INSIGHTS INTO RELATIONSHIPS AMONG RODENT LINEAGES BASED ON MITOCHONDRIAL GENOME SEQUENCE DATA

A Dissertation

by

LAURENCE JOHN FRABOTTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2005

Major Subject: Zoology

INSIGHTS INTO RELATIONSHIPS AMONG RODENT LINEAGES BASED ON MITOCHONDRIAL GENOME SEQUENCE DATA

A Dissertation

by

LAURENCE JOHN FRABOTTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Rodney L. Honeycutt
Committee Members,	James B. Woolley
	John W. Bickham
	James R. Manhart
Head of Department,	Vincent M. Cassone

December 2005

Major Subject: Zoology

ABSTRACT

Insights into Relationships among Rodent Lineages Based on Mitochondrial Genome Sequence Data.

(December 2005)

Laurence John Frabotta, B.S.; M.S., California State University, Long Beach

Chair of Advisory Committee: Dr. Rodney L. Honeycutt

This dissertation has two major sections. In Chapter II, complete mitochondrial (mt DNA) genome sequences were used to construct a hypothesis for affinities of most major lineages of rodents that arose quickly in the Eocene and were well established by the end of the Oligocene. Determining the relationships among extant members of such old lineages can be difficult. Two traditional schemes on subordinal classification of rodents have persisted for over a century, dividing rodents into either two or three suborders, with relationships among families or superfamilies remaining problematic. The mtDNA sequences for four new rodent taxa (*Aplodontia*, *Cratogeomys*, *Erethizon*, and *Hystrix*), along with previously published Euarchontoglires taxa, were analyzed under parsimony, likelihood, and Bayesian criteria. Likelihood and Bayesian analyses of the protein-coding genes converged on a single topology that weakly supported rodent monophyly and was significantly better than the parsimony trees. Analysis of the tRNAs failed to recover a monophyletic Rodentia and did not reach convergence on a stationary distribution after fifty million generations. Most relationships hypothesized in the likelihood topology have support from previous data.

Mt tRNAs have been largely ignored with respect to molecular evolution or phylogenetic utility. In Chapter III, the mt tRNAs from 141 mammals were used to refine secondary structure models and examine their molecular evolution. Both H- and L-encoded tRNAs are AT-rich with different %G and GC-skew and a difference in skew between H- and L-strand stems. Proportion of W-C pairs is higher in the H-strand and GU/UG pairs are higher in the L-strand, suggesting increased mismatch compensation in L-strand tRNAs. Among rodents, the number of variable stem base-pairs was nearly 75% of that observed across all mammals combined. Compensatory base changes were present only at divergences of 4% or greater. Neither loop reduction nor an accumulation of deleterious mutations, both suggestive of mutational meltdown (Muller's ratchet), was observed. Mutations associated with human pathologies are correlated only with the coding strand, with H-strand tRNAs being linked to substantially more of these mutations.

DEDICATION

I dedicate this body of work to my mom, who has always encouraged the scientist in me and supported me during my entire journey so far, to the memory of dad, who provided a stable and supportive home during my youth, and to a great pair of siblings, Mike F. and Jaime, and their families. To Judd and Nancy, who have treated me as their own and offered their support beyond measure. To my Grandmas, who have been waiting for this day for such a long time. I hope you're proud. To Hild, who tried to make it until I finished, but had to go. And lastly, to my soulmate and teammate, Colleen, who puts up with me, defends my nature, and inspires me to always keep at it, etc...

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the following people: my advisor, Rodney Honeycutt, for his previous hard work in rodent systematics, his knowledge and passion for systematics and molecular biology; my committee members, John Bickham (Wildlife & Fisheries Science), Jim Woolley (Entomology), and Jim Manhart (Biology) for their input, recommendations, and time. I am greatly indebted to Dr. Rich Broughton (University of Oklahoma Biological Survey) for the use of his lab, equipment, and materials for the shot-gun cloning of the mitochondrial genomes, Paulette (and Franz) Reneau (of the Broughton lab) for assistance in the shot-gun cloning protocol and a comfortable place to stay during my visit to OU, and Dr. Allison Gillaspy and staff of the OUHSC facility for plasmid preps and DNA sequencing. I thank Michael Dickens (Graduate Program in Genetics, TAMU) for modification and use of his PHROG scripts, Dr. Chris Elsik (Animal Science, TAMU) for helpful conversations on this project, Dr. Clare A. Gill (Animal Science, TAMU) for all of her lab assistance during the early portion of this project, and Dr. Marcus Young Owl (Anthropology, CSULB) for serendipitously introducing me to the controversy of rodent monophyly. A number of individuals and institutions provided samples from their collections and museums: John Bickham, Wildlife & Fisheries Sciences, TAMU; C. William Kilpatrick, Zadock Thompson Natural History Collections, University of Vermont; Terry Yates, Museum of Southwestern Biology, University of New Mexico; Duane Schlitter and Sue McLaren, Carnegie Museum of Natural History, Pittsburgh; Robert Baker, Museum of Texas Tech University; B. Hennie Erasmus, MacGregor Museum, South Africa). I thank Kathy

Dunn (Dalhousie University, Halifax, Nova Scotia), Matt Yoder (Entomology, TAMU), and Joe Gillespie (Entomology, TAMU) for their input on Chapter III, and Chris Walker and Andy Castillo for their help in sequencing the additional rodent tRNA genes. A special thanks to Drs. Anthony Cognato (Entomology, TAMU) and April Harlin-Cognato (Biology, TAMU), for their friendship, home, wine, and appreciation for life outside of B/CS, Texas. I also thank Drs. Nina Caris, Tom McKnight, Ira Greenbaum, and Tonna Harris-Haller (Lower Division Instruction, Biology, TAMU) for providing me with a source of employment and never-ending bemusement throughout much of this project.

Special thanks to my mom, Sally Tolle, who has never wavered in her support of my adventures. To Judd and Nancy Ingram, who have given me so much support, encouragement, and have always showed excitement for my work as if I were their own kid. And lastly, to my love, Colleen, who helped me throughout every aspect of this project, from long nights in the lab, to editing, for cooking many meals, laughing with me when I needed it, and always trying to convince me that mitochondrial genomes *were* interesting.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS viii

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER

I INTRODUCTION 1

II INSIGHTS INTO RELATIONSHIPS AMONG RODENT LINEAGES BASED ON MITOCHONDRIAL GENOME SEQUENCE DATA 20

Introduction 20
Materials and Methods 25
Taxon Sampling 25
Mitochondrial Genome Isolation by Long-PCR 26
Shot-gun Library Construction and Sequencing 27
Sequence Manipulation, Annotation, and Alignment 29
Phylogenetic Analysis 31
Results 36
Mitochondrial Genome Sequences 36
Phylogenetic Analysis 37
Discussion 52
Character Performance and Hypothesis Selection 52
tRNA Genes 56
Relationships Based on Present Dataset 57
Final Remarks 65

CHAPTER Page

III A COMPREHENSIVE ANALYSIS OF THE EVOLUTION OF MAMMALIAN MITOCHONDRIAL tRNA GENES 67

Introduction 67
Materials and Methods 71
Sequence Mining and Multiple Sequence Alignment 71
Scripted Manipulation of Data 73
PCR and Sequencing for Additional Rodent tRNAs 74
Results 75
Sequence Mining and Alignment 75
Additional Rodent tRNA Sequences 76
tRNA Secondary Structure 76
Nucleotide Composition of H- and L-Encoded tRNAs 88
Stem Positions 96
Loop Positions 102
Genome Position and Asymmetrical Mutation Bias 107
Variation Among Closely Related Species 110
Discussion 110
Conclusions 121

IV SUMMARY 123

LITERATURE CITED 128

APPENDIX 1 159

APPENDIX 2 160

APPENDIX 3 165

APPENDIX 4 166

APPENDIX 5 167

VITA 171

LIST OF TABLES

TABLE Page

1.1 Recent rodent families, common names, and number of genera and species 2

1.2 Major classification schemes of Rodentia 3

2.1 Taxa examined in this study. 33

2.2 Partitioned decay indices (by gene) for selected clades 39

2.3 Comparisons of topologies from different analyses using the Kishino-Hasegawa test 50

2.4 Comparisons to topologies from previous studies of higher-level rodent relationships using the Shimodaira-Hasegawa test 51

2.5 Range of *p*-values from relative rates tests (Tajima 1993) for select mitochondrial genes 54

3.1 Cumulative mean base-pair composition for the stem positions of the 22 mammalian mitochondrial tRNAs 100

3.2 Number and proportion of variable base-pair positions by stem region for the H-encoded and L-encoded mitochondrial tRNAs 101

3.3 Mean loop size (in nucleotides) of mammalian mt tRNAs 108

3.4 Number of mutations associated with human pathologies identified per tRNA gene 113

LIST OF FIGURES

FIGURE Page

2.1. Strict consensus of two equally-parsimonious trees recovered from all sites 38

2.2. One of two equally-parsimonious trees recovered from first and second positions only 41

2.3. Maximum-likelihood topology recovered from all codon positions under GTR+Γ+I..... 42

2.4. Maximum-likelihood topology recovered from MCMC analysis of all unambiguously aligned tRNA positions under a DNA-based 4-state model..... 46

2.5. Simplified topologies from previous molecular studies of higher-level relationships in Rodentia..... 47

3.1. Consensus secondary structure diagrams of mitochondrial tRNAs 77

3.2. Observed nucleotide composition bias across all positions, stem positions, and loop positions for mammalian mt tRNA genes..... 89

3.3. Observed nucleotide composition among all sites for each of the 141 taxa in this study. 91

3.4. GC- and AT-skew observed among all sites for the tRNA genes of each DNA strand for several mammalian orders.. 93

3.5. Observed nucleotide composition among stem positions for each of the 141 taxa in this study..... 94

3.6. GC- and AT-skew observed among stem positions for the tRNA genes of each DNA strand for several mammalian orders. 98

3.7. Observed nucleotide composition among all loop positions for each of the 141 taxa in this study 104

3.8. GC- and AT-skew observed among loop positions for the tRNA genes of each DNA strand for several mammalian orders 106

FIGURE Page

3.9. Observed frequencies of C (*A*) and T (*B*) nucleotides across all stem positions for four major mammalian lineages.… 111

CHAPTER I

INTRODUCTION

There is strong evidence of a catastrophic mass extinction at the Cretaceous-Tertiary (K-T) boundary, 65 million years before present (mybp) (Alvarez et al. 1984). Molecular evidence suggests that most of the nearly 20 orders of eutherian mammals were already differentiated by this time, surviving the extinction event and undergoing rapid diversification by filling ecological niches formerly occupied by dinosaurs (Kumar and Hedges 1998). Among eutherian mammals, no other order shows the levels of diversity in morphology, behavior, life history, or geographic distribution as the Rodentia. The order Rodentia is classified into 30+ families (Table 1.1) and represents nearly half of all living mammalian species (Wilson & Reeder 1993; Hartenberger 1998; Nowak 1999). For an overview of the major classification schemes of Rodentia, see Table 1.2. In 1758, Linnaeus classified both rodents and lagomorphs in the order Glires based on the presence of prominent gnawing incisors. This classification was maintained by most workers of the 19th Century (e.g., Cuvier 1817; Brandt 1855). de Blainville (1816) was the first to assign suborders to Glires and identified three suborders of rodents: Fouisseurs, Grimpeurs, and Marcheurs (muroids, sciurids, and hystricognaths, respectively). Brandt (1855), while continuing to recognize the order Glires, followed de Blainville in recognizing three groups of rodents and proposed the names Myomorpha, Sciuromorpha, and Hystricomorpha. Gidley (1912) was the first to

---

This dissertation follows the format and style of Molecular Biology and Evolution.

Table 1.1 Recent rodent families, common names, and number of genera and species

Family	Common Name	Genera/species
Aplodontidae	sewellel or mountain beaver	1/1
Sciuridae	squirrels	51/272
Gliridae (= Myoxidae)	dormice	9/27
Seleviniidae	desert dormouse	1/1
Castoridae	beavers	1/2
Geomyidae	pocket gophers	6/40
Heteromyidae	kangaroo rats & pocket mice	6/60
Anomaluridae	scaley-tailed squirrels	3/7
Pedetidae	springhare	1/1
Muridae	mice, rats, hamsters, voles	293/1310
Spalacidae	blind mole-rats	5/21
Dipodidae	jerboas	14/46
Zapodidae	jumping mice	3/5
Ctenodactylidae	gundis	4/5
Hystricidae	Old World porcupines	3/11
Petromuridae	dassie rat	1/1
Thryonomyidae	cane rats (ricecutters)	1/2
Bathyergidae	African mole-rats	6/17+
Abrocomidae	chinchilla rats	1/5
Agoutidae	pacas	1/2
Capromyidae	Antillean hutias	5(8)/7(26)
Caviidae	cavies	5/17
Chinchillidae	chinchillas & viscachas	3/6
Ctenomyidae	tuco-tucos	1/48
Dasyproctidae	agoutis & acouchis	2/13
Dinomyidae	pacarana	1/1
Echimyidae	spiny rats	19/73
Erethizontidae	New World porcupines	4/21
Hydrochaeridae	capybara	1/1
Myocastoridae	coypu	1/1
Octodontidae	degus & viscacha rats	9/13
	Total	462/2037

Modified from McKenna and Bell (1997) and Nowak (1999) with additions from Mares et al. (2000), Ingram et al. (2004), Norris et al. (2004). Parenthetical numbers for capromyids reflect historical extinctions.

Table 1.2 Major classification schemes of Rodentia

Author	Basis	Division
Linnaeaus (1758)	gnawing incisors	Glires[1]
Cuvier (1817)	presence of clavicle	Claviculata[1]
Brandt (1855)	zygomasseteric apparatus	Myomorpha / Sciuromorpha / Hystricomorpha[1]
Alston (1876)	number of incisors	Simplicidentata[1]
Zittel (1893)	zygomasseteric apparatus	added Protrogomorpha to Brandt's classification[1]
Tullberg (1899)	angle of lower jaw	Hystricognathi / Sciurognathi – divided Hystricognathi into
Wood (1937)	zygomasseteric apparatus	resurrected Protrogomorpha
Lavocat (1951)	angle of lower jaw	added Atypognathes to Tullberg's classification
Schaub (1953)	dentition	divided Hystricomorpha to Pentalophodonta and Nototrogomorpha
Wood (1955)	zygomasseteric apparatus	added Theridiomorpha, Castorimorpha, Caviomorpha, and Bathergomorpha to Brandt's classification - abandoned Protrogomorpha
Wood (1958)	zygomasseteric apparatus	re-recognized Protrogomorpha
Wood (1965)	morpho-grades	Caviomorpha / (Brandt's) Myomorpha / Protrogomorpha + fifteen families *incertae sedis*
Thaler (1966)	dentition	added Glirimorpha and Geomorpha to Brandt's classification
Lavocat (1969)	biogeography	removed Phiomorpha from Hystricomorpha
Bugge (1974)	cranial arteries	Brandt's classification (Myomorpha, Sciuromorpha, Hystricomorpha) plus Wood's Castorimorpha, Protrogomorpha, Caviomorpha, added Erethizontimorpha and Anomaluromorpha
Wood (1975)	cranial anatomy	readopted Tullberg's Sciurognathi / Hystricognathi, added Franimorpha
Chaline & Mein (1979)	paleontological anatomy	added Ctenodactylomorpha to Brandt's general scheme
Graur et al. (1991)	protein sequence	proposed rodent paraphyly
D'Erchia et al. (1996)	mtDNA genomes	supported rodent paraphyly
Reyes et al. (1998)	mtDNA genomes	increased taxon sampling - supported rodent paraphyly
Reyes et al. (2000)	mtDNA genomes	increased taxon sampling - supported rodent paraphyly
Murphy et al (2001a, b)	nuDNA and mtDNA	large data sets - strong evidence for rodent monophyly and Glires
Hudelot et al. (2003)	mtRNA	incorporated RNA-based models - supported rodent monophyly
Marivaux et al. (2004)	fossil dentition	major division: Ctenohystrica / Ischyromyiformes
Reyes et al. (2004)	mtDNA genomes	increased taxon sampling - supported rodent monophyly and Glires

[1]Classified with lagomorphs (rabbits, hares, pikas) in the order Glires prior to Gidley (1912).

recognize a separate order Lagomorpha, suggesting that no characters supported Glires and gliriform adaptations were convergent between the two groups. While some evidence suggests that rodents and lagomorphs are not sister-taxa (Hartenberger 1985; von Koenigswald 1985), an overwhelming amount of other evidence (including most molecular data; see Sullivan and Swofford 1997; Murphy et al. 2001a, 2001b; Hudelot et al. 2002; Reyes et al. 2004) suggests a close affinity between the Rodentia and Lagomorpha, therefore the Glires concept persists (Luckett 1985; Nedbal, Honeycutt, and Schlitter 1996; Reyes et al. 2000; Murphy et al. 2001a, 2001b; de Jong et al. 2003).

Brandt's detailed classification, based primarily on the morphology of the zygomasseteric structure of the skull (zygomatic arch, infraorbital foramen, and masseter muscles), has persisted in the literature to the present. von Zittel (1893) added a fourth suborder, the Protrogomorpha, to Brandt's scheme. The recognition of Protrogomorpha was largely ignored until Wood (1937). Tullberg (1899) considered the angle of mandible a more reliable character than the zygomasseteric structure. Tullberg's classification divided rodents into two suborders: Hystricognathi and Sciurognathi. In the former, Tullberg named two groups: the Hystricomorphi (all hystricomorphous rodents) and Bathyergomorphi (included only the African mole-rats, Bathyergidae that have protrogomorphous skulls). Since that time, nearly all subsequent classifications of rodents have followed either the Brandt or Tullberg schemes, with or without modification (Simpson 1945; Wilson and Reeder 1993; Nowak 1999), although alternative subordinal classifications have been put forth (Lavocat 1951; Wood 1955, 1959, 1965).

Within Rodentia, relationships among major lineages are confounded by conflicting results from analyses of morphological and molecular character sets. Much of the controversy in intraordinal classification of rodents has stemmed from either uncertainties in determining affinities of paleontological material to Recent rodent groups (Wood 1980; Vianey-Liaud 1985) or the placement of taxonomically-ambiguous groups (*e.g.*, Anomaluridae, Aplodontidae, Castoroidea, Ctenodactyloidea, Geomyoidea, Gliridae, and Pedetidae). Over the past century, each of these problematic rodent groups has been placed in numerous positions within the rodent phylogeny.

Taxa in the suborder Sciurognathi retain the symplesiomorphic sciurognathous condition. Hartenberger (1985) characterized Sciurognathi as a "wastebasket" assemblage wrought with convergences, divergences, and parallelisms, and few morphological or molecular data sets provide phylogenetic support for this group (but see Adkins et al. 2001). Among sciurognathous rodents, interfamilial relationships remain largely unresolved. Three rodent taxa (Ctenodactylidae, Anomaluridae, and Pedetidae) are both hystricomorphous (zygomasseteric apparatus) and sciurognathous (jaw angle). Simpson (1945) defined the superfamily Anomaluroidea to include the spiny-tailed squirrels (Anomaluridae) and the springhares (Pedetidae), both sciurognath hystricomorphs. Support for this superfamily has been shown with both morphology (Bugge 1985; Luckett 1985; von Koenigswald 1985) and molecular data (e.g., Montegelard et al. 2002). While their data supported recognition of the Anomaluroidea, Montgelard et al. (2002) were unable to establish this superfamily's affinity to other rodent groups. Others have suggested a close affinity between anomalurids and

ctenodactylids (Fischer and Mossman 1969) or collectively between anomalurids, ctenodactylids, pedetids and the hysticognaths. Recent molecular evidence suggests a sister-taxon relationship between Ctenodactylidae and Hystricognathi (Ctenohystrica *sensu* Huchon, Catzeflis, and Douzery 2000).

There is little paleontological and morphological evidence for the affinity of either Castoridae or Aplodontidae to other living rodent families. The sewellel, or mountain beaver of North America, is the only extant member of Aplodontidae and is protrogomorphous, leading some (Wood 1965; Bugge 1974) to place it in Zittel's Protrogomorpha, while others (e.g., Wood 1955) have placed it within the Sciuromorpha as sister to Sciuridae (Lavocat and Parent 1985; Sarich 1985; Vianey-Liaud 1985; Marivaux, Vianey-Liaud, and Jaeger 2004). Wood (1955) also introduced a new suborder, Castorimorpha, that included the sciuromorphous castoroids, Castoridae (beavers) and extinct Eutypomyidae, while others continued to classify these families as sciuromorphs (Hartenberger 1985). The hystricomorphous-sciurognathous ctenodactyloids of northern Africa and Asia have been placed as the stem rodent group (Korth 1984; Hartenberger 1985), as sister to the Geomyoidea (Hartenberger 1985), or more commonly placed in close relationship to the Hystricognathi (Luckett 1980, 1985; Huchon, Catzeflis, and Douzery 2000; Montegelard 2002; Marivaux, Vianey-Liaud, and Jaeger 2004) either alone or together with the anomalurids or pedetids, as previously stated. Thaler (1966) introduced another suborder, the Geomorpha, to group the Geomyidae (North American pocket gophers), Heteromyidae (pocket mice and kangaroo rats), and the fossil Eomyidae. Currently, this same grouping is recognized as the

superfamily Geomyoidea and has been supported as a natural group by both morphology (e.g., Fahlbusch 1985) and molecular data (Nedbal, Honeycutt, and Schlitter 1996; Matthee and Robinson 1997), but again, their affinity to other major rodent groups is controversial. Wilson (1949), Wood (1955, 1959), and Wahlert (1985) all included the Geomyoidea within the Myomorpha, while others have placed them within the Sciuromorpha (Simpson 1945; Luckett and Parent 1985; Vianey-Liaud 1985). Adding to these debates, recent molecular studies (Murphy et al. 2001a; Eizirik, Murphy, and O'Brien 2001) support novel arrangements of Pedetidae + Muridae and Castoridae + Dipodidae clades.

The Gliridae or Myoxidae (Dormice) have historically been placed within Myomorpha (Simpson 1945; Wahlert 1978; Carleton 1984; Sarich 1985) due to their myomorphous zygomasseteric structure, but more recent studies have shown that the "pseudo-myomorphy" (Vianey-Liaud 1985) of the glirids is derived from an ancestral protrogomorphous condition in contrast to a hystricomorphous-derived condition in true myomorphs. Support for an affinity of the glirids to sciuroids, particularly aplodontids or geomyoids, has been shown with morphological data (Bugge 1971, 1985; Lavocat and Parent 1985), paleontological data (Wood 1980; Flynn, Jacobs, and Lindsay 1985), and molecular data (Nedbal, Honeycutt, and Schlitter 1996; Murphy et al. 2001a; Gibson et al. 2005).

A recent study by Marivaux, Vianey-Liaud, and Jaeger (2004) examined 106 dental characters in 91 Tertiary fossil Glires taxa, with representatives from all major Paleogene groups of rodents (28 families and superfamilies) including both extinct

lineages and stem-group taxa leading to modern lineages. Their analyses recovered a monophyletic Glires and the deep dichotomy within Rodentia first proposed by Luckett and Hartenberger (1985) with a Ctenohystrica clade (*sensu* Huchon, Catzeflis, and Douzery 2000) (Ctenodactylidae, [1]†Chapattimyidae, †Yuomyidae, †Diatomyidae, †Tsaganomydiae, †Baluchimyinae, Hystricognathi) and a newly proposed Ischyromyiformes clade (†Ischryomyoidea, Aplodontoidea, Sciuroidea, †Theridiomorpha, Gliroidea, †Sciuravidae, †Zegdoumyidae, Anomaluridae, Muroidea, Dipodoidea, Geomyoidea, Castoroidea, †Cylindrodontidae). This dichotomy also produces a paraphyletic Sciurognathi.

In addition to these enigmatic rodent groups, the position of two families of porcupines, Erethizontidae and Hystricidae (New World and Old World porcupines, respectively) within Hystricognathi as well as the origins of the New World hystricognaths (Caviomorpha *sensu* Wood 1955) and Old World hystricognaths (Phiomorpha *sensu stricto* Lavocat 1973) has proven difficult. Bugge (1974) believed that the New World porcupines lack a close relationship to either the Caviomorpha or Phiomorpha and proposed a new suborder, Erethizontomorpha,. Wood (1965) proposed the suborder Bathy-Phiomorpha for the Old World hystricomorphs (Bathyergidae, Petromuridae, and Thryonomyidae) excluding the Old World porcupines, as the sole members of Hystricomorpha. Wood (1980, 1985) proposed an independent origin for the caviomorphs from his North American Franimorpha, and Patterson and Wood (1982) suggested a separate invasion of the Hystricidae into Africa. While Wood's proposal is

[1]† denotes extinct taxon

largely rejected by paleontological (Korth 1984, 1994; Wilson 1986) and morphological data (incisor enamel: Martin 1994; skeletal anatomy: Landry 1957; musculature: George 1985; fetal membranes: Luckett 1985; and middle ear anatomy: Lavocat and Parent 1985), the position of the two porcupine families remains problematic, even in light of recent molecular data. In many studies, each of these families is the most divergent lineage in their respective monophyletic group. Nedbal (1995) recovered Hystricidae sister to Caviomorpha in one analysis. Lavocat (1973) placed Hystricidae basal to other hystricomorphs, and Nedbal, Honeycutt, and Schlitter (1996), Huchon and Douzery (2001), and Rowe (2002) could not reject this scenario in favor of the hypothesis of reciprocal monophyly of the Phiomorpha and Hystricomorpha. The reciprocal monophyly hypothesis was later supported by Murphy et al. (2001a). Nedbal, Allard, and Honeycutt (1994) could not statistically reject an alternate tree with Erethizontidae sister to all other Hystricognathi. While not in conflict with the validity of the Hystricognathi as a natural group, this alternate topology requires a biogeographic scenario with two independent South American invasions. Additionally, Rowe (2002) could not reject the hypothesis of Erethizontidae as sister to other caviomorphs in favor of the more accepted hypothesis that nested erethizontids within Caviomorpha (Nedbal, Honeycutt, and Schlitter 1996; Adkins et al. 2001; Huchon and Douzery 2001; Murphy et al. 2001a).

More recently, the monophyly of the order Rodentia has been brought to question. Graur, Hide, and Li (1991) suggested that the guinea-pig-like rodents (Caviomorpha) form an outgroup to a clade containing Primates, Lagomorphs, and all

other Rodentia. Subsequent studies have focused on this issue (Allard, Miyamoto, and Honeycutt 1991; Graur et al. 1992; Hasegawa et al. 1992; Li, Hide, and Graur 1992; Li et al. 1992; Luckett and Hartenberger 1993; Ma et al. 1993; Cao, Adachi, and Hasegawa 1994; Cao et al. 1994; Frye and Hedges 1995). Saccone and colleagues (D'Erchia et al. 1996; Reyes, Pesole, and Saccone 1998; Reyes et al. 2000), in particular, have focused on the use of mitochondrial genome sequences to examine the issue of rodent monophyly, sparking further responses (Cao, Okada, and Hasegawa 1997; Sullivan and Swofford 1997; Huchon, Catzeflis, and Douzery 1999) and D'Erchia et al. (1996) have been criticized for making bold statements (such as rodent polyphyly) based on over-simplified models of sequence evolution. Sullivan and Swofford (1997) showed that a reanalysis of the D'Erchia et al. (1996) data set under parameter-rich models that incorporate among-site rate variation fit the data significantly better and fail to refute rodent monophyly. Philippe (1997) analyzed the D'Erchia et al. (1996) data set with additional mitochondrial genome sequences of the platypus, cat, and blue whale. Using the same analytical methods as D'Erchia et al. (1996), the most-parsimonious tree supported rodent monophyly and Glires. However, this topology was not supported by high bootstrap proportions and the choice of outgroups appeared to have a substantial impact on resulting topologies.

