo, M. 1996. Molecular evolutionary dynamics of cytochrome b in strepsirrhine primates: the phylogenetic significance of third-position transversions. *Mol. Biol. Evol.* **13**:1339–1350.