The debate over rodent monophyly continues with the majority of new data supporting both Rodentia and Glires. Recently, some of the largest data sets examined thus far have been applied to the origin and higher-level relationships of eutherian mammals. Cao et al. (2000) examined all available mitochondrial genome sequences for

eutherians (34 species), including 4 rodents (mouse, rat, dormouse, and guinea pig). This expanded data set supported rodent monophyly in contrast to D'Erchia et al. (1996). Liu et al. (2001) used 430 source phylogenies from previous morphological and molecular studies to construct supertrees with reasonable taxonomic samples of rodents. In all their analyses, rodent monophyly was supported and the combined supertree also suggests the clade Glires. Madsen et al. (2001; also Scally et al. 2001) analyzed 8655 nt (nuclear and mitochondrial) from all orders of eutherians under robust ML models. All resulting topologies supported monophyletic Rodentia and Glires clades. Murphy et al. (2001a; also Eizirik, Murphy, and O'Brien 2001) analyzed a separate 9779 nt (nuclear and mitochondrial) data set also supporting the monophyly of these two clades. The data sets of Madsen et al. (2001a) and Murphy et al. (2001a) were combined (16,397 nt) and analyzed under a complex ML model (GTR+Γ+I) with parametric bootstrapping and Bayesian inference (BI) (Murphy et al. 2001b). Both ML and BI analyses produced the same topology (supporting Glires and Rodentia) and nearly all clades were recovered with 100% Bayesian posterior probabilities.

While the majority of new molecular data (particularly nuclear genes) support the monophyly of both Rodentia and Glires, data from complete mitochondrial genome sequences continue to fail in recovering a monophyletic Rodentia with phylogenetic confidence, despite the use of complex likelihood models (Reyes, Pesole, and Saccone 1998; Reyes et al. 2000; Mouchaty et al. 2001; Lin, Waddell, and Penny 2002). This may be an artifact of taxon sampling confounded by the long branch attraction leading to the two available muroids (*Mus* and *Rattus*). The most recent analyses of mammalian

mitochondrial genome sequences included several additional myomorph rodents (*Nannospalax*, *Jaculus*, and *Volemys* (= *Microtus*)) and found strong support (100% posterior probabilities) for monophyletic Rodentia and Glires clades (Reyes et al. 2004).

To date, complete mitochondrial genomes have been published for 109 species of eutherian mammals: Afrosoricida (2), Carnivora (11), Cetartiodactyla (26), Chiroptera (7), Dermoptera (2); Edentata (2), Eulipotyphla (8), Hyracoidea (1), Lagomorpha (4), Macroscelidea (2), Primates (14), Perissodactyla (5), Pholidota (1), Proboscidea (2), Scandentia (1), Sirenia (1), Tubulidentata (1), and Rodentia (9). The nine published rodent taxa include: *Rattus* and *Mus* (Sciurognathi, Myomorpha, Muroidea, Muridae, Murinae), *Volemys* (= *Microtus*; Sciurognathi, Myomorpha, Muroidea, Muridae, Arvicolinae), *Nannospalax* (= *Spalax*; Sciurognathi, Myomorpha, Muroidea, Spalacidae), *Jaculus* (Sciurognathi, Myomorpha, Dipodoidea, Dipodidae), *Myoxus* (= *Glis*; Sciurognathi, Glirimorpha, Gliridae), *Sciurus* (Sciurognathi, Sciuromorpha, Sciuridae), *Thryonomys* (Hystricognathi, Phiomorpha, Thryonomyidae), and *Cavia* (Hystricognathi, Caviomorpha, Cavioidea, Caviidae). The sampling and analyses of rodent mitochondrial genome data have been based on only seven of the 30+ families recognized in the order, with three genomes from the family Muridae. Given that rodents make up nearly half of the diversity of eutherians, a more thorough and balanced taxonomic sampling is needed for a better resolution of the relationships within Rodentia and to potentially enhance mitochondrial DNA support for the Glires concept.

In Chapter II, I report the sequences for the complete mitochondrial genomes of four taxonomically-ambiguous rodent taxa: *Aplodontia rufa* (Sciurognathi,

Protrogomorpha, Aplodontidae), *Cratogeomys castanops* (Sciurognathi, Geomyoidea, Geomyidae), *Erethizon dorsatum* (Hystricognathi, Caviomorpha, Erethizontoidea, Erethizontidae) and *Hystrix africaeaustralis* (Hystricognathi, Hysticomorpha; Hystricidae). Combined with the previously published mitochondrial genomes for rodents, these taxa will provide better taxonomic sampling to address the relationships among several problematic rodent families. While I address the issue of rodent monophyly, the primary goal of this chapter is to explore intraordinal relationships within Rodentia with the use of complete mitochondrial (mtDNA) genome sequences, and to address the inadequacies of previous applications of mitochondrial genomic data sets to this topic. Allard, Honeycutt, and Novacek (1999) point out three major problems with previous studies examining the issue of rodent monophyly: a) the validity of assuming clock-like behavior of molecular data; b) the influence of rate heterogeneity, long branches, and taxon sampling; and c) lack of an appropriate model selection in ML to account for non-random patterns of nucleotide substitution. By utilizing both ML and Bayesian methods under appropriately complex models of sequence evolution, and analyzing an expanded data set, the issues of rate variation, poor taxon sampling, and model adequacy are addressed.

Authors using complete mitochondrial genome sequence data to examine mammalian relationships have focused almost exclusively on the concatenated heavy (H) strand encoded protein-coding "supergene" (COX1, COX2, COX3, ATP6, ATP8, ND1, ND2, ND3, ND4L, ND4, ND5, and cytochrome-*b*) and may include 12S and 16S rRNA gene (small subunit or SSU and large subunit or LSU, respectively) (Janke et al.

1994; Janke, Xu, and Arnason 1997; Xu and Arnason 1994; Xu and Arnason 1996; Arnason, Gullberg, and Janke 1997, 1999, 2004; Arnason et al. 2000, 2002; Reyes et al. 2000; Mouchaty et al. 2001; Arnason and Janke 2002; Reyes et al. 2004; but see Lin, Waddell, and Penny 2002). Individually, each of these genes has been shown to perform well at recovering divergences of a particular age and therefore, the protein-coding supergene may be useful across a wide range of evolutionary time. Slow-evolving genes, such as COX1, COX2, and cytochrome-*b*, have been used for divergence times up to ~ 100 million years before present (mybp) (Irwin, Kocher, and Wilson 1991; Adkins and Honeycutt, 1991; Adkins, Honeycutt, and Disotell 1996), while more rapidly-evolving genes (e.g., ND3 and ND4) have been applied to divergences dating to the Miocene and Oligocene (8-23 and 23-38 mybp, respectively) (Hogan, Davis, and Greenbaum 1997; Engel et al. 1998; Flores-Villela et al. 2000; Frabotta 2002).

The contribution of each gene and each of the three codon positions are therefore examined in the recovered phylogenies. In addition, the recovered phylogenies are statistically compared to recent studies of rodent relationships (D'Erchia et al. 1996; Adkins et al. 2001; DeBry and Sagel 2001; Murphy et al. 2001a; Jow et al. 2002; Montegelard et al. 2002; Reyes et al. 2004; Gibson et al. 2005).

Kumazawa and Nishida (1993) proposed the use of mt tRNAs for deep-level phylogenetic reconstruction among deuterostomes. Animal mt tRNAs show strong structural deviations from the canonical cloverleaf structure of their nuclear counterparts (Wolstenholme et al. 1987; Yokogawa et al. 1991; Janke et al. 1994; Moriya et al. 1994; Steinberg, Gautheret, and Cedergren 1994; Watanabe et al. 1994; Takemoto et al. 1995;

Janke, Xu, and Arnason 1997; Sprinzl et al. 1998; Dörner et al. 2001; Nilsson et al. 2003). Whereas many nuclear tRNAs are identical in taxa as varied as *Mus*, *Bos*, *Homo*, and *Xenopus* (e.g., $tRNA^{Phe}$: Sprinzl 1998), mt tRNAs exhibit relaxation of both secondary and tertiary structural constraints. Kumazawa and Nishida (1993) found up to 75% of all stem-forming base-pairs were variable among five animal taxa, and most changes occurred without immediate compensatory change. The reduction in tertiary constraints has been linked with simpler systems of transcription of mtDNA and recognition of mt tRNAs by enzymes during aminoacylation and protein synthesis (Wilson et al. 1985; Kumawaza et al. 1989, 1991). Because constraints on positions associated with tertiary structure are reduced, variation at these positions may be increased, particularly in the D-stem (Kumazawa and Nishida 1993). Without consideration for covariation, transitions in stem positions of mt tRNAs among five diverse deuterostomes showed a linear increase to saturation at ~25% (estimated divergences up to ~100 mybp) and transversions continued to accumulate linearly without reaching an asymptote (divergences in excess of 600 mybp) (Kumazawa and Nishida 1993). Miya and Nishida (2000) recovered an expected phylogeny for eight teleosts by analyzing the unambiguously aligned stem regions of mt tRNAs when the protein-coding genes (separate or combined) failed to recover this phylogeny.

Despite these compelling arguments for their application to phylogenetic studies, the analysis of mt tRNA data sets is often dismissed, particularly with respect to complete mitochondrial genome sequence data. Notable exceptions include numerous studies of non-mammalian taxa, such as teleost fishes (e.g., Miya and Nishida 2000;

Inoue et al. 2001; Miya, Kawaguchi, and Nishida 2001; Miya et al. 2003), chelonian reptiles (Kumazawa and Nishida 1995, 1999), and squamate reptiles (Kumazawa and Nishida 1995, 1999; Macey and Verma 1997; Macey et al. 1997, 1998, 1999a, 1999b, 2000; Macey, Schulte, and Larson 2000; Kumazawa et al. 1996, 1998, but see Janke et al. 2001). Only two studies to date have focused exclusively on the utility of mt RNA genes (rRNA and tRNA) in examining mammalian relationships (Jow et al. 2002; Hudelot et al. 2003). To explore the phylogenetic utility of mammalian mt tRNAs, the 22 mt tRNAs were aligned based on their secondary structures (shown in Chapter III) and analyzed under both DNA and compensatory RNA models of sequence evolution. The recovered phylogenies are compared to those recovered by the analyses of the protein-coding genes.

In addition to their utility as phylogenetics markers, mt tRNAs have received attention for their association with mitochondrial-linked human pathology (Goto, Nonaka, and Horai 1990; Wallace 1992, 1999; Larsson and Clayton 1995; Helm et al. 2000; Florentz and Sissler 2001; Sissler et al. 2004). To date, 20 mitochondrial disorders have been linked to over 100 point mutations in human mt tRNA genes (Mitomap 2005). To provide a better understanding of the level of variation present in the mt tRNAs of mammals and potential correlations to human pathology, Helm et al. (2000) surveyed the complete mitochondrial genome sequences of 31 mammals, sampling nine (of 17) extant orders from Eutheria, two orders from Metatheria (Marsupialia), and two monotremes. In this survey, Helm et al. (2000) produced robust sequence alignments based on putative secondary structure for the 22 mt tRNA genes,

and generated informative structural diagrams providing typical and consensus data for the 31 mammalian taxa examined.

In Chapter III, I extend the efforts of Helm and colleagues to characterize the structure of mammalian mitochondrial tRNAs. I mined public databases (OGRe: Jameson et al., 2003; NCBI *Entrez* Genome) for currently available mammalian complete mitochondrial genome sequences. I assembled a data set of the 22 tRNAs from 109 mammals, sampling all 17 recognized extant orders of eutherians, five orders of marsupials, and two monotremes. I revised consensus secondary structure diagrams for each of the 22 mt tRNAs of non-rodent mammals, adhering to more stringent constraints on helix and loop formation compared to Helm et al. (2000).

Rodents have been shown to have higher rates of sequence evolution in mitochondrial genes than other mammalian orders (Adkins, Honeycutt, and Disotell 1996; Nedbal, Honeycutt, and Schlitter 1996; but see Gissi et al. 2000). A primary motivation for the work presented here relates to the comparison between rodents and other eutherian mammals. Nedbal, Honeycutt, and Schlitter (1996) noted that mt 12S rRNA showed more variation across various groups of rodents than among outgroup mammals used to examine relationships among families in Rodentia. In addition, Honeycutt et al. (1995) and Adkins, Honeycutt, and Disotell (1996) demonstrated that rodents generally show higher rates of nucleotide substitution in mtDNA compared to non-rodents. In the case of the 12S rRNA gene, alignment was problematic, mainly because indels (predominately in loop regions) were present at a high rate. This implies minimal sequence conservation and makes reliable homology assignments difficult. To

determine whether these increased rates extend to the mt tRNAs, I determined the sequences of 11 clustered tRNAs (regions a (IQM), b (WANCY), and d (HSL): Kumazawa and Nishida 1993) for 32 additional rodent species, sampling 26 of the 32 extant families. These new sequences were combined with data from the nine available rodent mtDNA genomes to produce rodent-specific secondary structure diagrams.

As with all but one of the mitochondrial protein-coding genes (ND6), the majority of mt tRNAs are encoded by the heavy (H)-strand (Arg, Asp, Gly, His, Ile, Leu^{CUN}, Leu^{UUR}, Lys, Met, Phe, Ser^{AGY}, Thr, Trp, and Val), with the remaining eight (Ala, Asn, Cys, Gln, Glu, Pro, Ser^{UCN}, and Tyr) encoded by the light (L)-strand. Therefore unlike the protein-coding or rRNA genes, the mt tRNAs represent a suite of loci with reasonable representation on both the H- and L-strands of the genome, and thus offer an unique opportunity to examine potential differences in the evolution of these structurally similar genes, by comparing differences in nucleotide composition, mutation bias, and the potential effects of the asymmetrical nature of mitochondrial genome replication. Replication of mammalian mitochondrial DNA is asymmetrical with the two strands being synthesized from two distinct and distant replication origins (Clayton 1982). The H-strand origin of replication (O_H) is located in the main non-coding portion of the molecule: the D-loop of the control region. Replication begins with displacement of the parental H-strand by the replication bubble. The replication bubble continues approximately two-thirds (~ 11 kb) around the molecule until the L-strand origin of replication (O_L: located in the WANCY tRNA cluster of the mammalian mt genome) is exposed and L-strand synthesis begins in the opposite direction. Since replication of the

mitochondrial genome is slow (up to 2 hrs) (Clayton 1982), portions of the parental H-strand remain exposed as a single-stranded molecule for up to 80–100 m. During this time, the single-stranded H-strand is prone to mutation by hydrolytic deamination and oxidation, and the H-encoded tRNA genes are likely subject to the same directional mutational pressure observed in the protein-coding genes (Reyes et al. 1998; Bielawski and Gold 2002; Faith and Pollock 2003; Gibson et al. 2005). The purpose of Chapter III is to better characterize a comprehensive set of functional mammalian mt tRNA genes with emphasis on 1) variation and revision of secondary structure models; 2) nucleotide composition; 3) base-pair composition; 4) variation in stem and loop size among mammals; and 5) the potential effects of genomic position and the duration of single-strandedness on these features.

CHAPTER II

INSIGHTS INTO RELATIONSHIPS AMONG RODENT LINEAGES BASED ON MITOCHONDRIAL GENOME SEQUENCE DATA

INTRODUCTION

The order Rodentia is classified into 30+ families and represents nearly half of all living species of mammals (Wilson & Reeder 1993; Hartenberger 1998; Nowak 1999). Among eutherian mammals, no other order shows the levels of diversity in morphology, behavior, life history, or geographic distribution. In 1758, Linnaeus classified both rodents and lagomorphs in the order Glires based on the presence of prominent gnawing incisors. Two major schemes for the subdivision of major groups of rodents were proposed during the 19th Century, despite the continued recognition of Glires (Brandt 1855; Tullberg 1899). Brandt (1855) proposed the names Myomorpha, Sciuromorpha, and Hystricomorpha. Brandt's detailed classification was based primarily on the morphology of the zygomasseteric structure of the skull (zygomatic arch, infraorbital foramen, and origin and insertion of the masseter muscles). Tullberg (1899) considered the angle of mandible a more reliable character than the zygomasseteric structure. His resulting classification divided rodents into two suborders: Hystricognathi and Sciurognathi. In the former, Tullberg named two groups: the Hystricomorphi (all hystricomorphous rodents) and Bathyergomorphi (included only the protrogomorphous African mole-rats, Bathyergidae). Since that time, alternative subordinal classifications

have been put forth (Lavocat 1951; Wood 1955, 1959, 1965), but nearly all subsequent classifications of rodents have followed either the Brandt or Tullberg schemes, with or without modification (Simpson 1945; Wilson and Reeder 1993; Nowak 1999). Gidley (1912) was the first to recognize a separate order Lagomorpha, suggesting that no characters supported Glires and that gliriform adaptations were convergent between the two groups. While some evidence suggests that rodents and lagomorphs are not sister-taxa (Hartenberger 1985; von Koenigswald 1985), an overwhelming amount of other evidence, including most molecular data, suggests a close affinity between the Rodentia and Lagomorpha, therefore the Glires concept has persisted (Luckett 1985; Nedbal, Honeycutt, and Schlitter 1996; Reyes et al. 2000; Madsen et al. 2001; Murphy et al. 2001a; 2001b; Amrine-Madsen et al. 2003; de Jong et al. 2003; Reyes et al. 2004; Gibson et al. 2005).

Both nuclear (Madsen et al. 2001; Murphy et al. 2001a; 2001b; Amrine-Madsen et al. 2003; de Jong et al. 2003) and recent mitochondrial (Murphy et al. 2001a, 2001b; Hudelot et al. 2003; Reyes et al. 2004; Gibson et al. 2005) sequence data place Glires and Euarchonta (Dermoptera: colugos or flying lemurs; Scandentia, tree shrews; Primates) within Euarchontoglires. Nuclear data consistently place Euarchontoglires, in turn, as sister to Laurasiatheria, a group containing Cetartiodactyla (artiodactyls and cetaceans), Perissodactyla, Carnivora, Pholidota (pangolins), Eulipotyphla (true insectivores), and Chiroptera (bats) in a large Northern Hemisphere clade, Boreoeutheria. In some recent mitochondrial analyses, arrangements depicted by nuclear data are interrupted by the inclusion of xenarthan (armadillos, sloths, and

anteaters) or afrotherian clades (Reyes et al. 2004; Gibson et al. 2005; but see Jow et al. 2002; Hudelot et al. 2003). The superorder Afrotheria is an assemblage of morphologically diverse mammalian orders with Gondwanan origins: Afrosoricida (golden moles and tenrecs), Macroscelidea (elephant shrews), Tubulidentata (aardvarks), and the superordinal group, Paenungulata, containing the orders Sirenia (dugongs and manatees), Proboscidea (elephants), and Hyracoidea (hyraxes). While the monophyly of Paenungulata has been supported for nearly a century (Gregory 1910; Simpson 1945; Novacek 1992), the superordinal position of the other afrotherian orders has been historically problematic (Gregory 1910; McKenna 1975; Szalay 1977; Butler 1988). The monophyly of Afrotheria and the division of Afrotheria into two clades: paenungulate and non-paenungulate afrotherians (Afroinsectiphillia: Waddell, Kishino, and Ota 2001) are well supported by both nuclear and mitochondrial sequence data (Murphy et al. 2001b; Amrine-Madsen et al. 2003; Waddell and Shelley 2003).

For over a decade, monophyly of the order Rodentia has been debated. Graur, Hide, and Li (1991) suggested that the guinea-pig-like rodents (Caviomorpha) form an outgroup to a clade containing Primates, Lagomorpha, and all other members of Rodentia. Subsequent studies have focused on this issue (Allard, Miyamoto, and Honeycutt 1991; Graur et al. 1992; Hasegawa et al. 1992; Li, Hide, and Graur 1992; Li et al. 1992; Luckett and Hartenberger 1993; Ma et al. 1993; Cao, Adachi, and Hasegawa 1994; Cao et al. 1994; Frye and Hedges 1995). Saccone and colleagues (D'Erchia et al. 1996; Reyes, Pesole, and Saccone 1998; Reyes et al. 2000), in particular, used whole mitochondrial genome sequences to examine the issue of rodent monophyly, sparking

further responses related to appropriate treatment of inherent rate heterogeneity in the construction of phylogenetic trees and to issues of taxon sampling (Cao, Okada, and Hasegawa 1997; Sullivan and Swofford 1997; Huchon, Catzeflis, and Douzery 1999). Sullivan and Swofford (1997) demonstrated that a reanalysis of the mitochondrial genomes (from E'Erchia et al. 1996), under parameter-rich models that incorporated among-site rate variation, were more appropriate of patterns of variation in the data and failed to refute rodent monophyly. Philippe (1997) analyzed the D'Erchia et al. (1996) data set with additional mitochondrial genome sequences of the platypus, cat, and blue whale. Using the same analytical methods as D'Erchia et al. (1996), the most-parsimonious tree supported rodent monophyly and Glires. However, these clades were not supported by high bootstrap proportions, and the choice of outgroups appeared to have a substantial impact on resulting topologies.

While the majority of new molecular data (particularly nuclear genes) support the monophyly of both Rodentia and Glires, data from complete mitochondrial genome sequences have (until recently) failed to recover a monophyletic Rodentia with high levels of support in terms of bootstrap values, despite the use of complex likelihood models (Reyes, Pesole, and Saccone 1998; Reyes et al. 2000; Mouchaty et al. 2001; Lin et al. 2002). This may be an artifact of taxon sampling confounded by long branch effects leading to the muroid lineage, represented by *Mus* and *Rattus*. More recent analyses of mammalian mitochondrial genome sequences included additional dipodoid (jerboa) and muroid (blind mole-rat and vole) rodents and found strong support (100%

posterior probabilities) for monophyly of both Rodentia and Glires (Hudelot et al. 2002; Reyes et al. 2004).

Within Rodentia, relationships among major lineages are confounded by conflicting results from analyses of morphological and molecular character sets. Much of the controversy in intraordinal classification of rodents stems from either the assignment of paleontological material and determination of affinities to Recent rodent groups (Wood 1980; Vianey-Liaud 1985) or the placement of taxonomically-ambiguous groups (*e.g.*, Anomaluridae, Aplodontidae, Castoroidea, Ctenodactyloidea, Geomyoidea, Gliridae, and Pedetidae). Over the past century, each of these problematic rodent groups has been placed in numerous positions within the rodent phylogeny. In addition to these enigmatic rodent groups, the position of two families of porcupines, Erethizontidae and Hystricidae (New World and Old World porcupines, respectively) within Hystricognathi, as well as the origins of the New World hystricognaths (Caviomorpha *sensu* Wood 1955) and Old World hystricognaths (Phiomorpha *sensu* Lavocat 1973) has proven difficult.

Here, I report the sequences for the complete mitochondrial genomes of four taxonomically ambiguous rodent taxa: *Aplodontia rufa* (Sciurognathi, Protrogomorpha, Aplodontidae), *Cratogeomys castanops* (Sciurognathi, Geomyoidea, Geomyidae), *Erethizon dorsatum* (Hystricognathi, Caviomorpha, Erethizontoidea, Erethizontidae) and *Hystrix africaeaustralis* (Hystricognathi, Hysticomorpha; Hystricidae). Combined with the previously published mitochondrial genomes for rodents, these taxa provide better taxonomic sampling to address the placement of these problematic rodent families.

While I address the issue of rodent monophyly, the primary goal of this chapter is to explore the intraordinal relationships within Rodentia with the use of complete mitochondrial (mtDNA) genome sequences, and to address the inadequacies of previous applications of mitochondrial genomic data sets. Allard, Honeycutt, and Novacek (1999) pointed out three major problems with previous studies examining the issue of rodent monophyly: a) the validity of assuming clock-like behavior of molecular data; b) the influence of rate heterogeneity, long branches, and taxon sampling; and c) lack of an appropriate model selection in maximum likelihood (ML) analyses to account for non-random patterns of nucleotide substitution. By utilizing both ML and Bayesian methods under appropriately complex models of sequence evolution, and analyzing an expanded data set, the issues of rate variation, poor taxon sampling, and model adequacy are addressed.

MATERIALS AND METHODS

Taxon Sampling

DNA was isolated for the following four taxonomically ambiguous rodent taxa: *Aplodontia rufa* (H2370; Sciurognathi, Protrogomorpha, Aplodontidae), *Cratogeomys castanops* (H110; Sciurognathi, Geomyoidea, Geomyidae), *Erethizon dorsatum* (H5834; Hystricognathi, Caviomorpha, Erethizontoidea, Erethizontidae), and *Hystrix africaeaustralis* (H595; Hystricognathi, Hystricomorpha, Hystricidae). Total genomic DNA was isolated from frozen liver or skeletal muscle by proteinase-K digestion, followed by phenol/chloroform extraction (Sambrook, Fritsch, and Maniatis 1989). In

some cases, purified mtDNA (from CsCl gradients) was available from previous studies (e.g., Nedbal, Honeycutt, and Schlitter 1996).

Mitochondrial Genome Isolation by Long-PCR

Whole mitochondrial genomes were amplified using a long-PCR protocol (Cheng et al. 1994) and highly conserved primers designed to the 3'end of the 16S rRNA gene (S-LA-16S-H and S-LA-16S-L) (Miya and Nishida 2000). Long-PCR was performed on either a Perkin-Elmer 9700 or 2700 thermal cycler in 25 µL reactions containing 2.5 µL of 10X LA-*Taq* PCR Buffer (Takara: Fisher Scientific, Pittsburgh, PA), 3.5 µL dNTP (4mM), 2.0 µL of each primer (5 µM), 0.25 µL of 2.5 units of LA-*Taq* polymerase (Takara), and 11.75 µL of ddH_2O. Reaction conditions for "shuttle PCR" included an initial 2 m denaturation at 98ºC, followed by 30 cycles with denaturation at 98ºC for 10 s, annealing and extension combined at the same temperature (68ºC) for 16 m, and a final extension of 19 m. Amplification products were confirmed by electrophoresis (5µl) with a size standard marker (High DNA Mass Ladder: Cat # 10496-016, Invitrogen Life Technologies, Carlsbad, CA) in 1% agarose minigels, stained with ethidium bromide (EtBr), and visualized under UV light. Long-PCR products were excised from agarose gels and purified with QIAEX II gel extraction suspension (QIAGEN, Inc., Valencia, CA) and resuspended in 30 µL of 10 mM Tris-HCl (pH 8.0). To avoid shearing, all template DNA and long-PCR products were handled only with large-bore 20–200 µL pipette tips.

Shot-gun Library Construction and Sequencing

A Sonifier 450A sonicator (Branson Ultrasonics, Danbury CT) was used to shear purified long-PCR products into random fragments. The settings for sonication were: output = 5, duty cycle = constant, timer = hold. The sonication method requires some trial-and-error. Bursts of 5–10 s have been shown to produce random fragments 0.5-5.0 kb in length and one 10 s plus one 5 s burst are typically sufficient for 16-17 kb fragments. During sonication, the samples must be kept on ice to reduce uneven fragment distribution patterns.

The sheared DNAs were end-repaired using a simultaneous fill-in/kinase protocol to prepare the fragments for blunt-end cloning, with the following reaction mixture: 30 µL of DNA, 3.8 µL of 10X kinase buffer, 3.0 µL of 1 mM dNTP, 0.5 µL of 10 mM rATP, 0.2 µL (5U/µL) of Klenow DNA polymerase, 0.2 µL (5U/µL) of T4 DNA polymerase, and 0.5 µL (5U/µL) of T4 polynucleotide kinase. The repaired DNA was resuspended in 15 µL of 10 mM Tris-HCl (pH 8.0) and electrophoresed in 1% agarose with 1 Kb DNA ladder (Cat # 15615-016, Invitrogen Life Technologies) and stained with 0.01% EtBr for 15 m. Smears (from 0.4–4.0 kb) were excised from the agarose under UV light and purified using QIAmp MinElute gel extraction columns (QIAGEN, Inc.) and eluted with 10 µL of EB buffer (10 mM Tris-HCl, pH 8.0).

Plasmids (pGEM-3Z: Promega, Madison, WI) were linearized by digestion with *Sma*I in the following reaction mixture: 5 µL of (1 µg/µL) pGEM, 12U of *Sma*I (Promega) 3 µL of 10X *Sma*I buffer (Promega), 0.3 µL of 100X BSA, and 20.7 µL of ddH_2O, and incubated at 37ºC for 3–4 h. The linearized plasmids were

dephosphorylated with CIAP (calf intestinal alkaline phosphatase) by adding the following to the plasmid digestion: 5 µL (0.05U/µL) of CIAP (Promega), 4 µL of 10X CIAP buffer (Promega), 1 µL dd H_2O, and incubated at 37ºC for 15 m, then 56ºC for 15 m. An additional 5µL of CIAP was added and the incubation cycle was repeated. The phosphatase was inactivated by a final incubation at 65ºC for 30 m. Reactions were cooled and purified with a QIAquick PCR purification kit (QIAGEN, Inc.) and eluted in 45 µL of 10 mM Tris-HCl (pH 8.0).

Purified sheared DNA fragments were ligated into dephosphorylated, linearized pGEM using the LigaFast system (Promega) with the following reaction mixture: 2 µL (100 ng) of prepared vector, 10 µL (~ 80-200 ng) of DNA prep, 12.5 µL of 2X Rapid Ligation buffer, and incubated at 50ºC for 1 m. Three units of ligase were added to the reactions, thoroughly mixed, and centrifuged for 1 m. Ligation reactions were incubated at 4ºC for ~ 12 h, then frozen at -20ºC.

Ligated products were transformed into JM109 chemicompetent *E. coli* (Promega) in the following reaction mixture: 5 µL (10–50 ng) of ligation reaction, 100 µL of cells, 970 µL of SOC, 10 µL of 1M $MgCl_2$, and 20 µL of 1M glucose, mixed gently and incubated on ice for 10 m. Cells were heat-shocked at 42ºC for 45–50 s and returned to ice for 2 m. SOC media (900 mL) was added to each reaction and incubated in a shaker at 37ºC for 60 m, then allowed to stand at T_R for 3 h. 100 µL of cells were spread and grown on 8 cm Amp+ X-Gal LB agar plates, and allowed to grow for 18-24 h. Standard blue-white screening was used to select colonies with inserts. Selected

colonies were transferred into 96-well plates and cultured in 2X LB Amp+ broth with 1% freezing medium for permanent storage.

The plasmids from cell cultures were purified using a 96-well NeXPrep DNA purification kit and standard protocol (DeWalch Life Technologies, Houston, TX). Direct cycle-sequencing was done using the M13F primer and BigDye dye-labeled termination chemistry v3.0 (ABI Applied Biosystems, Inc., Foster City, CA) in the following reaction mixture: 5 µL of plasmid DNA, 1 µL of BigDye, 1.5 µL of 5X SDB buffer, 0.5 µL of 10µM primer, 2 µL of ddH_2O with 25 cycles of 97^oC for 30 secs, 50^oC for five sec, and 60^oC for two min. Excess terminator dye, oligonucleotides, and polymerase were removed by centrifugation at 1500 rpm through a sephadex G-50 matrix in a Multiscreen 96-well plate filter (Millipore, Billerica, MD). All sequencing reactions were analyzed on either an ABI 3700 automated sequencer or MegaBACE 1000 (Amersham Biosciences, Piscataway, NJ).

Sequence Manipulation, Annotation, and Alignment

Raw sequence data were pre-screened using Phrog (M. Dickens, unpublished), a set of Perl scripts that passes each sequence file through the following: 1) Phred (Ewing et al. 1998), a base-calling program that analyzes the peaks (height, spacing) to call bases and assigns a logarithmic quality score "phred score" linked to error probabilities to each base call; 2) phd2fasta, that converts each trace file to FASTA format (Pearson and Lipman 1988); 3) Cross_match (CodonCode, Dedham, MA) against the UniVec database (NCBI) to identify cloning vector sequence; 4) sequence trimming based on

low phred quality scores, or vector matching bases, or removal of sequence due to poor quality, length (< 75 nt), or 100% vector content; 5) Cross_match against *E. coli* to identify contamination from cloning host and removal of any sequences containing *E. coli*; and outputting the resulting sequences to a set of FASTA and phred scores files. Phrog output files for each taxon were imported into Sequencher v4.2–4.5 (GeneCodes, Ann Arbor, MI) and assembled into a single contiguous sequence (contig) against an appropriate scaffold mitochondrial genome (*Sciurus vulgaris*: AJ238588 or *Thryonomys swinderianus*: AJ301644). Using the annotations for the appropriate scaffold, contigs were proofed and genes were annotated by eye: 13 protein-coding genes (NADH subunits 1-6, cytochrome oxidase subunits I-III, ATPase 6 and 8, and cytochrome-*b*), 22 tRNA genes (Phe, Val, Leu^{UUR}, Ile, Gln, Met, Trp, Ala, Asn, Cys, Tyr, Ser^{UCN}, Asp, Lys, Gly, Arg, His, Ser^{AGY}, Leu^{CUN}, Glu, Thr, and Pro), and 2 rRNA genes (small subunit or 12S, and large subunit or 16S). Each protein-coding gene was aligned using the amino acid translation in T-Coffee v2.66 (Poirot et al. 2004; Notredame, Higgins, and Heringa 2000). The CORE (Consistency of Overall Residue Evaluation) index implemented in T-Coffee has been shown to be up to 40% more accurate than other multiple sequence alignment applications (e.g., ClustalW: Thompson, Higgins, and Gibson 1994) in terms of identifying correct blocks within difficult multiple sequence alignments (Notredame, Higgins, and Heringa 2000). T-Coffee had difficulty with the 5–10 3′ terminal residues of some genes. These were edited by eye in MacClade v3.07 (Maddison and Maddison 2002). The tRNA genes were aligned using the structure models presented in Chapter III. For the tRNA genes, regions in which positional homology assessments could not

be determined across all taxa were defined according to structural criteria as in Kjer (1997), and described as regions of alignment ambiguity (RAA), regions of slipped-strand compensation (RSC: Levinson & Gutman 1987), or regions of expansion and contraction (REC) following the methodology of Gillespie (2004); for reviews regarding rRNA sequence alignment see Schultes, Hraber, and LaBean (1999) and Hancock and Vogler (2000). The alignment was annotated with a stem-pairing mask equivalent to that utilized in either the program PHASE v2.0 beta (Jow et al. 2002; Hudelot et al. 2003) or MrBayes 3.1.1 (Ronquist and Huelsenbeck 2003). Ambiguously aligned regions were enclosed within square brackets. The twelve protein-coding genes encoded by the H-strand were concatenated to form a mtDNA protein "supergene" for analysis and converted to a PAUP* executable NEXUS file (Swofford 2002) (Appendix 1). The coding region for *ND6* (L-strand) is generally excluded for analysis purposes due to its heterogenous nucleotide composition and consistently poor phylogenetic performance (Zardoya and Meyer 1996; Miya and Nishida 2000; Mouchaty *et al.* 2001; Shevchuk and Allard 2001).

Phylogenetic Analysis

To reduce computational time and allow for more thorough analyses, the monophyly of Euarchontoglires was assumed and 24 euarchontoglire taxa (13 rodents, 3 lagomorphs, 1 tree shrew, 1 colugo, and 6 primates) and a perissodactyl outgroup (*Equus caballus*) were included in all analyses (Table 2.1). The horse was selected as an outgroup due to relatively short and uniform branch lengths in previous analyses (e.g.,

Gibson et al. 2005). The resulting multiple sequence alignment consisted of 10,926 unambiguous nucleotides.

Maximum-parsimony (MP), maximum-likelihood (ML), and Bayesian inference (BI) methods were used to analyze the data set. Under MP, all analyses were unweighted, with separate analyses for all positions, and with third codon positions excluded to evaluate potential effects of third position homoplasy on phylogeny reconstruction. In all cases, I used heuristic searches starting from a random tree, with 10,000 random addition sequences, and the following settings: TBR branch-swapping (Swofford and Begle 1993), saving all minimum length trees (MULPARS), zero-length branches collapsed to yield polytomies, and without the steepest descent option. Bootstrap proportions (BP) and decay indices (DI) were calculated as estimates of nodal support. Bootstrap analyses were done on all resulting trees using 1000 replicates, each with 10 random addition sequences and TBR branch-swapping. Both overall decay indices and partitioned decay indices (by gene: pDI) were calculated using TreeRot v.2b (Sorenson 1999) for the MP trees from each analysis.

Previous ML analyses of rodent mitochondrial genome sequences have used arbitrarily selected models of nucleotide substitution (e.g., HKY85: Cao *et al.* 1997, 1998; unnamed model: Reyes, Pesole, and Saccone 1998; TN93+Γ: Mouchaty *et al.* 2001). To determine the appropriate model of evolution for maximum-likelihood (ML) and Bayesian (BI) analyses, I used the ModelTest v3.7 (Posada and Crandall 1998; Posada and Buckley 2004) to statistically estimate the most-appropriate ML model for the given data set under the Akaike and Bayesian information criteria (AIC and BIC,

Table 2.1 Taxa examined in this study

Taxon		OGRe Abbrev.	Common Name	Accession
Outgroup	*Equus caballus*	EQUCAB	horse	X79547
Ingroup				
Dermoptera	*Cynocephalus variegatus*	CYNVAR	Malayan flying lemur	AJ428849
Primates	*Cebus albifrons*	CEBALB	white-fronted capuchin	AJ309866
	Gorilla gorilla	GORGOR	gorilla	D38114
	Homo sapiens	HOMSAP	human	AF347015
	Pan troglodytes	PANTRO	chimpanzee	D38113
	Papio hamadryas	PAPHAM	baboon	Y18001
	Pongo pygmaeus	PONPYG	orangutan	D38115
Scandentia	*Tupaia belangeri*	TUPBEL	northern tree shrew	AF217811
Lagomorpha	*Lepus europaeus*	LEPEUR	European hare	AJ421471
	Ochotona collaris	OCHCOL	pika	AF348080
	Oryctolagus cuniculus	ORYCUN	rabbit	AJ001588
Rodentia	*Cavia porcellus*	CAVPOR	domestic guinea pig	AJ222767
	Jaculus jaculus	JACJAC	lesser Egyptian jerboa	AJ416890
	Mus musculus	MUSMUS	house mouse	AY172335
	Myoxus glis	MYOGLI	fat dormouse	AJ001562
	Nannospalax ehrenbergi	NANEHR	Ehrenberg's mole-rat	AJ416891
	Rattus norvegicus	RATNOR	Norway rat	X14848
	Sciurus vulgaris	SCIVUL	Eurasian red squirrel	AJ238588
	Thryonomys swinderianus	THRSWI	greater cane rat	AJ301644
	Volemys (Microtus) kikuchii	VOLKIK	Taiwanese vole	AF348082
New Taxa[1]	*Aplodontia rufa*	APLRUF[1]	Sewellel or mountain beaver	H2370[1]
	Cratogeomys castanops	CRACAS[1]	yellow-faced pocket gopher	H110[1]
	Erethizon dorsatum	EREDOR[1]	North American porcupine	H5834[1]
	Hystrix africaeaustralis	HYSAFR[1]	South African porcupine	H595[1]

[1] For new rodent taxa, abbreviations are in the OGRe database format (Jameson et al. 2003) and accessions represent voucher numbers from the tissue collection of Rodney Honeycutt, Department of Wildlife & Fisheries Sciences, Texas A&M University.

respectively). ML analyses were performed using the model parameters estimated by the BIC and the following parameters: heuristic search, starting tree estimated by stepwise addition, 100 random addition sequences, and TBR branch-swapping. Non-parametric bootstrap proportions were estimated using 100 replicates and the ‘‘fast’’ stepwise addition option.

Bayesian inference (BI) methods use Markov chain processes and Monte Carlo simulations to estimate the posterior probabilities (PP) of a particular phylogenetic topology (and its nodes), given a particular data set and model of nucleotide evolution. Through a series of Markov chain iterations, the posterior probability of a node or tree of interest is estimated by the number of times that it is visited during the iterative analysis (Huelsenbeck *et al.* 2001). For BI, I used MrBayes v3.1.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) to estimate phylogenies and calculate posterior probabilities. For all BI runs, a flat Dirichlet probability density was used to estimate the priors for the following model parameters: nucleotide substitution matrix (Revmatpr), the stationary nucleotide frequencies (Statefreqpr), the alpha (α) shape parameter of the gamma distribution for rate variation (Shapepr), and the proportion of invariant sites (Pinvarpr). The flat Dirichlet distribution has recently been shown to perform better than a uniform distribution in specifying priors for these parameters (Zwickl and Holder 2005). Uniform priors were used for the topology (Topologypr) to allow equal probability for all distinct topologies and an exponential prior was used for branch lengths (Brlenspr) to allow for branch lengths from zero to infinity. I used three separate schemes to estimate run parameters: a) across all positions combined (as with ML in

PAUP*), b) all codon positions unlinked, and c) each gene unlinked (with all codon positions estimated together). The current implementation of MrBayes (v3.1.1) runs two independent Markov chain Monte Carlo (MCMC) simulations per analysis, each with four chains by default: one chain is "cold" and the other three are "heated" based on the Temp setting. MrBayes uses an incremental heating strategy where the posterior probability of each chain (i) is raised to the power of $1/(1 + i\lambda)$ with λ as the Temp setting (default = 0.20000). I ran separate simulations of 2×10^6 generations for three heating schemes (Temp = 0.20000, 0.30000, and 0.40000).

The MP, ML, and BI (majority-rule consensus) trees were compared using the S-H test (Shimodaira and Hasegawa 1999) in PAUP*. K-H tests (Kishino and Hasegawa 1989) were used to evaluate alternate hypotheses for the phylogenetics of Rodentia (Adkins et al. 2001; Murphy et al. 2001a; Hudelot et al. 2003; Reyes et al. 2004; Gibson et al. 2005). The original study that refuted rodent monophyly based on complete mitochondrial genome sequences (D'Erchia et al. 1996) was not considered due to their limited taxonomic sampling (*Mus*, *Rattus*, and *Cavia* only).

Additionally, the 22 tRNA genes were analyzed separately under MP and BI, using both 4-state DNA (GTR) and base-pair (Schöniger and von Haeseler 1994) RNA based models of sequence evolution. The 16-state model of Schöniger and von Haeseler (1994) considers all possible base-pairs: the four Watson-Crick pairs (AU, UA, CG, and GC), the stable GU and UG intermediates, and the additional non-pairing mismatches (AA, AC, CA, GG, AG, GA, CU, UC, AA, CC, GG, UU). Since a number of mismatches occur in high frequency in the helical portions of the mitochondrial tRNAs

(see Chapter III), an evolutionary model that considers the frequency of each mismatch class seems appropriate for the analysis of this dataset. The resulting topologies from the analyses of the tRNA genes were compared with those recovered for the protein-coding genes and from previous analyses of mt RNA genes (Hudelot et al. 2003).

RESULTS

Mitochondrial Genome Sequences

Structurally, the four new mitochondrial genomes described here were unremarkable. Each showed the typical vertebrate complement of 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes, a variable length control region, and an identifiable origin of L-strand replication (Ori_L) within the WANCY tRNA cluster, with no deviations from the mammalian mtDNA gene order. Several of the protein-coding genes showed unambiguous size variation compared to the other taxa examined: cytochrome oxidase I of *Aplodontia* had a codon insertion at nucleotide (nt) 460, and cytochrome II of *Cratogeomys* had a codon insertion at nt 631. The ND5 gene showed a number of codon indels: *Aplodontia* had a deletion of two codons at nt 625 and nt 628, *Cratogeomys* had a codon insertion at nt 52 and a deletion at nt 622, and *Hystrix* had two codon deletions (nt 37 and 49). The ND1 gene of *Cratogeomys* had an independent insertion at the second codon position, coding for a different amino acid residue than the second codon insertion found in muroid rodents (*Mus*, *Rattus*, *Nannospalax*, and *Volemys* (*Microtus*).

Phylogenetic Analysis

Protein-coding Genes

Parsimony

The maximum parsimony (MP) analysis of all codon positions recovered two equally-parsimonious trees (TL = 39,987; CI = 0.325; RI = 0.327; RC = 0.106) that differed only in the placement of the tree shrew *Tupaia belangeri* (Scandentia) as sister to Lagomorpha or sister to Glires (= Lagomorpha + Rodentia). The MP analysis of all positions recovered a monophyletic Rodentia (BP = 74%, DI = 16) with the strongest partitioned decay indices from ND4 and ND5 (pDI = 18 and 23.5, respectively) and recovery of several major groups: Myomorpha (BP = 77, DI = 28, COI pDI = 23, ND5 pDI = 11.5), Hystricognathi (BP = 87, DI = 28, ND5 pDI = 32.5), Caviomorpha (BP = 55, DI = 8, ND4 pDI = 9), and a "Sciuroid" clade (Sciuridae, Aplodontidae, Geomyoidea, and Gliridae: BP = 64, DI = 23, ND4 pDI = 14) (Fig. 2.1). Partitioned decay indices for major clades are shown in Table 2.2.

Parsimony analysis of first and second positions also recovered two equally-parsimonious trees with slightly improved character consistency (TL = 13,691; CI = 0.408; RI = 0.405; RC = 0.106). While one of the two trees was identical to the MP trees recovered from the analysis of all sites with Scandentia sister to a monophyletic Glires (BP = 53), the second tree was markedly different, with Scandentia sister to all other Euarchontoglires, and Lagomorpha sister to a clade of Primates + Dermoptera + Rodentia. Within this clade, a muroid + dipodoid clade was separated from other rodents and sister to a Dermoptera + Primates clade, with the remaining rodents

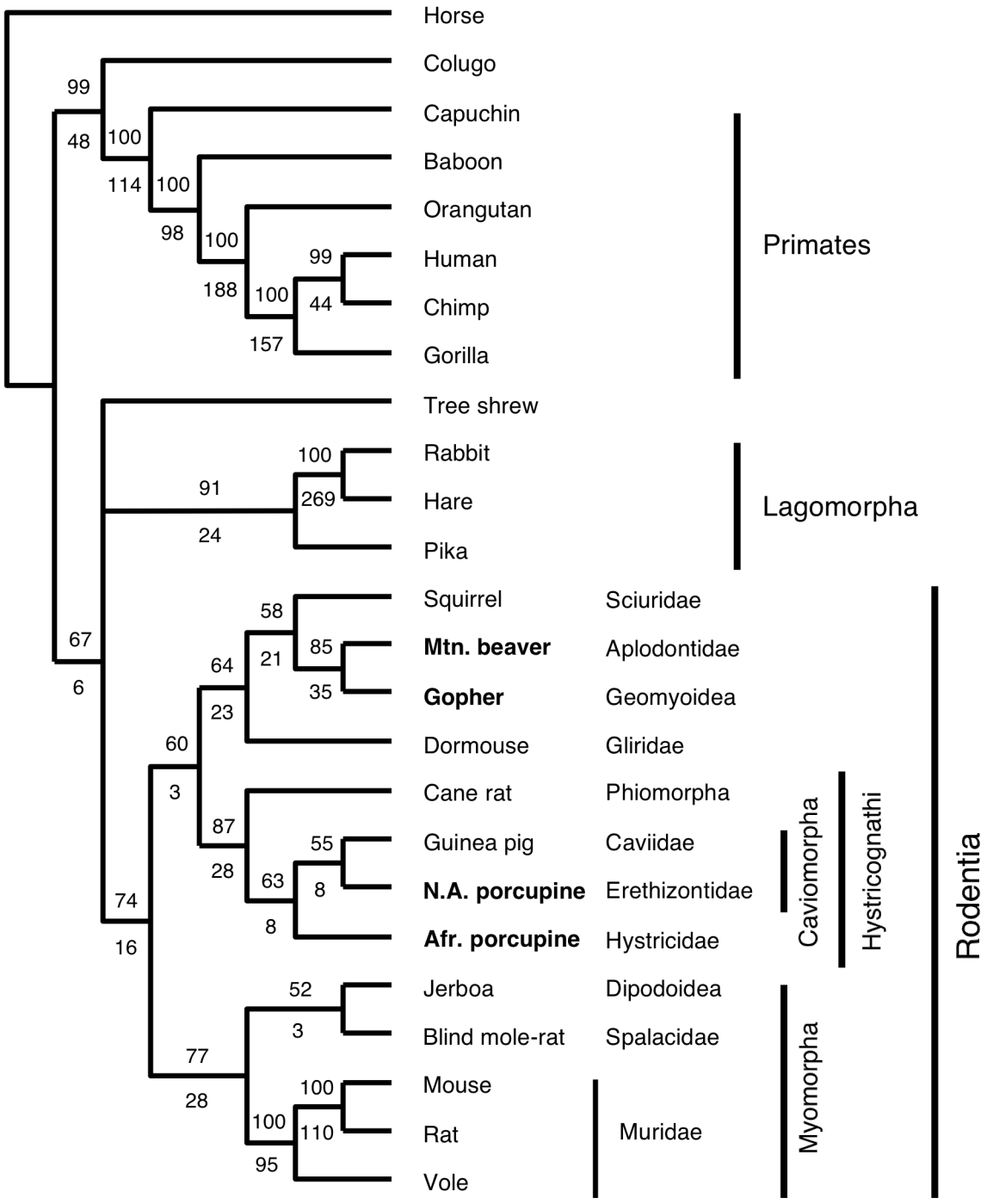


FIG. 2.1. —Strict consensus of two equally-parsimonious trees recovered from all sites (10,926 nucleotides; TL = 39,987; CI = 0.325; RI = 0.327; RC = 0.106). Values indicates bootstrap proportions (above branches) and decay indices (below branches).

Table 2.2 Partitioned decay indices (by gene) for selected clades

Clade	Total	ATP6	ATP8	COI	COII	COIII	Cyt-*b*	ND1	ND2	ND3	ND4L	ND4	ND5
Glires + Scandentia	6	4.5	0	5	-5	-3	3	2.5	-12	-7.5	-1	7	12.5
Rodentia	16	3.5	0	3	-5	-2	6	-2.5	-23	-9.5	4	18	23.5
Hystricognathi	28	4.5	1	6	-4	7	-8	-8.5	-23	-11.5	5	27	32.5
Hystricidae + Caviomorpha	8	2.5	0	2	-5	-7	3	1.5	-19	-8.5	-1	16	23.5
Caviomorpha	8	5.5	0	2	2	4	-2	0.5	-3	-1.5	0	9	-8.5
Muroidea + Dipodoidea	28	-16.5	5	23	-7	6	-2	14.5	-5	-0.5	0	-1	11.5
Muridae	95	-0.5	3	10	1	1	9	13.5	19	10.5	4	10	14.5
Lagomorpha	24	8.5	6	-15	-11	-7	8	12.5	7	1.5	6	-1	8.5
Dermoptera + Primates	48	6.5	2	-1	14	2	7	4.5	3	4.5	-8	1	12.5
Primates	114	3.5	4	12	-2	2	9	17.5	20	3.5	-4	11	37.5
OW monkeys + apes	98	3.5	5	2	24	-9	1	11.5	14	15.5	5	14	11.5
Hominoidea	188	8.8	4.3	16	21.3	16.3	10.3	11.8	23.7	6.8	1	24.3	43.2

forming a clade with the following subclades (Sciuridae: BP = 62, DI = 7 (Aplodontidae + Geomyoidea: BP = 84, DI = 14)) and (Gliridae: BP < 50, DI = 1 (Hystricognathi: BP = 99, DI = 29)) (Fig. 2.2). The fine structure for relationships among the Hysticognathi differed from the trees recovered by the analysis of all positions in two ways: 1) placement of hystricids with New World forms and 2) a sister-group relationship between hystricids and erethizontids. None of the relationships recovered from tree 2 (Fig. 2.2) were supported with bootstrap proportions > 50%.

Maximum Likelihood

Based on the Bayesian Information Criterion (BIC), the general time-reversible model (Yang, 1994), corrected for among-site rate variation using the discrete gamma distribution and a proportion of invariable sites (GTR+Γ+I), was significantly better than all simpler models (ModelTest v3.7: Posada and Buckley 2004; BIC = 335,591). ML analyses of all positions recovered a single topology (-*ln* L = 154,266) with Scandentia sister to a monophyletic Glires (BP = 82) and monophyly of Rodentia (BP < 50), Hystricognathi (BP = 98), and Caviomorpha (BP = 81) (Fig. 2.3). Maximum-likelihood analysis of first and second positions only recovered a single tree with an identical topology to the tree from all positions with nearly identical bootstrap proportions (-*ln* L = 67,920) (not shown).

To determine if the present taxon sampling was sufficient to overcome model inadequacy (Sullivan and Swofford 1997), the data set was analyzed under the simplistic

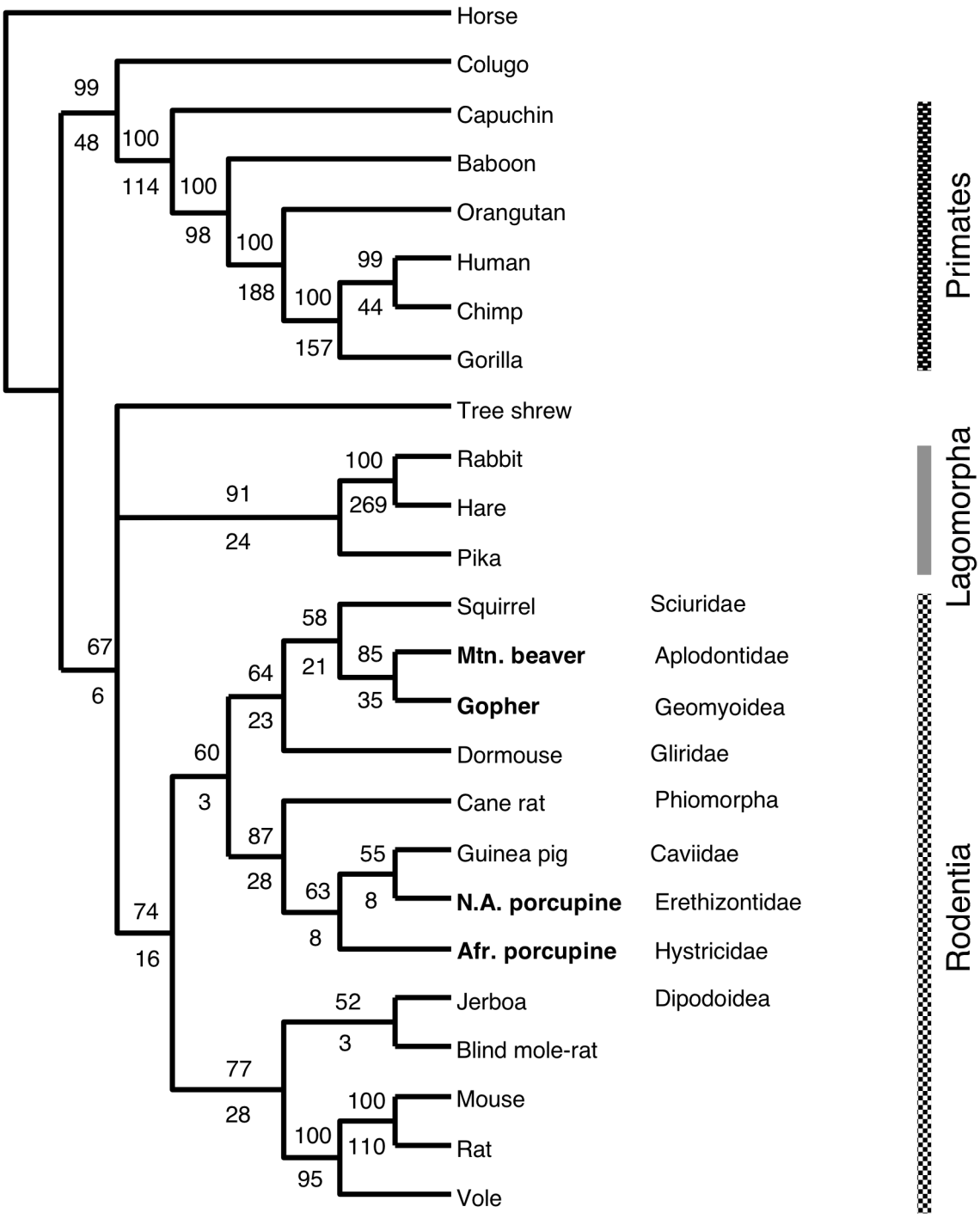


FIG. 2.2. —One of two equally-parsimonious trees recovered from first and second positions only (7,284 nucleotides; TL = 13,691; CI = 0.408; RI = 0.405; RC = 0.106). Values indicates bootstrap proportions (above branches) and decay indices (below branches).

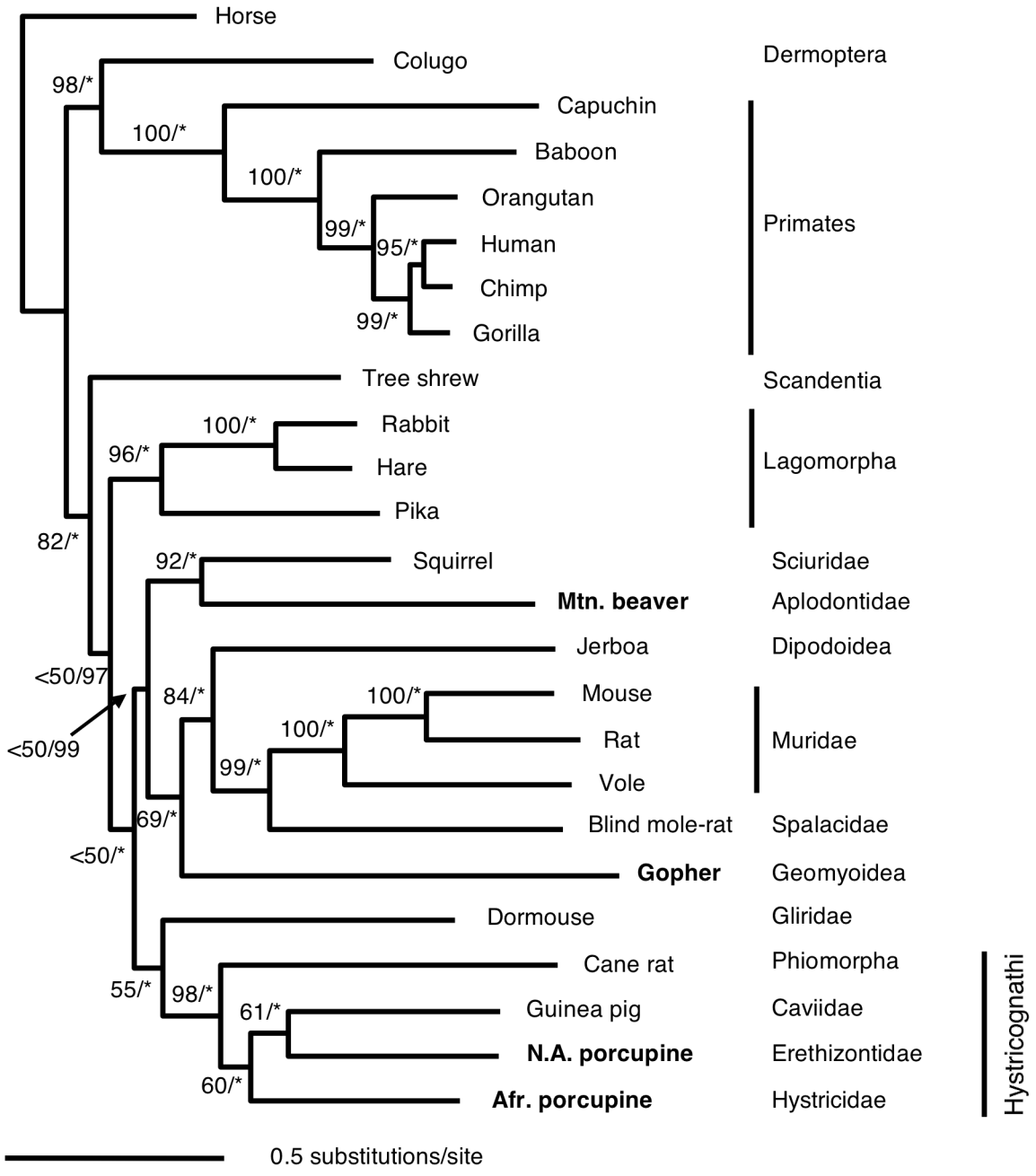


FIG. 2.3. —Maximum-likelihood topology recovered from all codon positions under GTR+Γ+I (-*ln* L = 154,267). Values above or below branches indicate non-parametric bootstrap proportions and Bayesian posterior probabilities, respectively (asterisks indicate PP of 100).

JC69 (Jukes and Cantor 1969) with equal base frequencies and equal substitution rates. While the topologies (not shown) differed from each other, and from more appropriately modeled analyses, ML analyses of all positions and first and second positions only, recovered monophyletic Glires and Rodentia, albeit with low bootstrap support (< 50 and <50–69, respectively), suggesting that taxon sampling alone has tremendous utility in determining phylogenetic relationships, particularly in diverse groups such as rodents.

Bayesian Analyses

At 2 x 10^6 generations, the standard deviation of split frequencies (estimated between the two runs per simulation) was sufficiently close to zero to assume that two runs had converged onto a stationary likelihood distribution. MCMC runs of all nucleotide positions converged within 100,000 generations under all three model estimating schemes: 1) all positions combined, 2) unlinked codon positions, and 3) unlinked genes. Regardless of the modeling scheme, only two nodes were recovered with less than 100% posterior probabilities (PP): 1) the node defining Glires (PP = 97–98%) and 2) the node uniting a Sciuridae + Aplodontidae clade with a (Geomyoid (Muroid + Dipodoid)) clade (PP = 99). While the partitioned model estimates improved the likelihood scores obtained, all methods converged on the identical topology recovered under ML. As partitioned ML analyses are not presently available in PAUP*, the model parameters estimated under GTR+Γ+I across all positions appear to model this data set adequately.

tRNA Genes

Unweighted MP analysis of the unambiguously aligned positions of the mt tRNA genes recovered two equally-parsimonious trees (not shown) that failed to recover either a monophyletic Glires or Rodentia, but some major groups were recovered: muroids plus dipodoids (*Jaculus* (*Nannospalax* (*Volemys* (*Mus* + *Rattus*)))), Hystricognathi minus *Thryonomys* (*Hystrix* (*Erethizon* + *Cavia*)), and "sciuroids" with a nested hystricognath, *Thryonomys*: (*Myoxus* (*Thryonomys* (*Sciurus* (*Aplodontia* + *Cratogeomys*)))). Among rodent taxa, only the clades within Muridae were recovered in greater than 51% of bootstrap replicates: (*Volemys*: BP = 97 (*Mus* + *Rattus*: BP = 99)). In contrast to the lack of recovery of Glires or Rodentia, the parsimony analysis of the tRNA genes recovered a monophyletic Primates with the expected branching order and bootstrap support of 93% or greater.

Of the numerous Bayesian analyses of the tRNA gene data, analyzed under the covarion model (all tRNA genes combined or separated by strand, and three Temp settings each), none of the runs from the MCMC simulation converged, based on the standard deviation of the split frequencies (SD > 0.10), even at higher Temp settings and up to 10×10^6 generations. This suggests that modeling this rather small dataset under an RNA-based compensatory model may require an order of magnitude of additional generations to converge on a stationary posterior distribution.

A MCMC simulation estimated under a 4-state general-time reversible (DNA) model converged in 2×10^6 generations, based on the standard deviation of the split frequencies (SD = 0.004). This simulation recovered a monophyletic Glires with low

support (PP = 59), Lagomorpha sister to a Muroid + Dipodoid + Geomyoid clade with moderate support (PP = 87), and the remaining rodents in another monophyletic clade with moderate support (PP = 83) (Fig 2.4). As with the parsimony analysis of the tRNA genes, the Bayesian analysis recovered a Dermoptera + Primates clade with strong support (PP = 97) with the expected branching order within Primates, all with 100% posterior probabilities. While the effective sample size (ESS = 1915) of the -*ln* L values sampled during the DNA-model MCMC simulation suggests adequate mixing between the Markov chains, this non-compensatory model of sequence evolution can not adequately reflect the type of substitutions that occur in strongly structural molecules, such as mt tRNAs. Thus, any conclusions drawn from the resulting topology should be considered with caution.

Evaluation of Recovered Trees and Comparisons with Previous Hypotheses

The topologies recovered from the MP, ML, and Bayesian analyses of the protein-coding genes (all sites and 1st and 2nd positions only) and the tRNA genes (DNA models only) were compared using the Kishino-Hasegawa test. The ML topology, recovered from ML analysis (all sites and 1st and 2nd positions only) and Bayesian analyses (all positions combined, partitioned by gene, or partitioned by codon position) was significantly better than trees recovered in other analyses (Table 2.3). The ML of taxa (Fig 2.5) and evaluated using the S-H test (Table 2.4). The ML tree recovered by the present dataset was significantly better in half of the comparisons. In one case, the

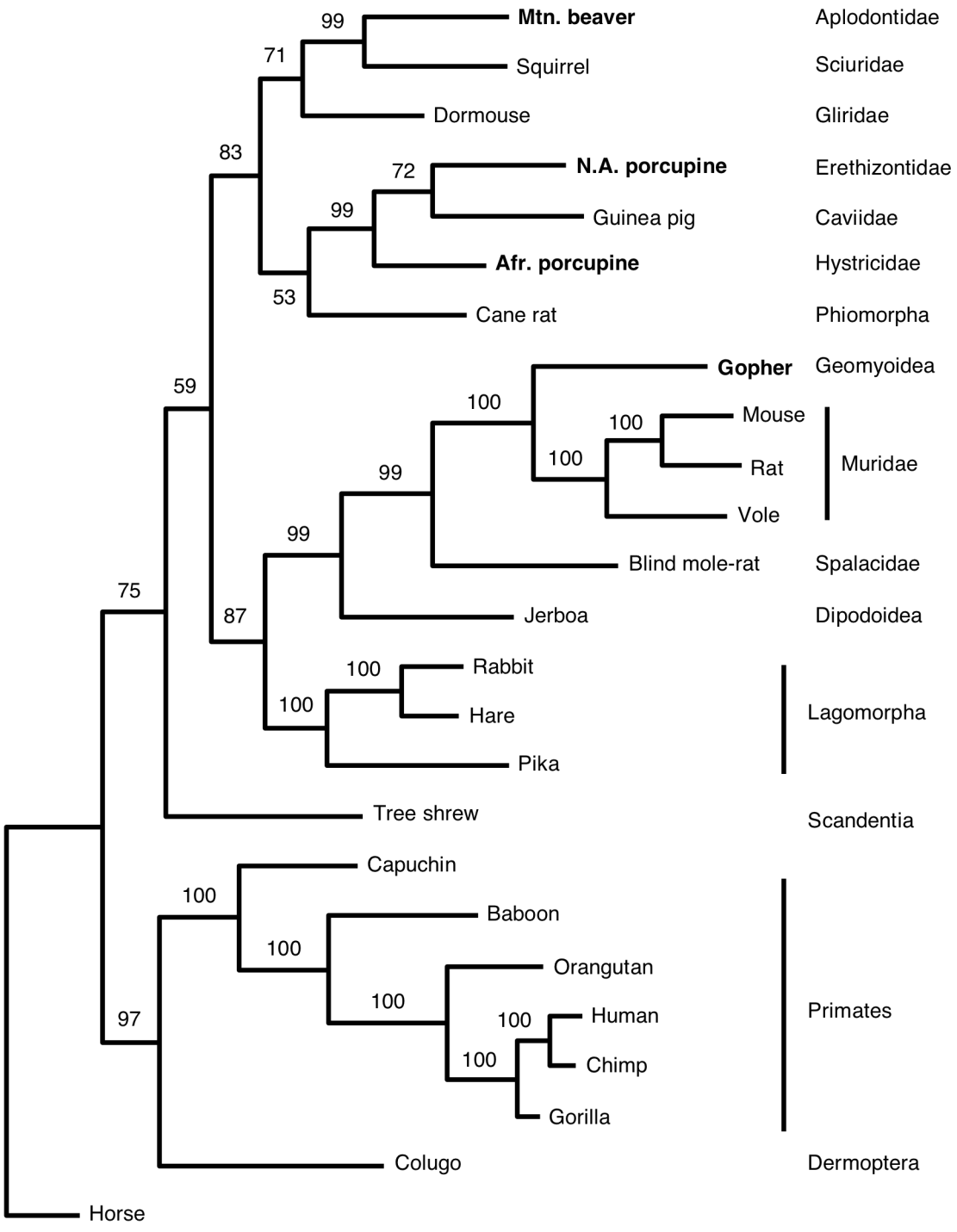


FIG. 2.4. —Maximum-likelihood topology recovered from MCMC analysis of all unambiguously aligned tRNA positions under a DNA-based 4-state model (*-ln* L = 10,827; ESS = 1915; 2 x 10^6 generations). Values above or below branches indicate posterior probabilities.

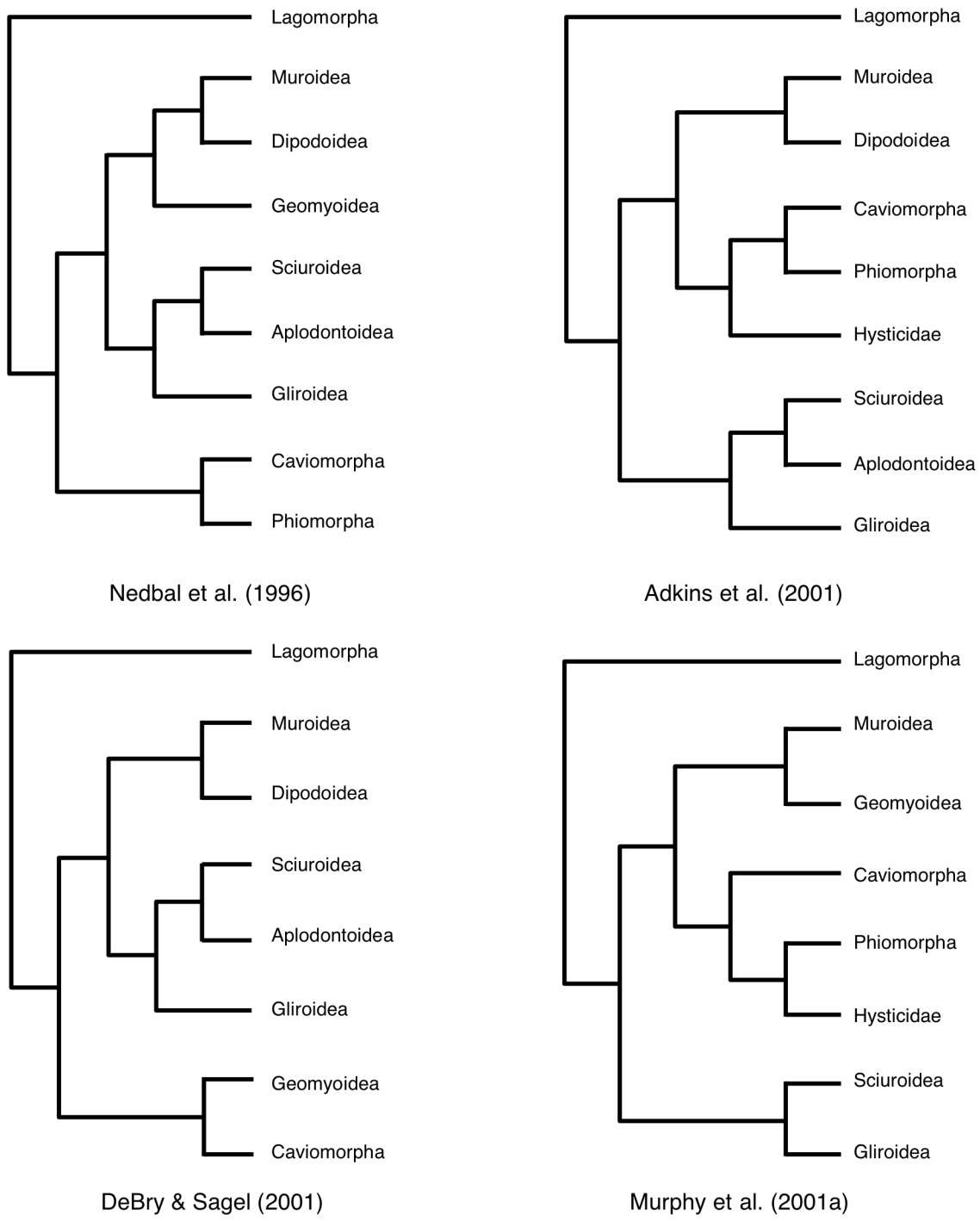
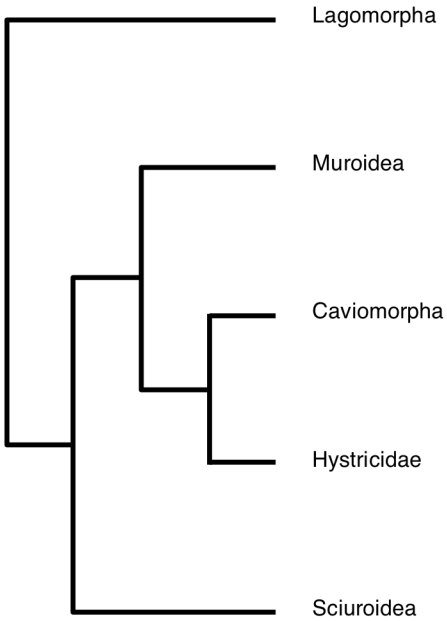
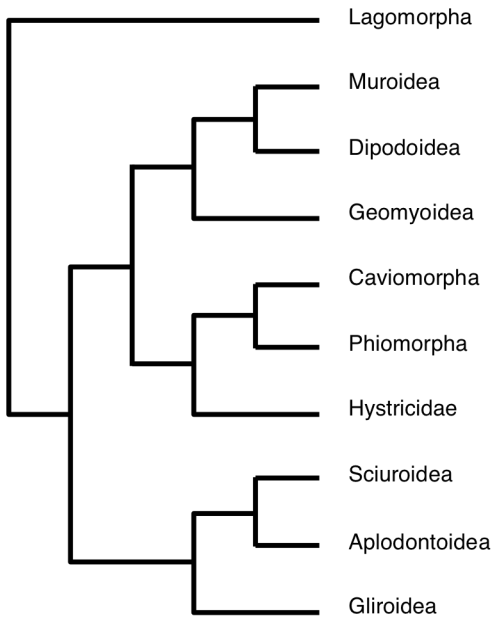


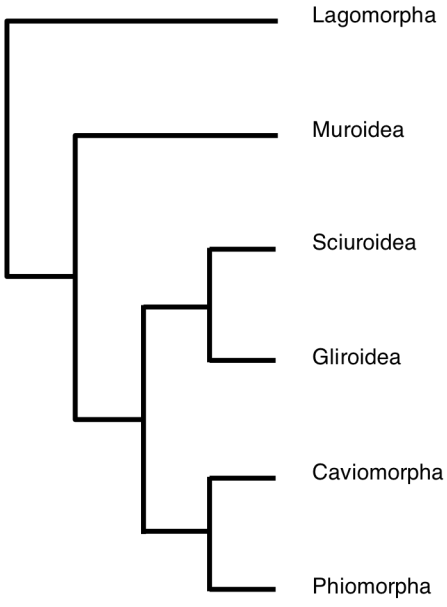
FIG. 2.5. —Simplified topologies from previous molecular studies of higher-level relationships in Rodentia (compared with the present data using the Kishino-Hasegawa test –see Table 2.4).



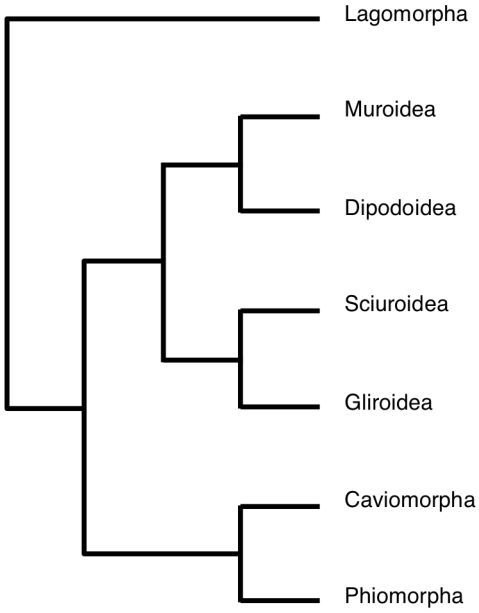
Murphy et al. (2001b)



Huchon et al. (2002)



Hudelot et al. (2003)



Reyes et al. (2004)

FIG. 2.5 —continued.

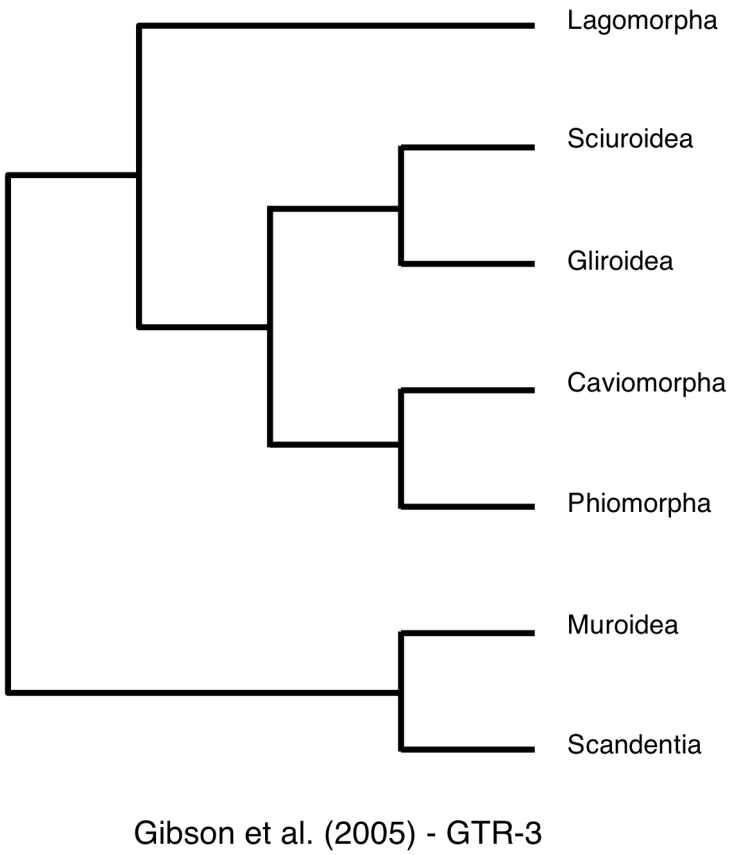


FIG. 2.5 —continued.

Table 2.3 Comparisons of topologies from different analyses using the Kishino-Hasegawa test

Analysis	*-ln* L	Difference in *-ln* L	*P*
ML - all sites (GTR+G+I)	154,267	-	-
ML - 1st & 2nd positions only	154,267	0.00	1.000
ML - all sites (JC)	154,364	97.48	< 0.05
ML - 1st & 2nd positions only (JC)	154,305	38.42	< 0.05
MP - all sites - tree 1	154,572	305.35	< 0.05
MP - all sites - tree 2	154,510	242.69	< 0.05
MP - 1st & 2nd positions only - tree 1	154,350	83.02	< 0.05
MP - 1st & 2nd positions only - tree 2	154,298	31.29	< 0.05
MP - tRNA - tree 1	154,572	305.35	< 0.05
MP - tRNA - tree 2	154,510	242.69	< 0.05
BI - tRNA - DNA model	154,365	98.58	< 0.05

ML = maximum likelihood, GTR = general time-reversible model, JC = Jukes-Cantor model, MP = maximum parsimony, BI = Bayesian inference

Table 2.4 Comparisons to topologies from previous studies of higher-level rodent relationships using the Shimodaira-Hasegawa test[1]

Study	Number of taxa compared[2]	*-ln* L of ML tree (this study)	*-ln* L	Difference in *-ln* L	*P*
Nedbal et al. (1996)	16	106,748	106,772	23.69	**0.020**
Adkins et al. (2001)	15	101,103	101,120	16.76	0.084
DeBry & Sagel (2001)	14	96,781	96,819	37.80	**0.005**
Murphy et al. (2001a)	14	94,560	94,590	29.84	**0.008**
Murphy et al. (2001b)	11	76,533	**76,522**	**+11.12**	**0.039**[3]
Montgelard et al. (2002)	14	96,781	96,791	10.17	0.185
Huchon et al. (2002)	16	106,748	106,775	26.59	**0.024**
Hudelot et al. (2003)	12	84,940	84,951	10.38	0.169
Reyes et al. (2004)	13	91,344	91,363	19.21	**0.023**
Gibson et al. (2005) - (GTR-3)	13	90,860	90,877	17.24	0.114

[1]Trees simplified to match taxonomic groups between each comparison. [2]Number of taxa compared includes ingroup and outgroup taxa. [3]The simplified topology (8 rodent taxa) from Murphy et al. (2001b) was significantly better than the simplified topology for the present study.

simplified ML presented here was significantly worse than the topology for rodent relationships proposed in Murphy et al. (2001b). The branching pattern of Sciuroidea (or Sciuroidea + Gliroidea) as sister to all remaining rodents, proposed in the two Murphy et al. (2001a, b) datasets, was incongruent with all other hypotheses, including the present dataset. This comparison included only eight of the thirteen available rodent taxa, and this poor sampling may have a strong influence on the results of the S-H test (Table 2.4).

DISCUSSION

Character Performance and Hypothesis Selection

While the topologies recovered by parsimony analysis were worse, based on the Kishino-Hasegawa criterion, than the ML tree, the decay indices (particularly the partitioned decay indices) are still useful as a measure of performance of each gene in recovering particular clades of interest. Partitioned decay indices (Baker and DeSalle 1997; Baker, Yu, and DeSalle 1998) provide a way to determine how different, non-overlapping data partitions contribute to the overall decay index for any particular node. Partitioned decay indices are calculated in the same manner as the overall decay indices, by subtracting the tree length of the most-parsimonious tree from the shortest tree for a particular data partition. Values for individual partitioned decay indices can be either positive or negative, but the sum of partitioned indices for any given node must sum to the overall decay index for that node. In the recovery of Rodentia and key intraordinal

clades, ND4 and ND5 performed very well with positive indices for all but one major clade (Myomorpha: ND4 pDI = -1) and ATP6 contributed to the recovery of all non-myomorph clades, while COII, cytochrome-*b*, ND2, and ND3 all performed poorly in the recovery of rodent clades (Table 2.2). Since the two MP trees (Fig 2.1) differed in the position of the tree shrew (*Tupaia*: Scandentia) with respect to Glires (Rodentia + Lagomorpha), no decay indices could be calculated for the recovery of Glires. However, the same genes contributed to the recovery of the Scandentia + Glires clade (Table 2.2). The pattern of recovery performance is markedly different for the recovery of clades related to Primates, with ATP6, ATP8, cytochrome-*b*, ND1, ND2, ND3, ND4, and ND5 all contributing to the recovery of all primate-related clades. No gene showed universally negative decay indices in the Primates clades (Table 2.2) and primates showed a higher overall retention index for all positions in the present dataset, compared to rodents (RI = 0.44 and 0.25, respectively). Relative rate tests (Tajima 1993) across the four genes that performed poorly in resolving Rodentia (COII, Cyt-*b*, ND2, ND3) revealed significant rate variation both within the Scandentia + Glires clade, and between this clade and Primates, particularly for COII and Cyt-*b* (Table 2.5). Rodents (and lagomorphs) showed significant intraordinal rate variation in COII, Cyt-*b*, and ND2, while primates showed weakly significant rate variation only in Cyt-*b*. These results are consistent with previous comparisons of rates of mitochondrial gene evolution between rodents and primates. Honeycutt et al. (1995) showed that COII and cytochrome-*b* were phylogenetically inconsistent with each other and previous

Table 2.5 Range of *p*-values from relative rates tests (Tajima 1993) for select mitochondrial genes

Taxa	COII	Cyt-*b*	ND2	ND3
Rodents vs. Rodents	1.0–0.00008***	1.0–0.000065***	1.0–0.000818*	1.0–0.0201
Rodents vs. Lagomorphs	1.0–0.00027**	0.9446–0.000007****	0.7077–0.000017****	1.0–0.0116
Rodents vs. Scandentia	0.9230–0.00014**	1.0–0.0035	0.7971–0.00324	0.1083–0.000174**
Rodents vs. Primates	0.0624–0.000008****	1.0–0.0000000****	0.9549–0.0148	1.0–0.1037
Lagomorphs vs. Lagomorphs	0.5175–0.1470	0.6737–0.0190	0.5870–0.0941	0.8026–0.3657
Lagomorphs vs. Scandentia	0.7675–0.2876	0.2012–0.000475***	0.9065–0.1498	0.1515–0.0339
Lagomorphs vs. Primates	0.0068–0.000001****	0.1902–0.0000000****	0.6534–0.000458**	0.6041–0.0811
Scandentia vs. Primates	0.00033**–0.000006****	1.0–0.00105*	0.5832–0.0444	0.0148–0.0017
Primates vs. Primates	0.9042–0.1083	0.8307–0.000516**	1.0–0.0523	0.9115–0.4561

All *p*-values Bonferroni-corrected for multiple comparisons

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.005$

hypotheses for eutherian relationships and contributed this inconsistency primarily to lineage-specific variation in nucleotide substitution rates. In a comparison of whole mtDNA, Weinreich (2001) found significant deviations in the ω (d_N/d_S: ratio of differences in nonsynonymous vs. synonymous substitutions) estimated for species-pairs of primates and rodents in all protein-coding genes but ND4L (mean $\omega_{rodent}/\omega_{primate}$ = 0.206).

It has been argued (Goldman 1993; Gaut and Lewis 1995; Sullivan and Swoford 1997; Huelsenbeck 1998; Lemmon and Moriarty 2003; but see Yang 1997; Bruno and Halpern 1999) that selection of an adequate model of sequence evolution is essential in model-based phylogenetic methods (neighbor-joining, maximum likelihood, Bayesian inference, etc.). Based on the Bayesian Information Criterion (BIC) estimated in ModelTest (Posada and Crandall 1998), the present dataset was analyzed under one of the most complex DNA models available in most analytical software (GTR + Γ +I estimated across all sites). Since the ML analysis of 1st and 2nd positions recovered the same topology as the ML analysis of all sites, one might argue that third positions are not contributing phylogenetic information to the recovery of relationships. On the contrary, this illustrates that the GTR + Γ +I model is adequately modeling the homoplasy expected in third codon positions over the divergence times seen in the present dataset. To determine if adequate modeling was essential, as suggested by Sullivan and Swofford (1997) in a reduced but similar dataset (D'Erchia et al. 1996) or if the crucial factor in the question of rodent monophyly relates to increased taxon sampling, I analyzed the present dataset under the simplistic Jukes-Cantor model (2 free

parameters: equal base frequencies and equal substitution rates). In analyses of all sites and 3rd positions excluded, monophyletic Glires and Rodentia were recovered. Low support values (non-parametric bootstrap), however, suggest that model selection is also an important factor in accurate phylogenetic estimates.

As the ML topology (Fig 2.3) was statistically more likely than those recovered by parsimony methods, and was also recovered as the most-likely tree in the Bayesian analyses (Table 2.4), I selected this topology as the best hypothesis of phylogenetic relationships for the present dataset. While the results of Shimodaira-Hasegawa tests comparing the ML tree from the present dataset with previously published hypotheses of higher-level relationships within Rodentia were mixed, the present ML tree was statistically preferred in both comparisons that used all 13 rodent taxa (Table 2.5).

Current software packages available for Bayesian phylogenetic analysis (MrBayes, PHASE, etc.) allow the estimate of model parameters on a per-partition basis. Estimating model parameters across gene partitions or codon position significantly increased the likelihood scores of the best tree recovered in each scheme, but these analyses converged on the same topology recovered under the maximum likelihood optimization criterion (Fig 2.3).

tRNA Genes

In their phylogenetic utility in recovering relationships within the highly divergent Rodentia, the mt tRNA genes appear to be poor phylogenetic markers. This is most likely due to the high levels of variation observed within rodent mt tRNA gene

sequences compared to all other available mammals (including monotremes and marsupials; see Chapter III). As observed in rRNA genes across highly divergent taxa, the conservation of secondary structure of mt tRNA may be maintained at the expense of reduced primary structure (Gutell 1996). In contrast to their poor behavior in recovering rodent relationships, the mt tRNAs recovered the expected relationships among the included primates under both MP and Bayesian methods. While these conclusions are based on nucleotide-based (DNA) analyses and not base-pair (RNA) analyses, overall relationships should not differ substantially under a method that uses base-pairs. With current computer hardware (single or dual processor G5 workstations), full Bayesian analyses of even a moderately sized dataset (presented here) under appropriate compensatory base-pair models appears intractable due to the high number of free parameters in these models (lack of convergence after 2×10^7 generations: SD of split frequencies = 0.09). In its current iteration (v2.01 beta), PHASE (Jow et al. 2002) is not ported for parallel processing and MrBayes (v3.1.1: Ronquist and Huelsenbeck 2003) requires a commercial third-party application (Pooch: Dauger Research, Huntington Beach, CA) to run as a parallel application, making options limited.

Relationships Based on Present Dataset

Glires

The monophyly of the cohort Glires has been supported by the morphological and paleontological communities for over two centuries (Linnaeus 1758; Brandt 1855; Tullberg 1899; Simpson 1945; Li and Ting 1985; Luckett 1985; Novacek 1985;

Mossman 1987; Luckett and Hartenberger 1993; but see Gidley 1912). With few exceptions (Lin, Waddell, and Penny 2002; Reyes et al. 2004), analyses of complete mitochondrial genomes have failed to show strong support for the monophyly of a superordinal Glires (Rodentia + Lagomorpha) (D'Erchia et al. 1996; Reyes et al. 2000; Mouchaty et al. 2001; Lin, Waddell, and Penny 2002). In contrast, recent studies employing nuclear sequence data have confirmed the monophyly of Glires (Madsen et al. 2001; Murphy et al. 2001a, 2001b; Amrine-Madsen et al. 2003). As indicated by Sullivan and Swofford (1997), the discrepancy between mitochondrial and nuclear gene analyses may be explained by failure on the part of previous mtDNA studies to consider appropriate models that accommodate heterogeneity associated with among site variation and differential rates of evolution. Results presented here at least partially support the claims of Sullivan and Swofford (1997). For instance, Bayesian analyses retrieved a monophyletic Glires supported by high posterior probabilities (PP = 100), MP analyses failed to support monophyly of Glires, and ML analyses, even with the most appropriate model recovered a monophyletic Glires but with low bootstrap support (BP < 50). The discrepancy in support between posterior probabilities produced by Bayesian analysis and non-parametric bootstrap values derived from ML analysis appears to be a consistent pattern. Suzuki, Glazko, and Nei (2002) used four-taxon simulations to assess the statistical confidence of posterior probabilities and maximum likelihood (non-parametric) bootstrap proportions, and concluded that PP values were excessively high, while BP values were only slightly low. Another study (Erixon et al. 2003) simulated five-taxon trees and showed that BP values were conservative under the

correct model of sequence evolution, but cautioned that erroneous conclusions could be made using PP values under underparameterized models. In a recent empirical study employing three mt RNA genes (12S, 16S, and $tRNA^{Val}$) for 23 snake species, Wilcox et al. (2002) critically examined the same issue through the analysis of 120 simulated datasets derived from their ML topology and model parameters. Their conclusions, based on a real dataset, showed that Bayesian PP values were far better estimates of phylogenetic support than non-parametric bootstrap values, in contrast to Suzuki, Glazko, and Nei (2002).

Rodentia

Morphologically, rodent monophyly has never been an issue (Cuvier 1917; Brandt 1855; Tullberg 1899; Simpson 1945; Luckett and Hartenberger 1993). All rodents share a large suite of shared derived characteristics associated with cranial and post-cranial skeleton, soft tissue, dentition, and jaw. In particular, the anatomical features associated with gnawing, combined with chewing by the cheekteeth include: 1) ever-growing incisors with enamel restricted to the labial surface; 2) loss of canine teeth and a diastema that separates the incisors from the cheekteeth; and 3) the zygomasseteric structure of the masseter muscles and the infraorbital foramina of the skull (Marivaux, Vianey-Liaud, and Jaeger 2004). However, accurate diagnoses of higher relationships among rodent families is a major issue. This problem has resulted in numerous subordinal groupings based on different interpretations of fossils and morphological traits (see Chapter I). Whole mitochondrial genome sequences have challenged rodent

monophyly and stand in contrast to most recent phylogenetic studies based on nuclear gene sequences. At the same time, different combinations of nuclear sequences have not provided a highly resolved phylogeny for rodent families, especially with respect to the placement of the families (Adkins et al. 2001; DeBry and Sagel 2001; Huchon et al. 2002; Rowe 2002).

As with Glires, the level of support for the monophyly of Rodentia differs between likelihood and Bayesian analyses (BP < 50, PP = 100). Given the overwhelming body of evidence supporting Rodentia, this relationship is considered valid. Among rodents, three major clades were recovered: 1) Sciuridae + Aplodontidae; 2) Gliroidea + Hystricognathi; and 3) Geomyoidea (Dipodoidea + Muroidea), each discussed below.

Sciuroidea (Sciuridae + Aplondontidae)

The sister-relationship between the Sciuridae and Aplodontidae has been proposed since Wood (1955) with support from auditory anatomy (Lavocat and Parent 1985; Vianey-Liand 1985), albumin immunology (Sarich 1985), and numerous molecular studies (Nedbal, Honeycutt, and Schlitter 1996; Adkins et al. 2001; DeBry and Sagel 2001; Huchon et al. 2002; Montgelard et al. 2002). The present study adds to the growing body of evidence for this relationship with strong support (BP = 92, PP = 100).

Hystricognathi

The monophyly of Hystricognathi was well supported in the present dataset (BP = 98, PP = 100), but the relationships within the suborder require further discussion. The Caviomorpha (*Erethizon* and *Cavia*) formed a monophyletic group with moderate to strong support (BP = 61, PP = 100). The lack of taxon sampling of complete mitochondrial genome sequences within the Caviomorpha prevents examination of the position of the Erethizontidae with respect to other caviomorphs. The Erethizontidae show a number of uniquely derived morphological characters compared with other caviomorphs, such as dental characters (Wood and Hermanson 1985) and the retention of an internal carotid artery (Bugge 1985). These autoapomorphic characteristics have led some to place the New World porcupines in their own superfamily, Erethizontoidea (Patterson and Wood 1982) or suborder, Erethizontomorpha (Bugge 1974). In addition, some have proposed an independent invasion of the erethizontid ancestor, separate from that of the remaining caviomorph stock (Wood 1980). Recently, Rowe (2002), in an analysis of two nuclear genes (transthyretin (TTR) and growth hormone receptor (GHR)) and the mt 12S rRNA gene, found conflicting results in the placement of *Erethizon.* The proposal of Bugge (1974) that the Erethizontidae were the most basal of the Caviomorpha lineages was equivocal to the alternative with Erethizontidae sister to Cavioidea (Rowe and Honeycutt 2002).

Phiomorpha

The monophyly of the Old World hystricognath rodents (Phiomorpha *sensu stricto* Lavocat 1973) was refuted in the present analysis with moderate to high support (BP = 61, PP = 100). As stated in Chapter I, there has been much controversy in determining the relationships among Old World hystricognath rodents. Wood (1965) proposed Bathy-Phiomorpha to include the Old World Thryonomyidae, Petromuridae, and Bathyergidae, to the exclusion of the Old World porcupines, Hystricidae. Patterson and Wood (1985) proposed an independent invasion of Africa by an Asian hystricid ancestor, resulting in Hystricidae as the sole members of Hystricomorpha. The term Hystricomorpha comes with historical baggage back to Brandt (1855). To avoid taxonomic confusion among workers, Luckett and Hartenberger (1993) recommended the use of Bathy-Phiomorpha and Hystricidae to refer to the whole of Old World hystricognaths. Rowe (2000) found no evidence for a monophyletic Phiomorpha *sensu stricto*, and could not determine whether the Bathy-Phiomorpha or Hystricidae represented the sister-taxon to the Caviomorpha. Each of these hypotheses: 1) monophyletic Phiomorpha; 2) Bathy-Phiomorpha sister to Caviomorpha; and 3) Hystricidae sister to Caviomorpha, have gained support from recent molecular studies (Murphy et al. 2001a; Adkins et al. 2001 and Huchon et al. 2002; and Rowe 2002, respectively). In the present analysis, the Old World porcupine, *Hystrix*, was recovered as sister to the Caviomorpha, consistent with the combined analyses of TTR, GHR, and vWF (von Willebrand factor) by Rowe (2002).

Gliroidea

In the present analysis, the dormice (Gliroidea) were weakly to strongly supported as sister to the Hystricognathi (BP = 55, PP = 100). Classically, glirids were classified as close relatives of Myomorpha based on their myomorphous zygomasseteric structure (Simpson 1945; Wahlert 1978), but some molecular studies also support this affinity (Sarich 1985). Vianey-Liaud (1985) showed that the "pseudo-myomorphy" of the glirids is derived from an ancestral protrogomorphous condition, in contrast to a hystricomorphous-derived condition in true myomorphs. This led other investigators to assign glirids to "Sciurognathi", closely related to sciurids or aplodontids (Bugge 1971, 1985; Wood 1980; Flynn, Jacobs, and Lindsay 1985; Lavocat and Parent 1985). Recent molecular studies have lent support for a sciuroid affinity for glirids (Nedbal, Honeycutt, and Schlitter 1996; Murphy et al. 2001a; Huchon et al. 2002; Gibson et al. 2005). Others remain skeptical about either placement, instead suggesting that the Gliridae, †Theridiomyidae, and Anomaluridae may form a separate clade from either "Sciurognaths" or Ctenohystrica (= Ctenodactylidae + Hystricognathi) *sensu* Huchon, Catzeflis, and Douzery (2000) (Flynn et al. 1985). While at first glance, a Gliridae + Hystricognathi clade may appear to be without support from other biological or paleontological data, this placement may simply be a reflection on the lack of representatives in the present dataset from other key major rodent groups, Ctenodactylidae or Anomaluroidea *sensu* Simpson (1945) (= Anomaluridae + Pedetidae).

Myomorpha

The monophyly and relationships within the Myomorpha *sensu stricto* are generally without controversy and supported by numerous morphological, paleontological, and molecular data. The monophyly of the Muridae (*Mus*, *Rattus*, and *Volemys*) and Muroidea (Muridae + Spalacidae (*Nannospalax*)) was strongly supported in the present dataset (BP = 100 and 99, PP = 100 and 100, respectively). The sister-group relationship between the Dipodoidea (*Jaculus*) and Muroidea, referred collectively as Myomorpha or Myodonta, is strongly supported by morphological (Bugge 1985), paleontological (Vianey-Liaud 1985), and molecular (Nedbal, Honeycutt, and Schlitter 1996; Adkins et al. 2001; Huchon et al. 2002; Reyes et al. 2004) data. The present dataset lends further support for this relationship with strong bootstrap proportions and posterior probabilities (98 and 100%, respectively).

Geomyoidea

The Geomyoidea is comprised of the Geomyidae (pocket gophers) and Heteromyidae (pocket mice and kangaroo rats). This superfamily is strongly supported by morphological (Fahlbusch 1985) and molecular data (Nedbal, Honeycutt, and Schlitter 1996; Matthee and Robinson 1997; Huchon et al. 2002) and I assume here that *Cratogeomys* is representative of this superfamily. An affinity between the Geomyoidea and Myomorpha has been proposed for decades (Hill 1937; Wilson 1949; Wood 1955; Wood 1959; Wahlert 1978), and the two lineages are linked by similarities in numerous fetal membrane characters to the exclusion of sciurids (Luckett 1985) and the carotid

arterial pattern of heteromyids is consistent with that of muroids (Bugge 1985). Fossil geomyoids show strong affinities with *Adelomys* (Muroidea) (Lavocat and Parent 1985). In contrast, serum albumin data suggests that Geomyoidea is as distant from Myomorpha as it is from the Sciuridae + Aplodontidae clade (Sarich 1985).

The placement of the Geomyoidea in molecular studies has been extremely varied. Both Nedbal, Honeycutt, and Schlitter (1996) and Huchon et al. (2002) place the Geomyoidea in a clade with Anomaluroidea that is sister to Myomorpha. Their hypotheses differ in that Huchon et al. (2002) include Castoroidea (beavers) in this clade, while Nedbal, Honeycutt, and Schlitter (1996) recover Castoroidea sister to a (Gliroidea (Ctenodactylidae (Sciuridae + Aplodontidae))) clade. One recent molecular study (DeBry and Sagel 2001) placed the Geomyoidea sister to Hystricognathi (represented by two caviomorph families: Chinchillidae and Erethizontidae), but as with the present study, key groups such as Ctenodactylidae and Anomaluroidea were missing from their dataset. The present dataset recovered a Geomyoidea + Myomorpha clade with weak to strong support (BP = 55, PP = 100), but this may either reflect a true affinity between the groups directly or may be the result of the lack of important intermediate taxa (e.g., Anomaluroidea and Castoroidea).

Final Remarks

The present dataset has increased the number of rodent complete mitochondrial genome sequences by 44%, focusing on taxonomically important and historically problematic taxa and lends support for a number of previously recognized intraordinal

relationships. Lack of adequate taxon sampling remains problematic and work continues in the Honeycutt lab to generate complete mitochondrial genome sequences for Ctenodactylidae, Castoridae, Pedetidae (Anomaluroidea), and a number of additional caviomorph taxa.

CHAPTER III

A COMPREHENSIVE ANALYSIS OF THE EVOLUTION OF MAMMALIAN MITOCHONDRIAL tRNA GENES

INTRODUCTION

A nuclear tRNA (yeast $tRNA^{Ala}$) was the first nucleic acid for which a secondary structure was determined (Holley et al. 1965), and another nuclear tRNA (yeast $tRNA^{Phe}$) was the first for which a three-dimensional tertiary model was described (Kim et al. 1973, 1974). Since then, studies of tRNA molecules have generally included aspects of their secondary and tertiary structural features. Compilations of these small molecules and their genes from nuclear, organellar, and prokaryote sources were initiated by Sprinzl and colleagues in 1984 with annual reports in *Nucleic Acids Research* (see Sprinzl et al. 1998) and continue as an updated online resource (http://www.uni-bayreuth.de/departments/biochemie/trna/).

As a large number of tRNAs accumulated, patterns of rigid secondary and tertiary structural constraints emerged as the norm for tRNAs found among prokaryotes, eukaryote nuclear genomes, and plant organelles. In contrast, animal mitochondrial (mt) tRNAs show strong structural deviations from the canonical cloverleaf structure of their nuclear counterparts (Wolstenholme et al. 1987; Yokogawa et al. 1991; Janke et al. 1994; Moriya et al. 1994; Steinberg, Gautheret, and Cedergren 1994; Watanabe et al. 1994; Takemoto et al. 1995; Janke et al. 1997; Sprinzl et al. 1998; Dörner et al. 2001;

Nilsson et al. 2003). Whereas many nuclear tRNAs are identical in taxa as varied as *Mus*, *Bos*, *Homo*, and *Xenopus* (e.g., $tRNA^{Phe}$: Sprinzl et al. 1998), animal mt tRNAs are marked with relaxation of both secondary and tertiary structural constraints. Among seven animal taxa (*Mus*, *Rattus*, *Bos*, *Homo*, *Xenopus*, *Gallus*, and *Strongylocentrotus*), Kumazawa and Nishida (1993, 1995) reported that up to 75% of all stem-forming base-pairs were variable with a high proportion of non-Watson-Crick pairs, and proposed that many changes may occur without immediate compensatory base change (CBC). Similarly, Jukes (1995) determined that none of the 20 invariant or semi-invariant sites identified by Sprinzl et al. (1998; their fig. 3.1) were invariant among *Mus*, *Rattus*, *Bos*, *Homo*, and *Xenopus*. The reduction in tertiary structure constraints in animal mt tRNAs has been linked with the simpler systems of transcription of mtDNA and recognition of mt tRNAs by enzymes during aminoacylation and protein synthesis (Wilson et al. 1985; Kumazawa et al. 1989; Kumazawa et al. 1991). Because constraints on positions associated with tertiary structure are reduced, variation at these positions may be increased, particularly in domains with limited function, such as the D-stem (Kumazawa and Nishida 1993). Without consideration for base-pair covariation, transitions in stem positions of mt tRNAs among seven diverse deuterostomes maintain linearity to saturation at ~25% (estimated divergences up to ~100 mybp), and transitions continue to accumulate linearly without reaching an asymptote (divergences in excess of 600 mybp) (Kumazawa and Nishida 1993).

As with all but one of the mitochondrial protein-coding genes (ND6), the majority of mt tRNAs are encoded by the heavy (H)-strand (Arg, Asp, Gly, His, Ile,

Leu^{CUN}, Leu^{UUR}, Lys, Met, Phe, Ser^{AGY}, Thr, Trp, and Val), with the remaining eight (Ala, Asn, Cys, Gln, Glu, Pro, Ser^{UCN}, and Tyr) encoded by the light (L)-strand. Replication of mammalian mitochondrial DNA is asymmetrical with the two strands being synthesized from two distinct and distant replication origins (Clayton 1982). The H-strand origin of replication (O_H) is located in the main non-coding portion of the molecule: the D-loop of the control region. Replication begins with displacement of the parental H-strand by the replication bubble. The replication bubble continues approximately two-thirds (~ 11 kb) around the molecule until the L-strand origin of replication (O_L: located in the WANCY tRNA cluster of the mammalian mt genome) is exposed and L-strand synthesis begins in the opposite direction. Since replication of the mitochondrial genome is slow (up to 2 hrs) (Clayton 1982), portions of the parental H-strand remain exposed as a single-stranded molecule for up to 80–100 m. During this time, the single-stranded H-strand is prone to mutation by hydrolytic deamination and oxidation, and the H-encoded tRNA genes are likely subject to the same directional mutational pressure observed in the protein-coding genes (Reyes et al. 1998; Bielawski and Gold 2002; Faith and Pollock 2003; Gibson et al. 2005).

Mutations in mt tRNAs have received attention for their association with mitochondrion-linked human pathology (Goto, Nonaka, and Horai 1990; Wallace 1992, 1999; Larsson and Clayton 1995; Schon, Bonilla, and DiMauro 1997; Helm et al. 2000; Florentz and Sissler 2001; Sissler et al. 2004). To date, 20 mitochondrial disorders have been linked to over 100 point mutations in human mt tRNA genes (Mitomap 2005). To provide a better understanding of the level of variation present in mt tRNAs of mammals

and potential correlations to human pathology, Helm et al. (2000) surveyed complete mitochondrial genome (mt genome) sequences of 31 mammals, sampling nine orders from Eutheria, two orders from Metatheria (Marsupialia), and two monotremes. They produced robust sequence alignments based on putative secondary structures for the 22 mt tRNA genes and generated informative structural diagrams providing typical and consensus data for the 31 mammalian taxa examined.

As the number of complete mt genome sequences has grown, few authors have examined mt tRNAs, instead focusing nearly exclusively on protein-coding and/or rRNA genes (*e.g.*, Xu and Arnason 1994; Arnason, Gullberg, and Janke 1997; Arnason et al. 2002; Janke et al. 1997; Reyes et al. 1998; Arnason and Janke 2002; Gibson et al. 2005). Currently, the complete mt genomes have been sequenced for all 17 extant orders of eutherians, five orders of marsupials, and two monotremes. In animal mt genomes, however, the tRNAs are the only loci with reasonable representation on both the H- and L-strands of the genome, and thus offer an unique opportunity to examine potential differences in the evolution of these structurally similar genes. In the present study, I extended the efforts of Helm and colleagues to characterize the structure of mammalian mitochondrial tRNAs. I mined public databases (OGRe: Jameson et al. 2003; NCBI *Entrez* Genome) for the currently available complete mammalian mitochondrial genome sequences. I assembled a data set of the 22 mt tRNA genes from 109 mammals, sampling 17 orders of eutherians, five orders of marsupials, and two monotremes. We also compare the mt tRNA genes between rodents and other eutherian mammals. Nedbal, Honeycutt, and Schlitter (1996) noted that mt 12S rRNA showed more variation

across various groups of rodents than among outgroup mammals used to examine relationships among families in Rodentia. In addition, Honeycutt et al. (1995) and Adkins, Honeycutt, and Disotell (1996) demonstrated that rodents generally show higher rates of nucleotide substitution in mtDNA compared to non-rodents. In the case of the 12S rRNA gene, alignment was problematic, mainly because indels (predominately in loop regions) were present at a high rate. This implies minimal sequence conservation and makes reliable homology assignments difficult. To determine whether these increased rates extend to the mt tRNAs, I added the sequences of 11 clustered tRNAs (regions a (IQM), b (WANCY), and d (HSL): Kumazawa and Nishida 1993) for 32 additional rodent species to the data set, sampling 26 of the 32 extant families of rodents.

The purpose of this paper is to better characterize a comprehensive set of functional mammalian mt tRNA genes with emphasis on 1) variation and revision of secondary structure models; 2) nucleotide composition; 3) base-pair composition; 4) variation in stem and loop size among mammals; and 5) the potential effects of genomic position and the duration of single-strandedness on these features.

MATERIALS AND METHODS

Sequence Mining and Multiple Sequence Alignment

The 22 tRNA genes of 109 complete mammalian mitochondrial (mt) genomes were extracted and compiled from the OGRe (Jameson et al. 2003 – current version of OGRe available at http://ogre.mcmaster.ca) and NCBI *Entrez* Genome databases. The data set included representatives from all 17 extant orders of eutherian mammals, 5

orders of metatherians, and 2 monotremes (for taxa and their GenBank accession numbers, see Appendix 2).

Sequences were exported into Word® v2001 (Microsoft Corp.) for manual alignment according to secondary structure, following the convention of Kjer (1995) with slight modifications (Gillespie 2004; Gillespie et al. 2004) (alignment available in Appendix 3 and at the *jRNA* website: http://hymenoptera.tamu.edu/rna). For each tRNA gene, sequences were converted to RNA and vertically aligned by eye, focusing primarily on positions of the four helical regions. Alignments followed the secondary structural models of Kumazawa and Nishida (1993), Sprinzl et al. (1998), and Helm et al. (2000), with the exception of $tRNA^{Lys}$ of the metatherians, for which the model of Nilsson et al. (2003) was adopted. Highly variable regions of tRNA sequences, namely the D- and T-domains, were evaluated in the program MFOLD v3.1 (Zuker 2003; http://bioinfo.math.rpi.edu/~zukerm/), which folds RNA sequences based on free energy minimizations (Mathews et al. 1999; Zuker, Mathews, and Turner 1999). These free energy based predictions were used to facilitate the search for potential base-pairing stems, which were confirmed only by the presence of compensatory base changes (CBC) across a majority of taxa in my alignment. Regions in which positional homology assessments could not be determined across all taxa were defined according to structural criteria of Kjer (1997), with modifications following Gillespie (2004). For reviews regarding rRNA sequence alignment, see Schultes, Hraber, and LaBean (1999) and Hancock & Vogler (2000). The $tRNA^{Lys}$-like sequence for the nine metatherian taxa

presented were considered separately due to their unusual structure (Janke et al. 1994, 1997; Dörner et al. 2001; Nilsson et al. 2003). This alignment is available upon request.

Scripted Manipulation of Data

The concatenated structural alignment consisted of 1693 unambiguously aligned nucleotides. This was converted to a PAUP* executable NEXUS file (Swofford 2002). A separate index file was created for the identification of each white-space delimited block and each stem-pair in the alignment. The alignment was annotated with a stem-pairing mask equivalent to that utilized in the program PHASE v1.1 (Jow et al. 2002; Hudelot et al. 2003). The mask distinguishes between pairing and non-pairing regions of the sequences. Ambiguously aligned regions were enclosed within brackets. I used Perl scripts (scripts, source code, and descriptions are available at *jRNA* website) that parse the alignment, index file, and a pairing statement (equivalent to the alignment mask) and return various input file formats (HTML formatted, color-highlighted alignments, summary statistics on base-pair composition and covariation, base-pair frequency tables, column and region base composition, nucleotide composition for non-pairing sites, and loop nucleotide range values). All correlation coefficients were calculated using Minitab v10.51 Xtra (Minitab, Inc.). Transition:transversion (TS:TV) ratios were estimated in MacClade v4.05 (Maddison and Maddison 2002), for all unambiguously aligned positions across a composite tree based on topologies presented by Arnason, Gullberg, and Janke (2004), Gibson et al. (2005), and Rowe (2003) (Nexus tree file available in Appendix 4).

For each tRNA, an inferred secondary structure diagram was determined for both non-rodent mammals and rodents. On all structure diagrams, putative tertiary interactions are shown. Structural diagrams were drawn manually and follow a simplified version of the convention of Helm et al. (2000).

PCR and Sequencing for Additional Rodent tRNAs

To further explore variation of mt tRNAs among rodents, sequences were determined for 11 of the 22 tRNAs for 32 additional rodent taxa (Appendix 2). Total genomic DNA was isolated from frozen liver or skeletal muscle tissue by proteinase-K digestion, followed by phenol/chloroform extraction (Sambrook, Fritsch, and Maniatis 1989). In some cases, purified mtDNA was available from previous studies, e.g., Allard and Honeycutt (1992). The polymerase chain reaction (PCR) was used to amplify three clusters of mitochondrial tRNAs as described in Kumazawa and Nishida (1993): a 527 base pair (bp) fragment bounded by the ND1 and ND2 genes, amplifying $tRNA^{Ile}$, $tRNA^{Gly}$, and $tRNA^{Met}$; a ~700 bp fragment bounded by the ND2 and COI genes, amplifying $tRNA^{Trp}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, and $tRNA^{Tyr}$; and a ~950 bp fragment bounded by the ND4 and ND5 genes, amplifying $tRNA^{His}$, $tRNA^{Ser(AGY)}$, and $tRNA^{Leu(CUN)}$. Amplification was performed using universal primer pairs: AH4641 and AL4160, BH5937 and BL5347, and DH12625 and DL11778, respectively (Kumazawa and Nishida 1993). Reaction conditions consisted of an initial denaturation at 94ºC for five min, followed by 30 cycles of a 94ºC for 30 sec, 55ºC for 1 min, and 72ºC for 30 sec, with a final extension at 72ºC degrees for 10 min. PCR amplifications were purified

by QIAquick Spin PCR purification spin columns, following a standard protocol (Qiagen Inc., Valencia, CA). Both strands of the PCR product were sequenced using the initial PCR primers, allowing confirmed full coverage of the tRNA genes. Cycle sequencing reactions were performed using ABI Prism BigDye Terminator chemistry (PE Applied Biosystems Inc., Foster City, CA) and included 25 cycles of 97ºC for 30 sec, 50ºC for five sec, and 60ºC for two min. Excess terminator dye, oligonucleotides, and polymerase were removed by centrifugation at 3000 rpm through a Sephadex G-50 matrix (Sigma Co.). Sequencing reactions were analyzed on an ABI 377 automated sequencer. Sequence data were imported into Sequencer v3.0–4.0 (GeneCodes Inc.) for alignment and annotation of each tRNA for each new specimen. The new rodent sequences were added to their respective alignments for analysis.

RESULTS

Sequence Mining and Alignment

Based on complete alignments, annotated sequences from 89 taxa required editing to conform to the 5' and 3' end of the acceptor stem of each tRNA gene (for a list of edits, contact the corresponding author). For 73 taxa (138 sequences), nucleotides were removed from the annotated sequences and for 62 taxa (103 sequences), nucleotides were added to the annotated sequences by direct examination of the complete genome sequences in GenBank. In 2 cases (the $tRNA^{Gly}$ of *Urotrichus talpoides*: accession AB099483 and $tRNA^{Lys}$ of *Macaca mulatta*: AY612638), the sequences were incorrectly annotated in GenBank. Through pair-wise alignment with

closely related taxa (*Mogera wogura* and *Macaca sylvanus*, respectively), the correct sequences were determined and added to the alignments. For several tRNAs (*Bos indicus*: $tRNA^{Asp}$; *Thylamys elegans*: $tRNA^{Phe}$, $tRNA^{Pro}$; *Dromiciops gliroides*: $tRNA^{Phe}$, $tRNA^{Pro}$; *Arctocephalus forsteri*: $tRNA^{Pro}$), only partial sequences were originally submitted at publication.

Additional Rodent tRNA Sequences

The sequencing effort focused on three major clusters of tRNAs (regions a (IQM), b (WANCY), and d (HSL): Kawazawa and Nishida 1993), resulting in sequences for 11 tRNAs of 32 additional rodent taxa, sampling 26 of the 32 extant families (Nowak 1999). The 11 tRNAs consist of 5 "heavy" tRNAs (encoded by the light strand) and 6 "light" tRNAs, including $tRNA^{Ser(AGY)}$ which bears a D-replacement loop (de Bruijn and Klug 1983; Yokogawa et al. 1991; Helm et al. 2000).

tRNA Secondary Structure

Secondary structure diagrams for both non-rodent mammals and rodents for each tRNA species illustrate the differences between the taxon-based models (fig. 3.1). My structural models deviate from the standardized secondary structure of Sprinzl et al. (1998), adopted by Kumazawa and Nishida (1993) and Helm et al. (2000), to conform to a minimum number of 3 nucleotides in the formation of hairpin loops, as hairpins of two

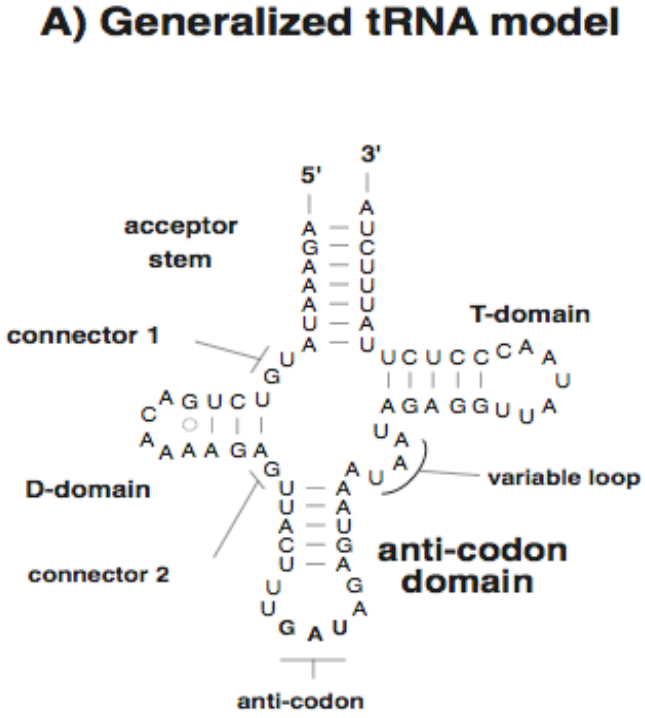


FIG. 3.1. —Consensus secondary structure diagrams of mitochondrial tRNAs (inferred from the alignment of 100 mammalian and up to 41 rodent taxa, left and right respectively). tRNAs are listed in their order along the molecule: (*A*) Generalized tRNA model; (*B*) H-strand encoded tRNAs; (*C*) L-strand encoded tRNAs. Numbering along each tRNA follows Sprinzl et al. (1998). Positions show the consensus of all nucleotides present at > 10% and follow IUPAC-IUB nomenclature (Moss 2005). Positions with conserved purines or pyrimidines are designated by a square and positions where transversions were present are designated by a circle. Secondary interactions follow the convention of Cannone et al. (2002), with bars for Watson-Crick pairing, small dark dots for G-U/U-G pairing (if > 50% base-pairs), open circles for G-A/A-G pairing (if > 50% base-pairs), or dark circles for other classes of mismatch pairing (if > 50% of base-pairs). Dark arrows represent insertion events and open arrows represent deletion events present in some sequences. Dashed lines represent potential tertiary interactions. Variation in the D- and T-loops is indicated by the total range observed in each loop (includes nucleotide positions shown).

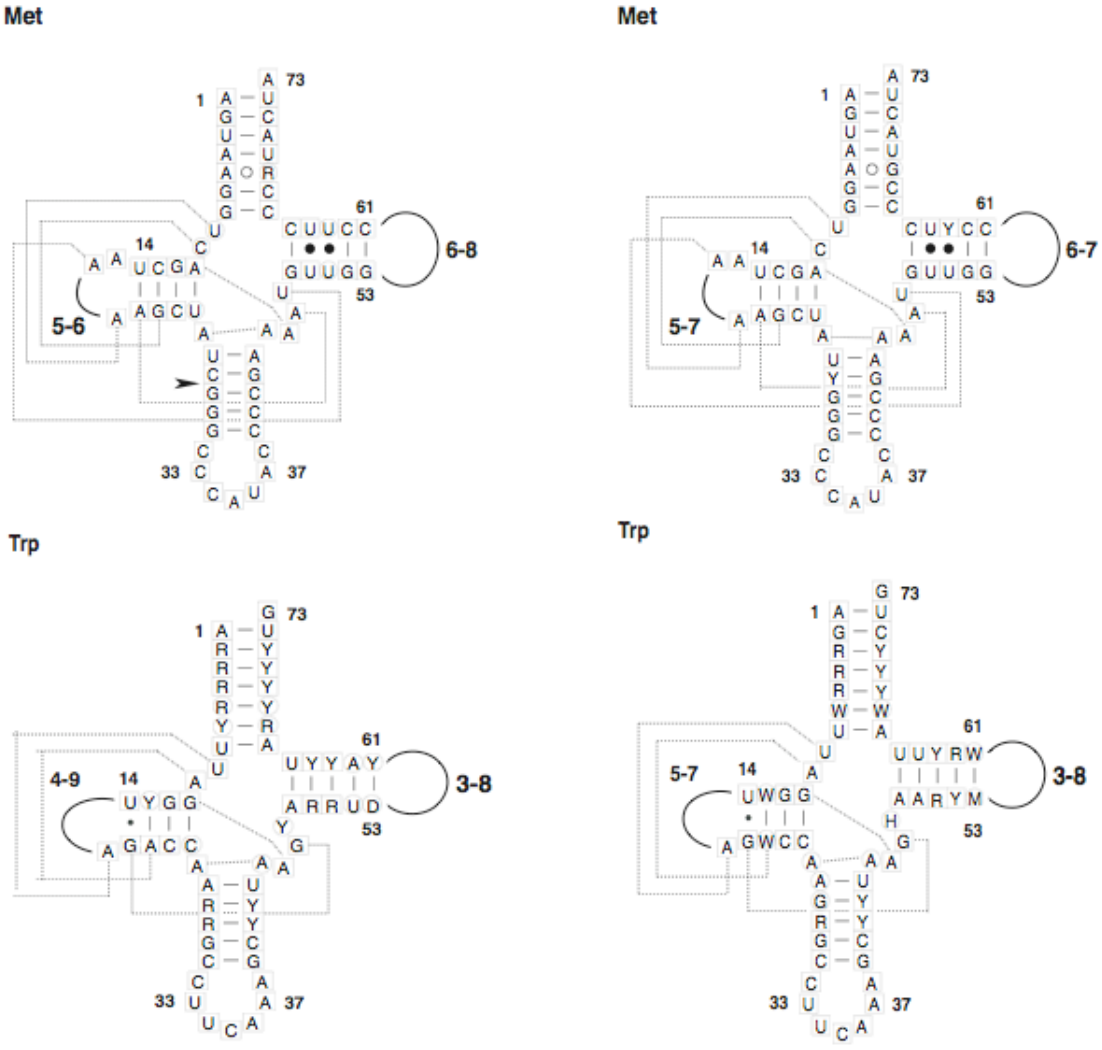


FIG. 3.1. —continued.

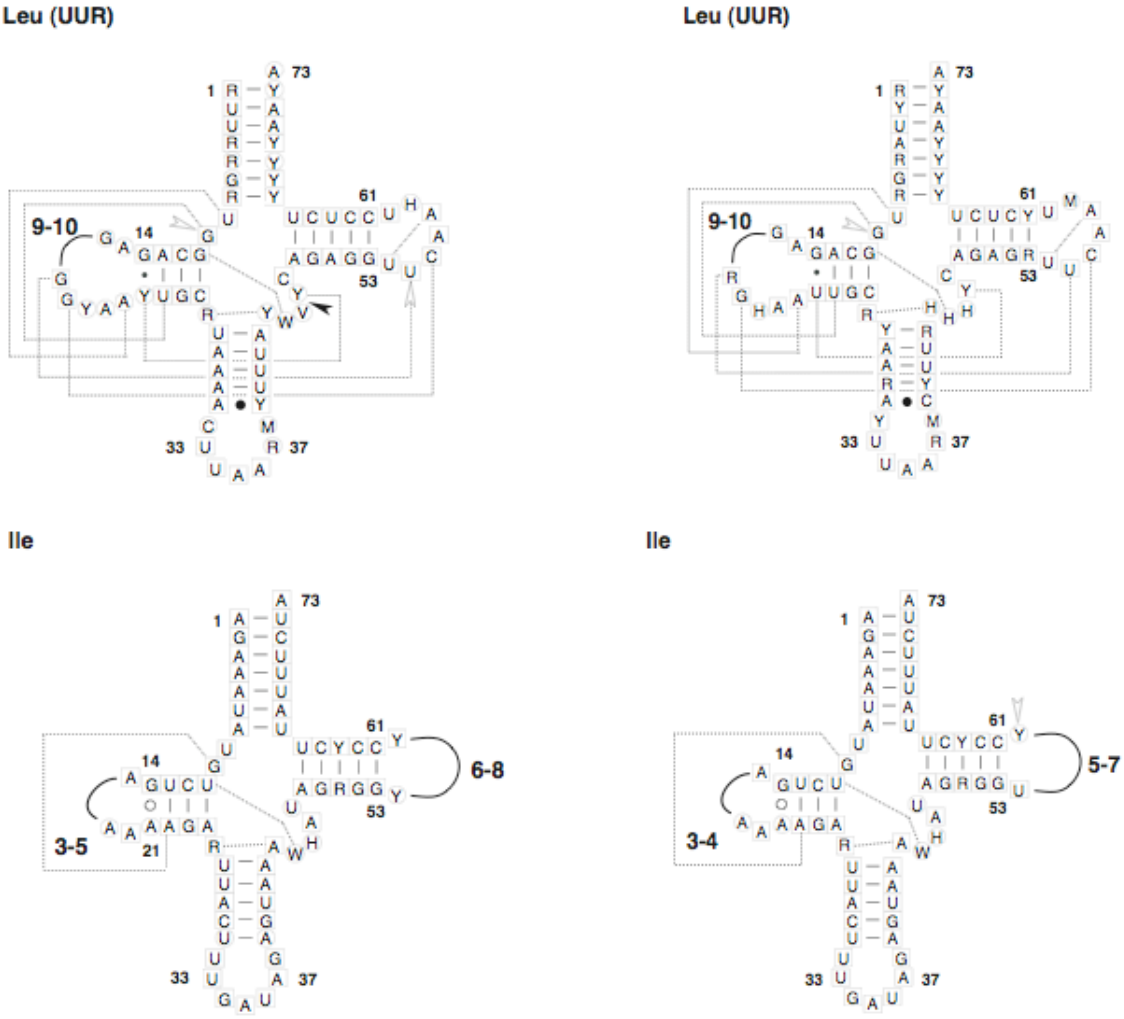


FIG. 3.1. —continued.

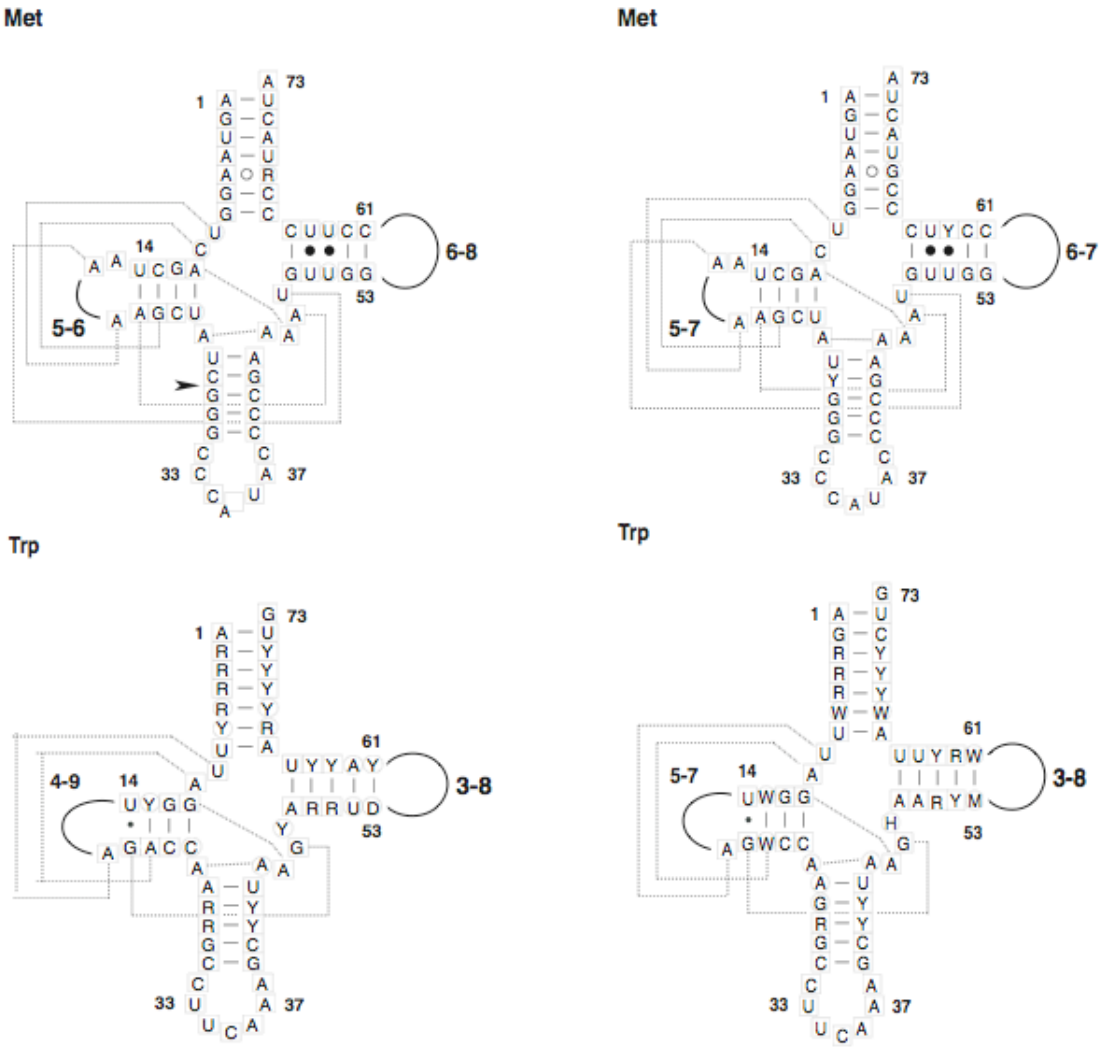


FIG. 3.1. —continued.

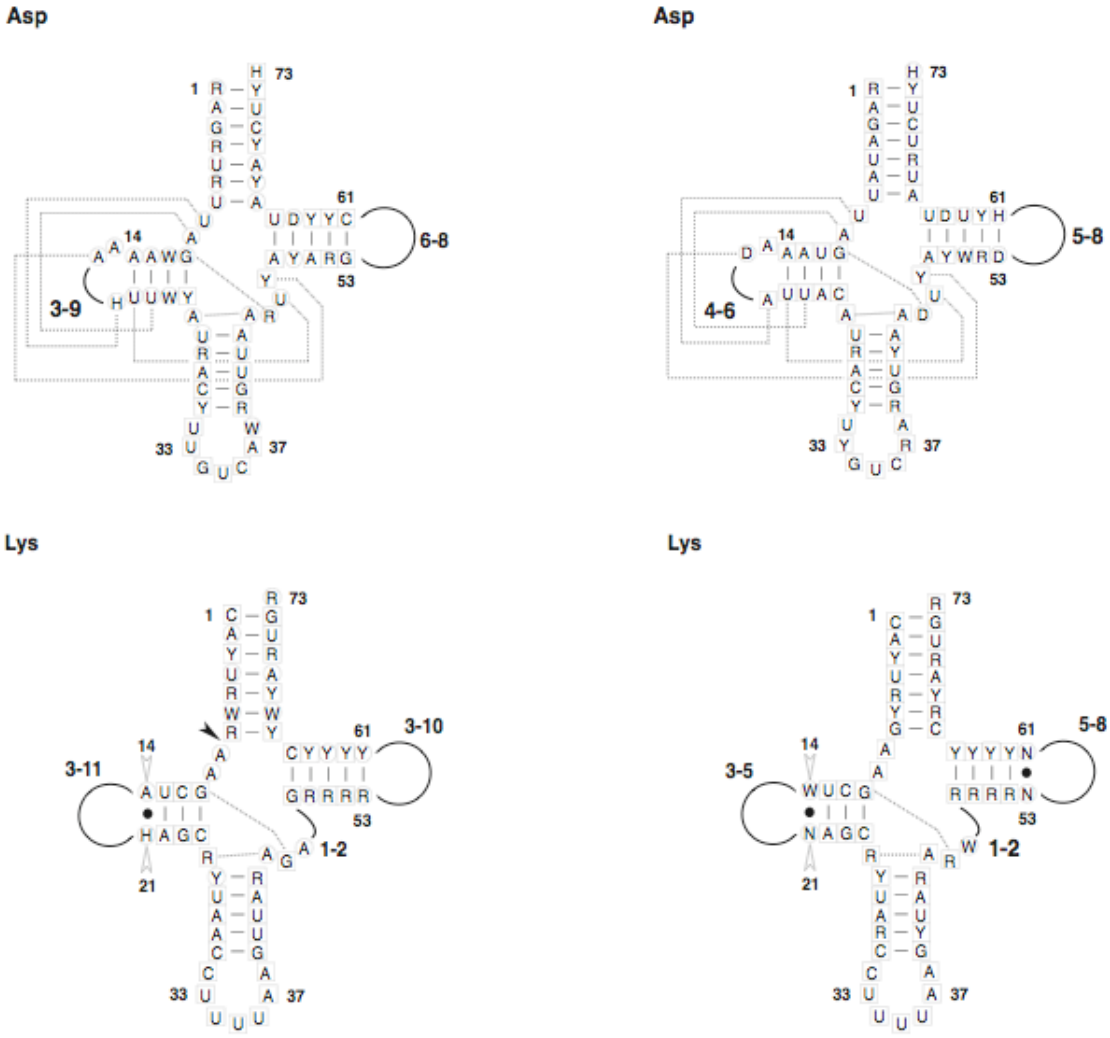


FIG. 3.1. —continued.

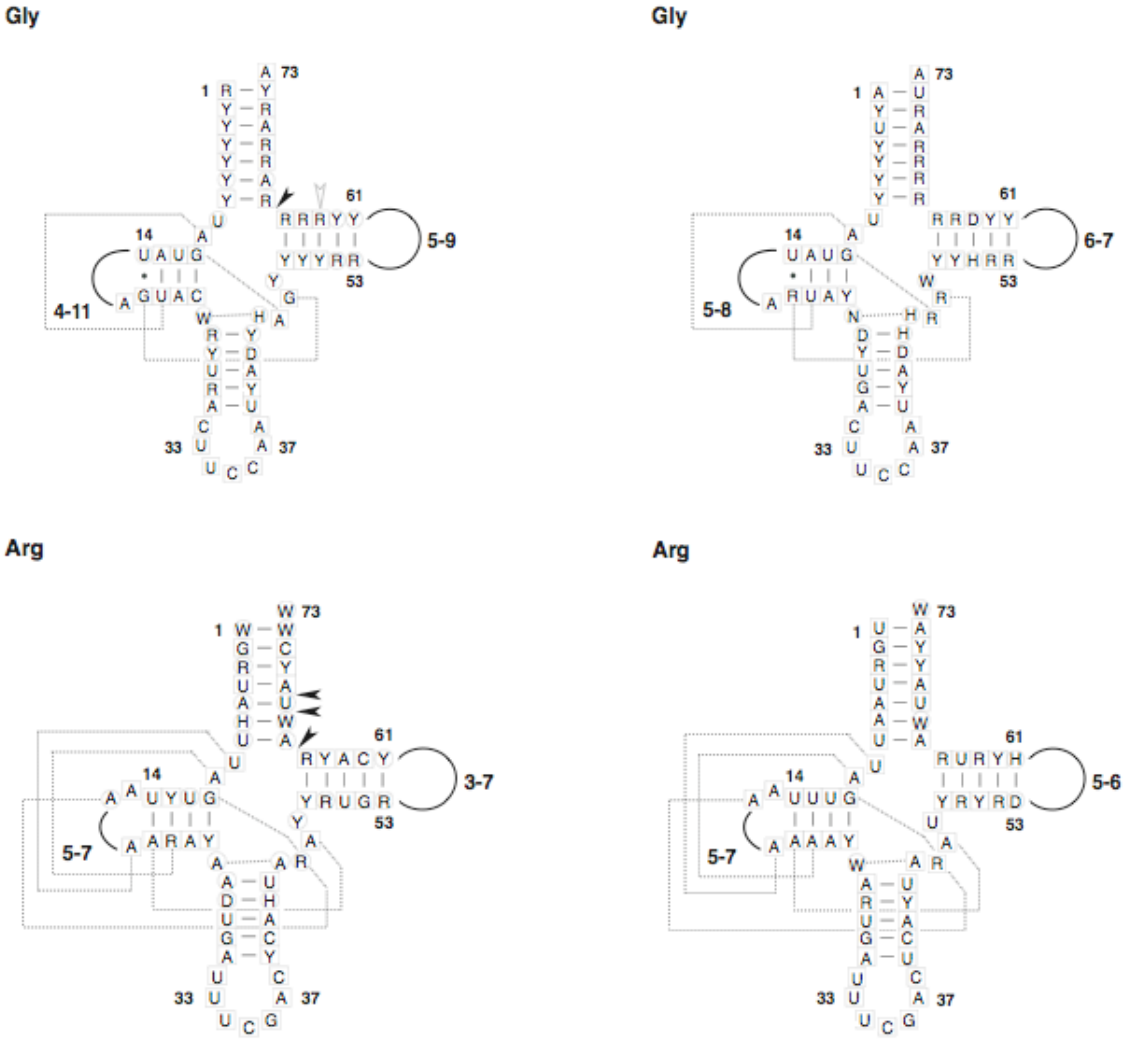


FIG. 3.1. —continued.

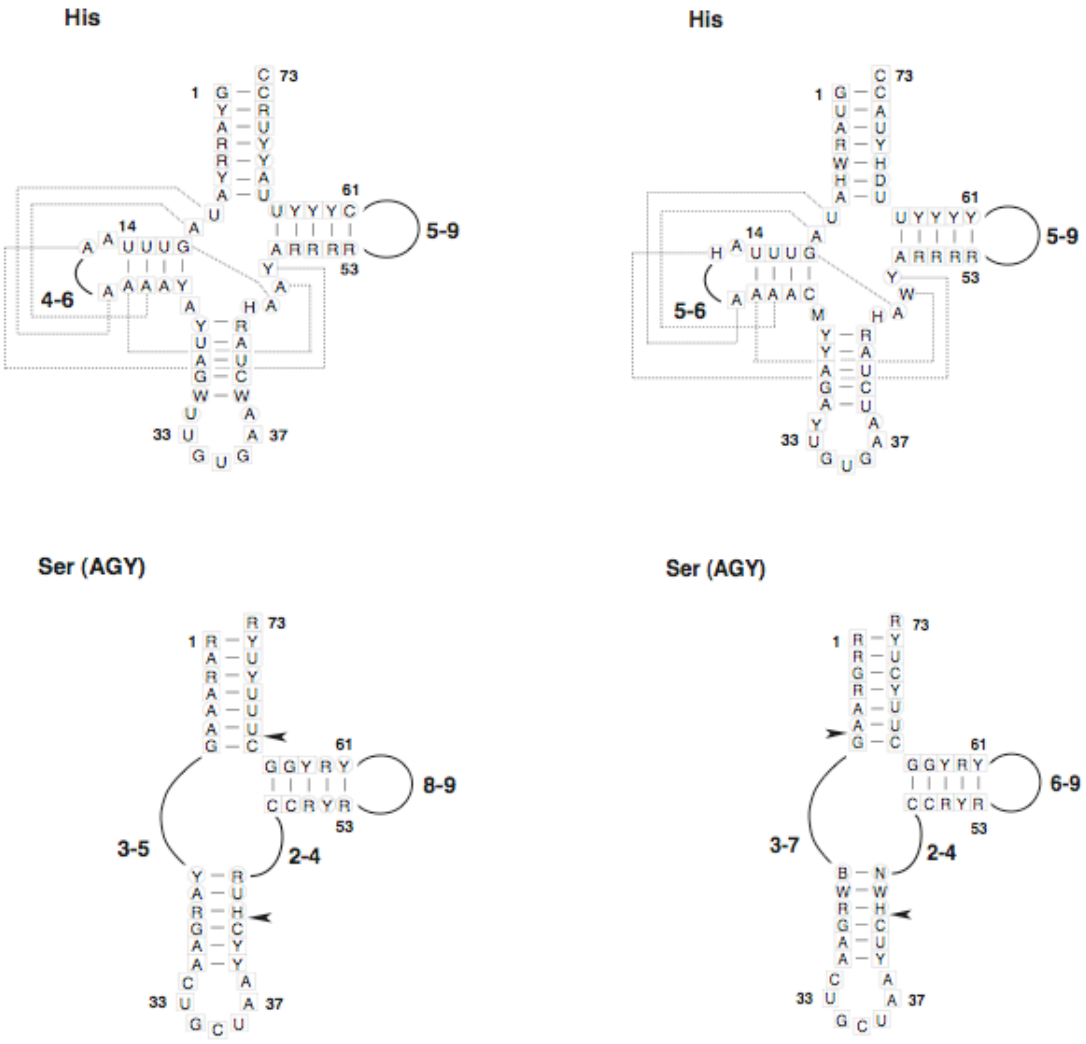


FIG. 3.1. —continued.

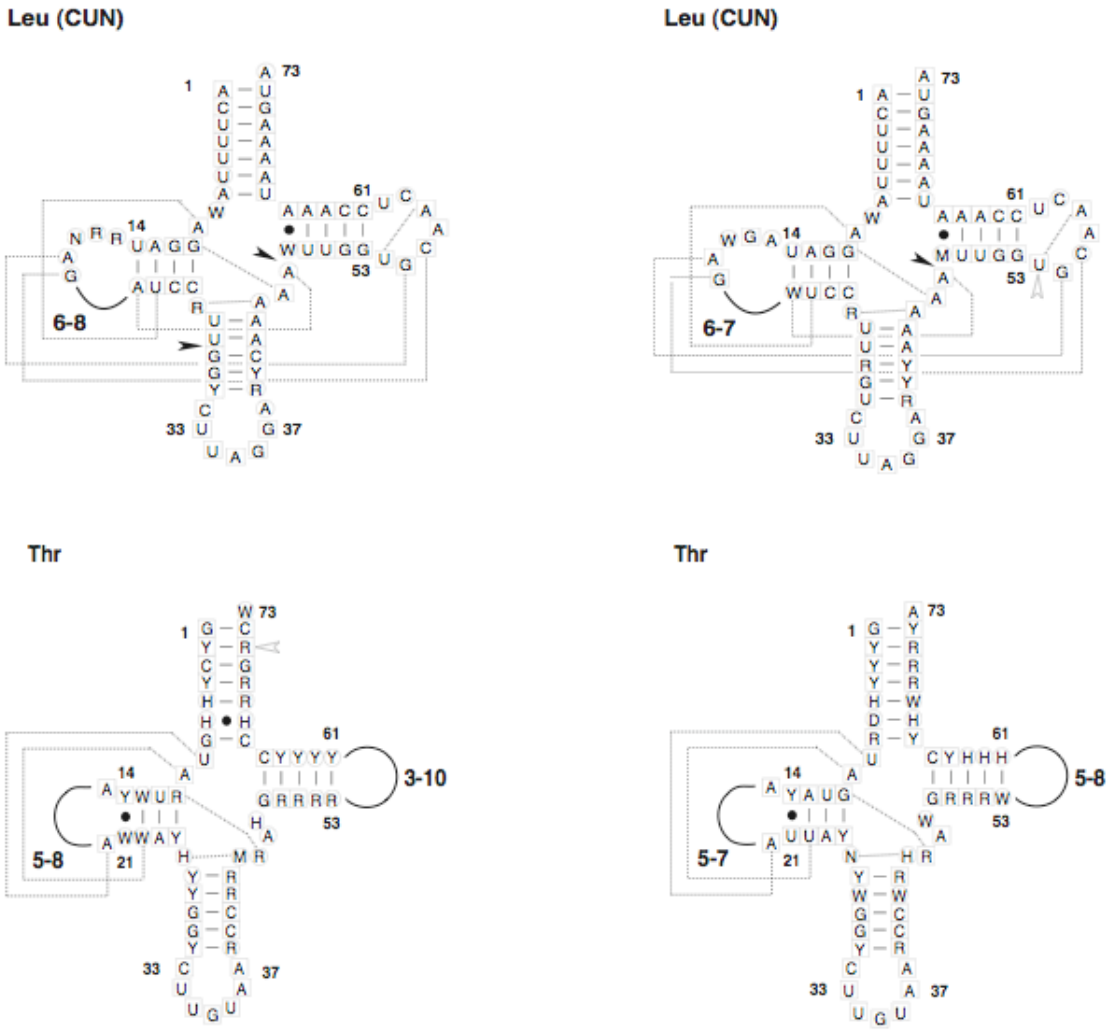


FIG. 3.1. —continued.

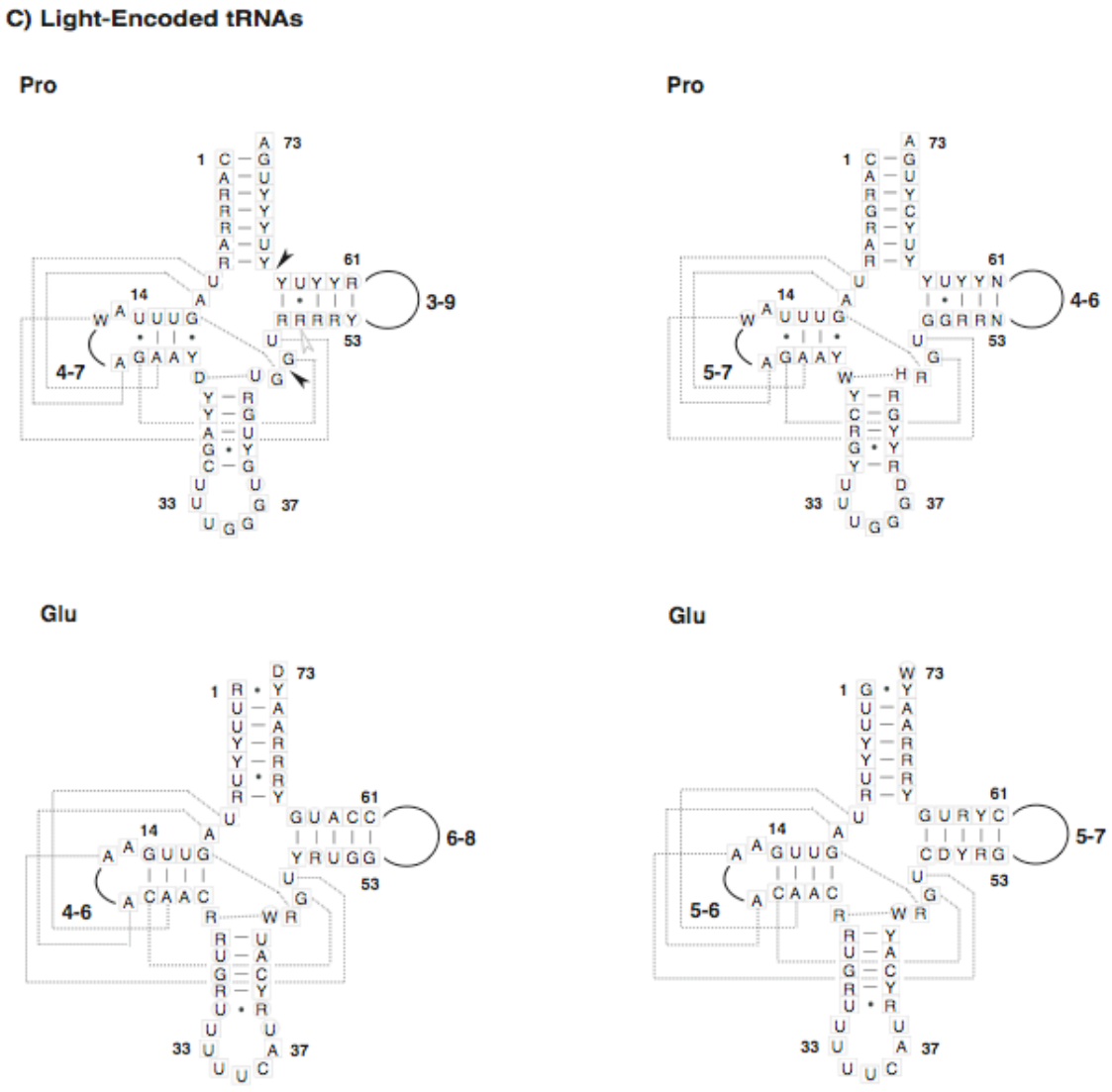


FIG. 3.1. —continued.

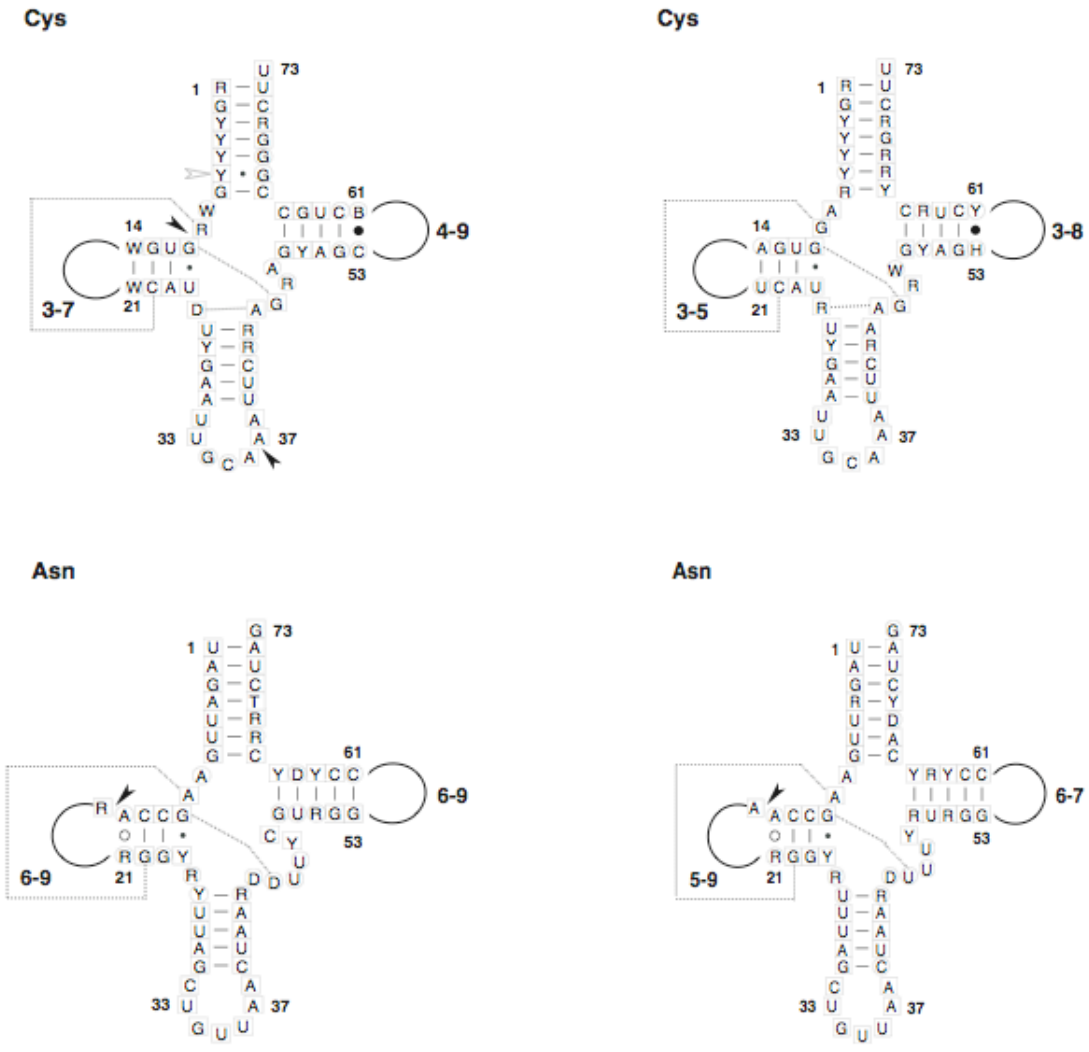


FIG. 3.1. —continued.

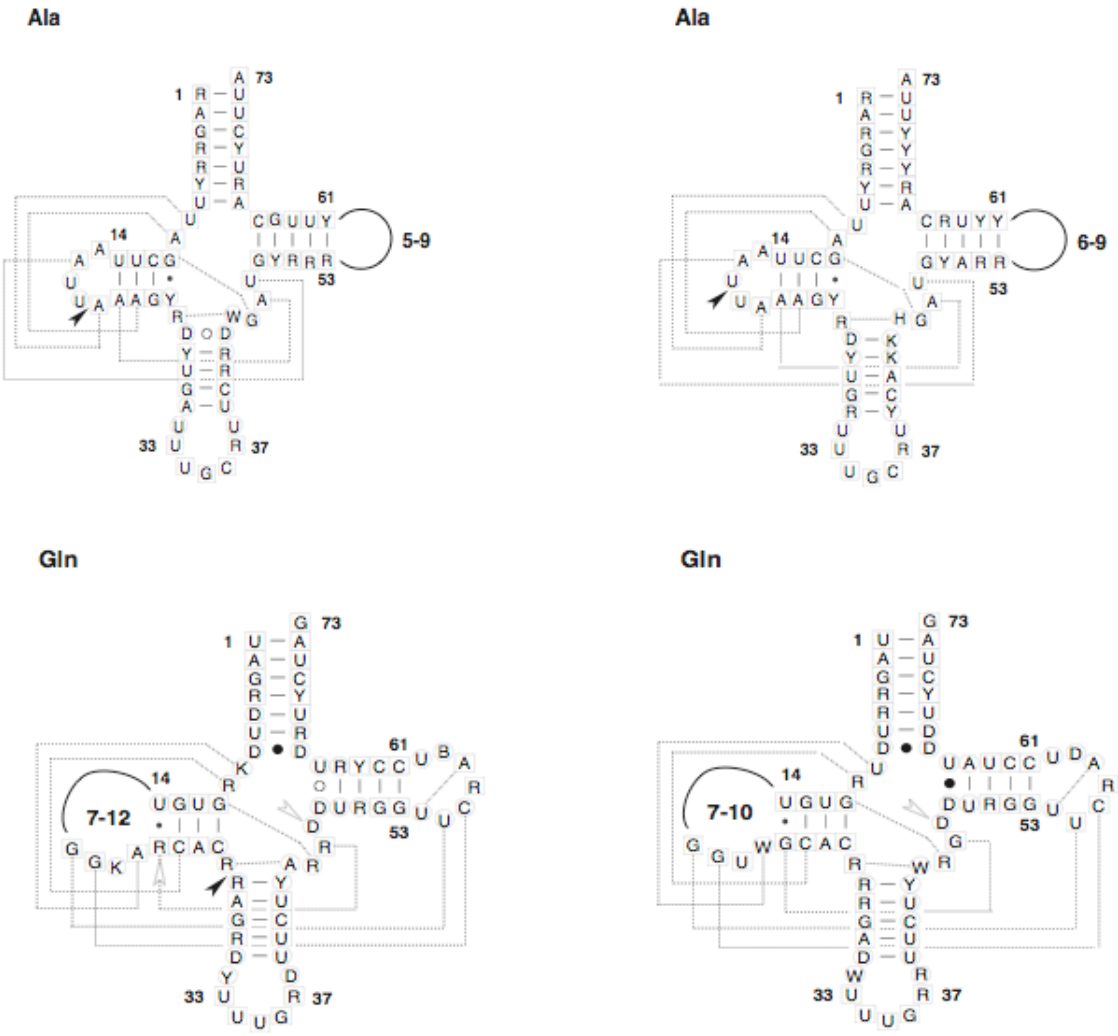


FIG. 3.1. —continued

bases are rare even under laboratory conditions (Groebe and Uhlenbeck 1988). This results in stem base-pair formulae that differ from those proposed in standardized models (seven, four, five, and five base-pairs for the acceptor-stem, D-stem, anti-codon-stem, and T-stem, respectively).

Nucleotide Composition of H- and L-Encoded tRNAs

Unlike protein-coding genes of the mammalian mitochondrial genome, where less than 5% of the nucleotides (ND6 only) are L-encoded, over one-third of the mitochondrial tRNA genes (8 of 22) are L-encoded. To better reflect the nucleotide composition of tRNAs, all analyses and statistics reported are based on the sense strand for each tRNA gene. Nucleotide composition for all sites showed significant variation in both H- and L-encoded tRNA genes (fig. 3.2). For all positions (stems and loops), the proportion of positions that showed only transitions ranged from 0.47 for $tRNA^{Asp}$ to 0.91 for $tRNA^{Met}$, both H-strand encoded tRNAs (data not shown).

The mean overall base composition of H-strand tRNA genes was A = 0.36 (33–40%), C = 0.19 (16–23%), G = 0.16 (13–18%), T = 0.29 (26–32%), with %GC = 0.35 (31–40%) (fig. 3.2). Following the convention of Gibson et al (2005), figure 3.3 shows the base composition of all H-encoded sites for the 141 taxa examined, sorted by increasing levels of C. There is a negative correlation between the change in percent C and the percentages of A and T (r = -0.475, $P < 0.001$ and r = -0.906, $P < 0.001$, respectively) (fig. 3.3*A*). There is also a weak positive correlation between the

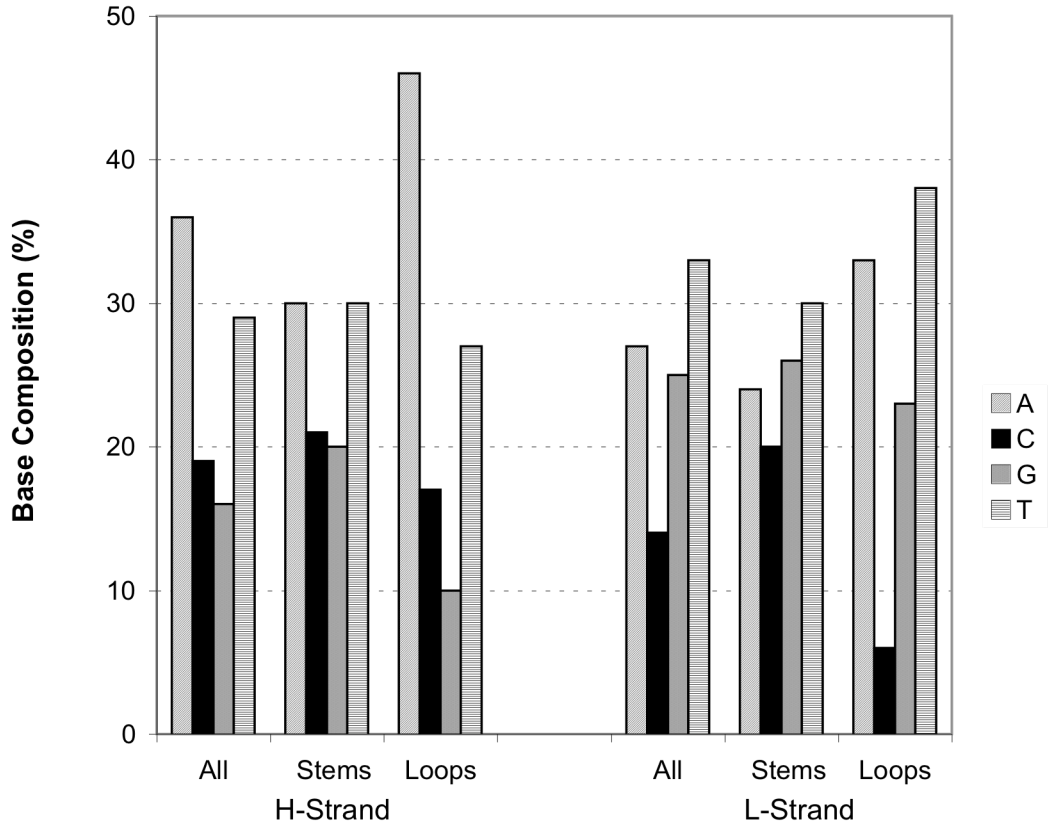


FIG. 3.2. —Observed nucleotide composition bias across all positions, stem positions, and loop positions for mammalian mt tRNA genes. tRNA genes are separated by their coding strand.

percentage of C and the percentage of G ($r = 0.244$, $P < 0.05$) (fig 3.3*A*.). Among the rodent taxa, the correlation between %C and %G is stronger than among non-rodents ($r = 0.448$, $P < 0.01$) (data not shown). The TS:TV ratio estimated for all unambiguous sites of H-encoded tRNAs was 2.12.

For the L-strand tRNA genes, the variation is more extreme with a mean overall base composition of A = 0.27 (21–33%), C = 0.14 (10–18%), G = 0.25 (20–30%), T = 0.33 (30–39%), with %GC = 0.39 (33–46%) (fig. 3.2 and Appendix). As with the H-strand tRNA genes, L-encoded tRNAs exhibit a strong negative correlation between change in %C and both %A or %T ($r = -0.689$ and $r = -0.478$, both at $P < 0.001$) and a strong positive correlation between the %C and %G ($r = 0.509$, $P < 0.001$) (fig. 3.3*B*). The TS:TV ratio for all unambiguous sites of L-encoded tRNAs was higher (2.57) than observed for H-encoded tRNAs.

Mean %GC observed in the H-encoded tRNA genes is consistent with that reported for all eukaryotic mitochondrial tRNAs (%GC = 0.34, Higgs 2000), while the mean %GC of the L-strand tRNA genes is 4–5% higher. The bias in nucleotide composition of each strand (as %GC) is influenced differently. In the H-encoded tRNAs, %GC is strongly correlated with %C (mean $r = 0.903$, $P < 0.001$), while in the L-encoded, %GC is strongly correlated with %G (mean $r = 0.935$, $P < 0.001$). Figure 3.4 shows biases in the distribution of nucleotides between the two DNA strands, measured as AT- and GC-skews calculated by the following formulae: AT-skew = (A-T)/(A+T); GC-skew = (G-C)/(G+C) (Perna and Kocher 1995).

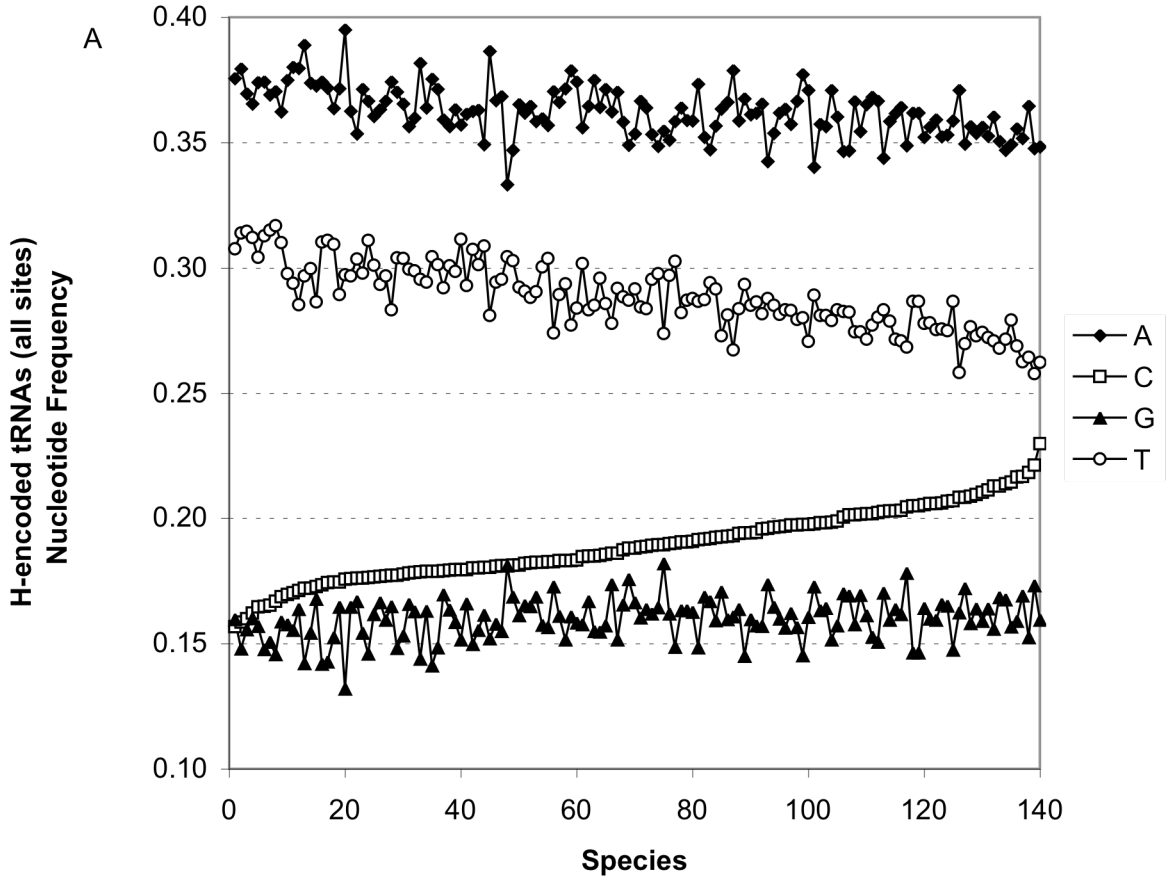


FIG. 3.3. —Observed nucleotide composition among all sites for each of the 141 taxa in this study. The species are ordered by increasing percentage of C: (*A*) H-encoded tRNAs; (*B*) L-encoded tRNAs.

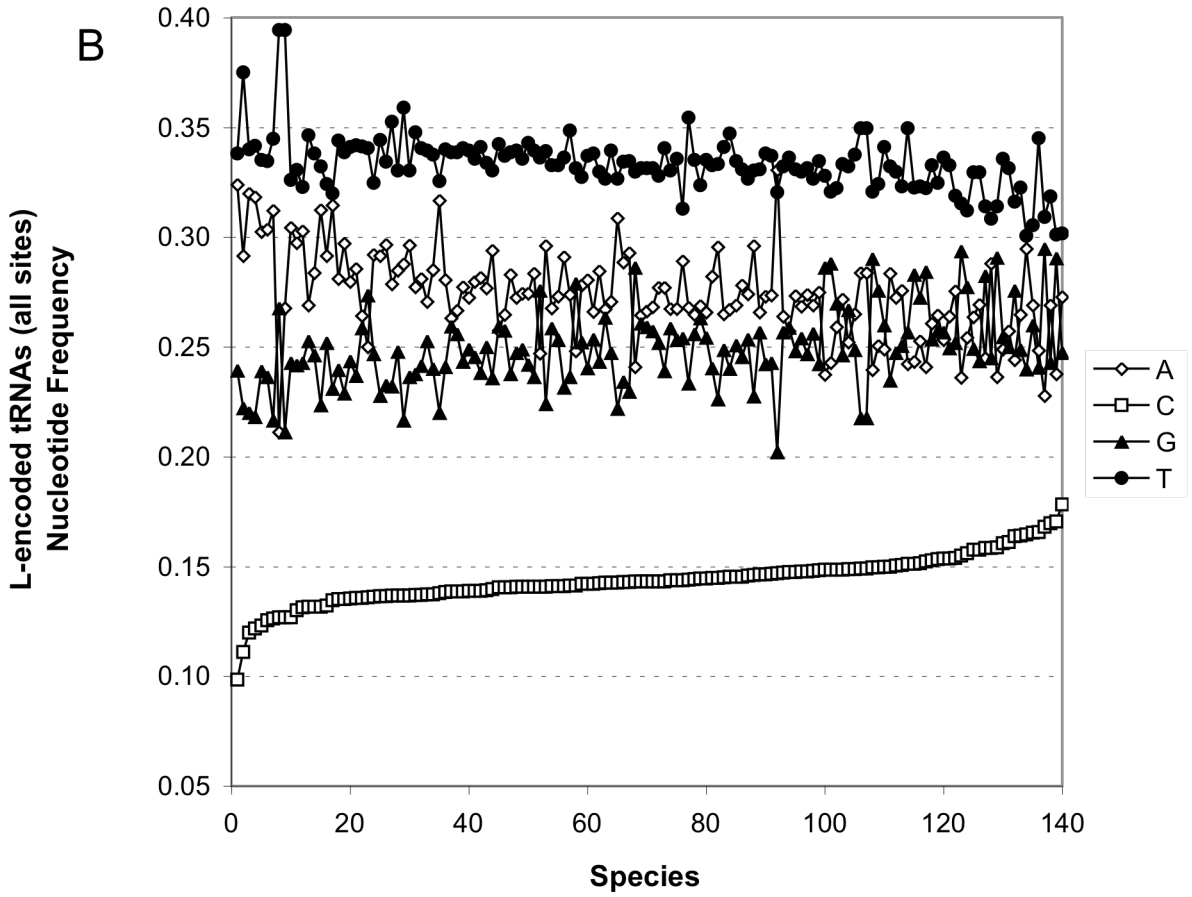


FIG. 3.3. —continued.

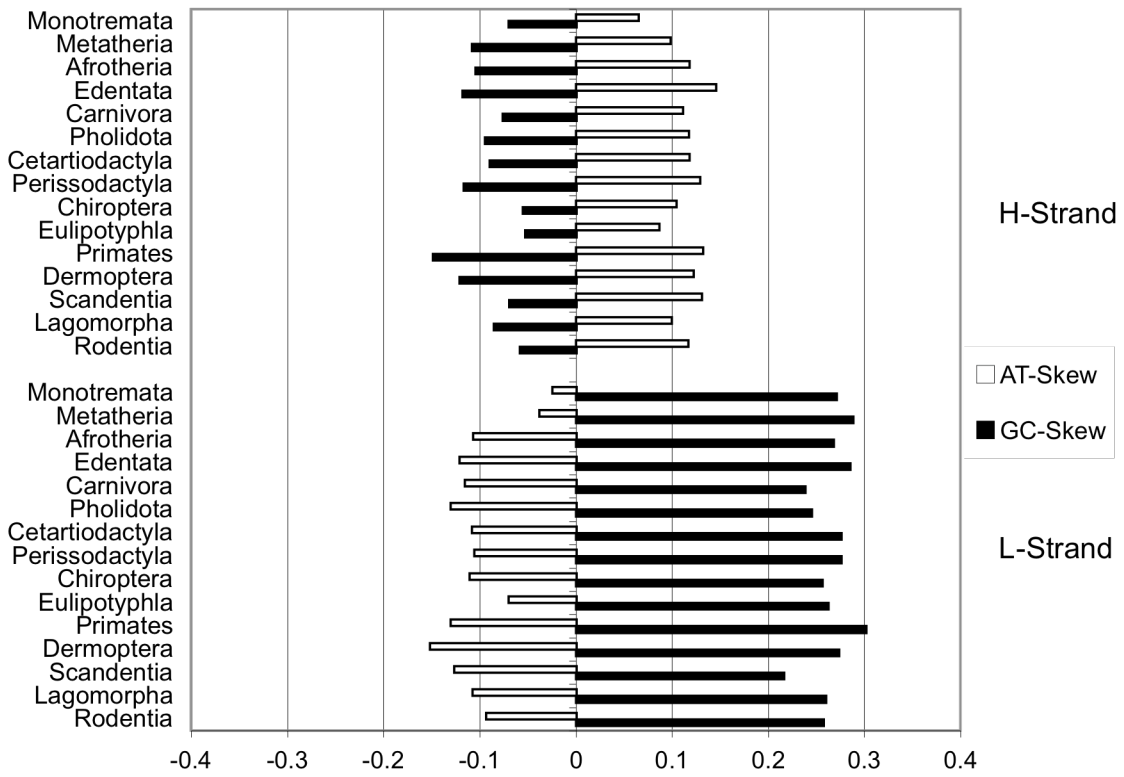


FIG. 3.4. —GC- and AT-skew observed among all sites for the tRNA genes of each DNA strand for several mammalian orders. Values for H-encoded tRNA genes above, L-encoded tRNA genes below. The superorder Afrotheria includes the following 6 orders: Afrosoricida, Hyracoidea, Macroscelidea, Proboscidea, Sirenia, and Tubulidentata.

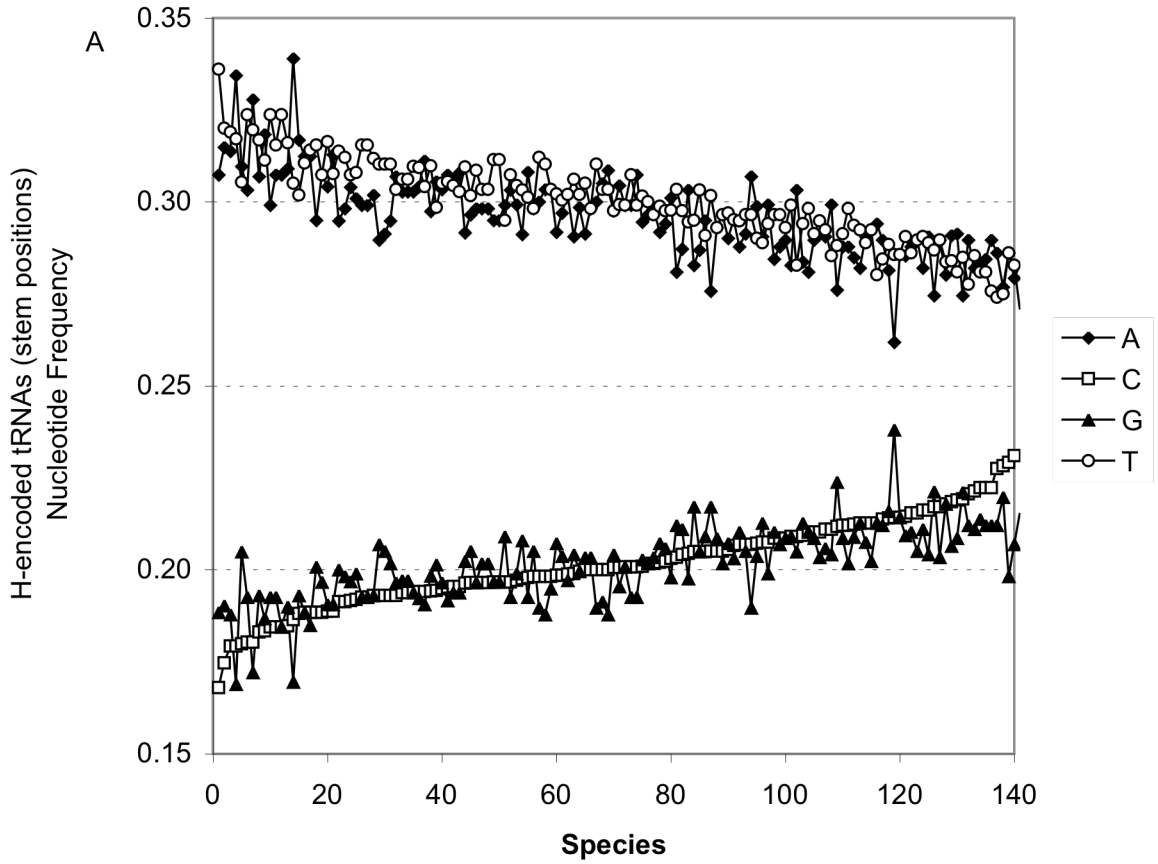


FIG. 3.5. —Observed nucleotide composition among stem positions for each of the 141 taxa in this study. The species are ordered by increasing percentage of C: (*A*) H-encoded tRNAs; (*B*) L-encoded tRNAs.

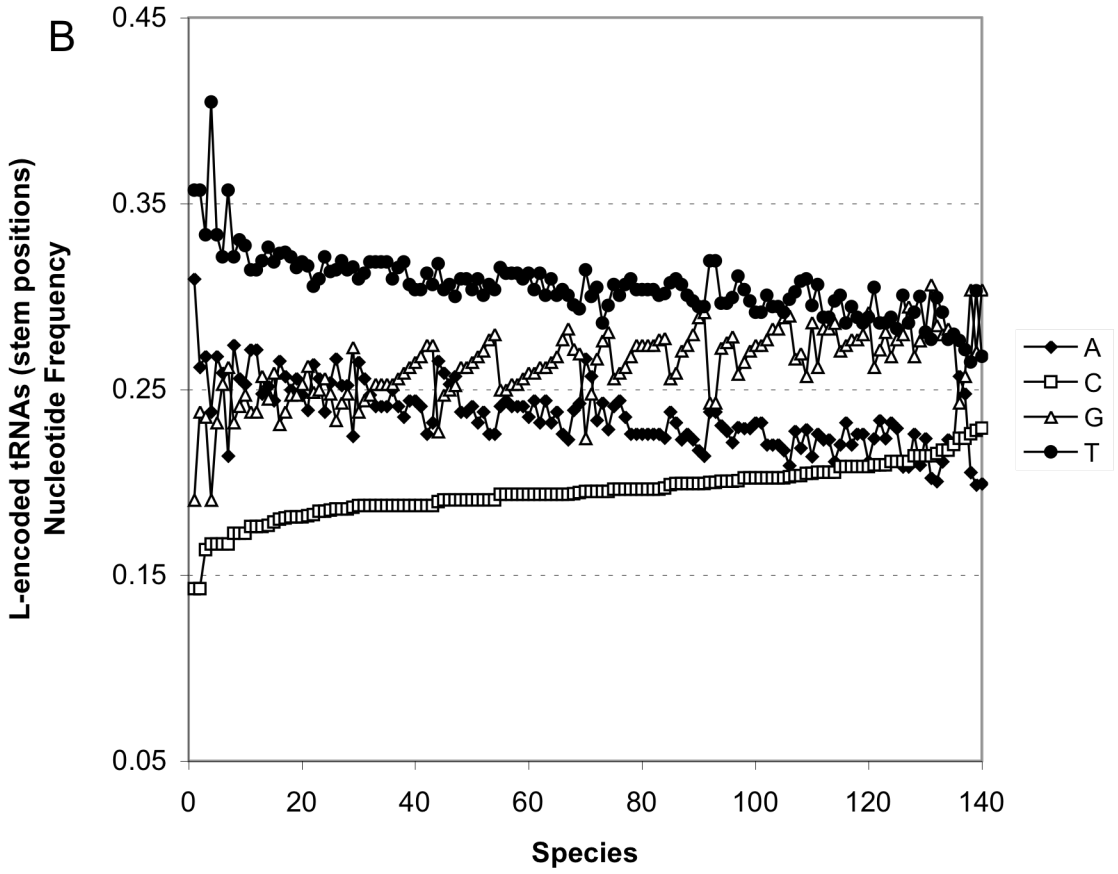


FIG. 3.5. — continued.

Stem Positions

Nucleotide Composition

Base composition of H-encoded stem regions does not meet the prediction of low G content and high A content, as observed across the entire molecule in previous studies (Reyes et al. 1998; Boore 1999), with average frequencies of: A = 0.30 (26–34%), C = 0.20 (17–23%), G = 0.20 (17–24%), and T = 0.30 (27–34%), with %GC = 0.40 (35–45%) (fig. 3.5*A*). There is a strongly significant negative correlation between the change in percent C and the percentage of both A and T (r = -0.789 and r = -0.940, respectively, both at $P < 0.001$) and a strong positive correlation between percentage of C and G (r = 0.701, $P < 0.001$). Change in %A and %T have nearly identical slopes, as do the slopes of change in %C and %G (fig. 3.5*A*).

For L-strand encoded tRNA helices, patterns of nucleotide frequency were quite different (fig. 3.5*B*). Average base frequencies were: A = 0.24 (20–31%), C = 0.20 (14–23%), G = 0.26 (19–31%), and T = 0.30 (27–41%), with %GC = 0.45 (33–53%) (fig. 3.5*B*). Again, percentage of C is significantly negatively correlated with both A and T and positively correlated with percentage of G (data not shown, all $P < 0.001$), but with very different patterns from those observed for H-strand encoded helices (fig. 3.5*A*). In the H-encoded helices, the %GC is correlated %G in rodents and correlated with %C in non-rodents (r = 0.950 and 0.945, respectively). In L-encoded helices of both rodents and non-rodents, %GC is correlated with %C (r = 0.911, and 0.954, respectively). L-encoded stems showed moderate to high proportions of transition-only sites (0.63 to 0.88), while the H-encoded tRNAs show a large variation, with highly conserved

tRNAs, such as $tRNA^{Met}$ and $tRNA^{Ile}$ showing very high ratios (0.93 and 0.92, respectively) and, at the other extreme, $tRNA^{Asp}$ with very low proportion of transition-only sites (0.49). The mean TS:TV ratio for H- and L-encoded stem positions were markedly different: 5.86 and 7.82, respectively. Both the AT- and GC-skew observed in the stem positions of the H-encoded tRNA genes showed no obvious pattern, with all values not significantly different from zero (fig. 3.6). This is in drastic contrast to the pattern observed in L-encoded tRNA stem positions, where the AT-skew ($\bar{x}$ = -0.13) has the same direction and magnitude as that observed for all L-strand tRNA sites. The GC-skew of the L-encoded stem positions has the same direction as that observed for all L-encoded tRNA sites, but with a lower magnitude ($\bar{x}$ = 0.15) (fig. 3.6).

Base-Pair Composition in tRNA Stems

The cumulative mean base-pair compositions for all stem regions of the 22 mammalian mt tRNAs are shown in Table 3.1. In the combined stem positions for all tRNAs (462 base-pairs), rodents had 187 invariant base-pair positions, yet showed frequency variation at 15 of the 83 invariant base-pair positions observed in non-rodent mammals. Of the variable base-pairs, non-rodents showed consistent distribution across both DNA strands with 35% of the variable pairs in the acceptor arm, 15% in the D-arm, 24% in the anti-codon arm, and 25% in the T-arm (Table 3.2). Variation in the base-pairs of rodents, however, differed among tRNAs and between the two strands, with H-encoded stems having 36% in the acceptor arm, 8% in the D-arm, 23% in the anti-codon arm, and 34% in the T-arm and L-encoded stems having 39% of the variable base-pairs

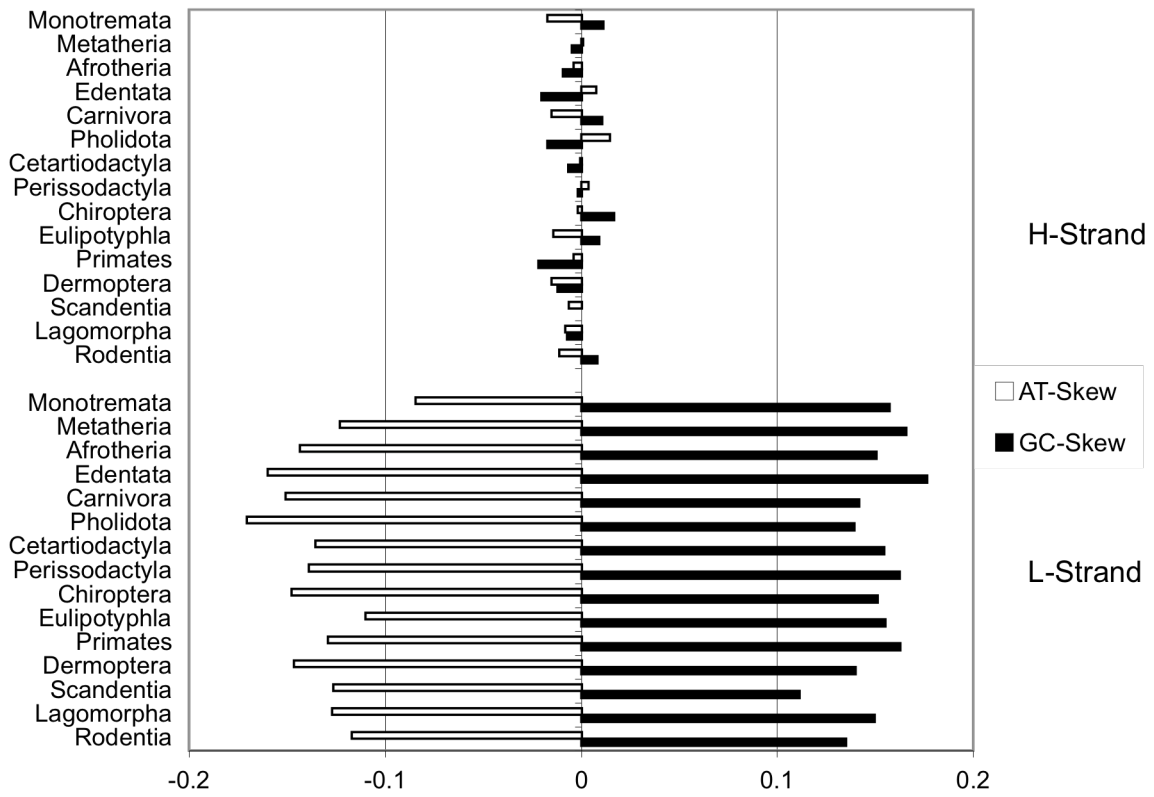


FIG. 3.6. —GC- and AT-skew observed among stem positions for the tRNA genes of each DNA strand for several mammalian orders. Values for H-encoded tRNA genes above, L-encoded tRNA genes below. The superorder Afrotheria includes the following 6 orders: Afrosoricida, Hyracoidea, Macroscelidea, Proboscidea, Sirenia, and Tubulidentata.

in the acceptor arm, 10% in the D-arm, 25% in the anti-codon arm, and 26% in the T-arm (Table 3.2).

The proportion of GU/UG in L-encoded tRNAs is substantially higher in both rodents and non-rodents than in the H-encoded tRNAs (Table 3.1). Among non-rodents, GU was present at 157 base-pair positions and UG at 125 base-pair positions. Of these, 53 (34%) and 36 (29%) respectively, were found at positions where otherwise the base-pairs strictly covary. Among rodent tRNAs, there were 86 positions with GU and 50 with UG base-pairs, with 33 (38%) and 27 (54%) at otherwise strictly covarying base-pairs positions. Among non-rodent mammals, four base-pair positions were invariant for non-Watson-Crick (WC) pairs (2 UG pairs and 2 mismatch pairs). Three of four of these mismatch pairs occur in D-arm positions (Ile G13-A22, Pro U13-G22, and Trp U13-G22), with the remaining mismatch pairs occurring in the T-arm of methionine (U50-U64). Rodent tRNAs share three of the four invariant mismatch pairs, lacking an invariant Ile G13-A22 pair. In addition, rodents exhibit four unique invariant mismatch pairs; two in D-arm positions (Leu^{UUR} G13-U22 and Thr U13-U22), one in T-arm (Pro G50-U64), and one in the anti-codon stem (Leu^{UUR} A30-C40). Overall mismatch frequencies were consistent between rodents and non-rodents (data not shown), but were different between H-encoded (6%) and L-encoded (3%) tRNAs (Table 3.1).

Table 3.1 Cumulative mean base-pair composition for the stem positions of the 22 mammalian mitochondrial tRNAs

	Watson-Crick pairs	GU/UG Pairs	Mismatch Pairs*
H-Encoded			
Phenylalanine	0.92	0.01	0.07
Valine	0.92	0.01	0.07
Leucine^{UUR}	0.90	0.06	0.04
Isoleucine	0.95	0.01	0.04
Methionine	0.86	0.00	0.14
Tryptophan	0.90	0.06	0.04
Aspartic Acid	0.95	0.03	0.02
Lysine	0.92	0.02	0.06
Glycine	0.91	0.06	0.03
Arginine	0.96	0.03	0.01
Histidine	0.96	0.01	0.03
Serine^{AGY}	0.94	0.01	0.05
Leucine^{CUN}	0.93	0.01	0.06
Threonine	0.87	0.02	0.11
mean	0.92	0.02	0.06
L-Encoded			
Proline	0.79	0.21	0.00
Glutamic Acid	0.87	0.12	0.01
Serine^{UCN}	0.82	0.17	0.01
Tyrosine	0.90	0.05	0.05
Cysteine	0.83	0.12	0.05
Asparagine	0.86	0.08	0.06
Alanine	0.79	0.16	0.05
Glutamine	0.80	0.13	0.07
mean	0.83	0.13	0.04

*Possible mismatch pairs are AA, AC, AG, CA, CC, CU, GA, GG, UC, UU.

Table 3.2 Number and proportion of variable base-pair positions by stem region for the H-encoded and L-encoded mitochondrial tRNAs

Non-Rodents				
	H-encoded		*L-encoded*	
Stem	Variable Positions	Proportion	Variable Positions	Proportion
A	87	0.35	46	0.36
D	38	0.15	20	0.16
A-C	60	0.24	31	0.24
T	61	0.25	31	0.24
Total	246		128	

Rodents				
	H-encoded		*L-encoded*	
Stem	Variable Positions	Proportion	Variable Positions	Proportion
A	60	0.36	39	0.39
D	13	0.08	10	0.1
A-C	39	0.23	25	0.25
T	57	0.34	26	0.26
Total	169		100	

Abbreviations: A = acceptor arm; D = dihyrouridine arm; A-C = anti-codon arm; T = T-pseudouridine arm.

Loop Positions

Nucleotide Composition in Loops

Average nucleotide composition of the combined loop regions (including ambiguous nucleotides) for H-encoded tRNA genes were: A = 0.46 (43–51%), C = 0.17 (11–24%), G = 0.10 (7–13%), and T = 0.27 (22–33%), with %GC = 27% (21–33%) (fig. 3.7*A*). Among non-rodents, there were significant negative correlations between the change in percentage of C and the percentage of T (r = -0.861, $P < 0.001$) and between change in %A and %G (r = -0.471, $P < 0.001$). In rodents, a different pattern emerges with a negative correlation between %C and %T (r = -0.542, $P < 0.001$), and moderately negative correlations between change in %A with %T (r = -0.451, $P < 0.01$) and %G (r = -0.451, $P < 0.01$). In all taxa, %GC is correlated with %C (r = 0.765 and r = 0.901 for rodents and non-rodents, respectively). The proportion of transition-only sites of unambiguously aligned H-encoded loops ranged from 0.41 in $tRNA^{Asp}$ to 1.0 in the "bizarre" $Serine^{AGY}$ tRNA, which shows no transversions in its loop positions ($\bar{x}$ = 0.73). The observed TS:TV ratio in unambiguously aligned positions of H-strand loops was 1.07.

Variation in L-encoded loop positions (including ambiguous nucleotides) were quite different from that observed in the H-encoded tRNAs (fig. 3.7*A*) with very low %C, and higher %G and %T (fig. 3.7*B*). Average frequencies were: A = 0.33 (21–43%), C = 0.07 (3–11%), G = 0.23 (16–31%), and T = 0.37 (31–45%), with %GC = 0.29 (22–37%). There are strong negative correlations between the change in percent A and both %G and %T (data not shown, all $P < 0.001$) and between change in %C and %T (r = -

0.303, $P < 0.01$). The proportion of transition-only sites observed in the unambiguously aligned loop positions of L-stranded tRNAs was moderate to high: 0.44 to 0.80 (mean = 0.67). The estimated TS:TV ratio of unambiguously aligned positions in L-encoded loops was 1.21, higher than observed in H-encoded loops.

Directionality of the AT- and GC-skews of H-strand loop nucleotides are reflective of the skew measures calculated from all tRNA sites of H-encoded tRNA genes, but the magnitudes are much larger ($\bar{x}$ = 0.26 and -0.29, respectively) (fig. 3.8). Skew measures of the L-strand loop are also consistent in directionality with those from all L-strand tRNA sites with AT-skews being slightly less negative ($\bar{x}$ = 0.06) and GC-skews being substantially higher ($\bar{x}$ = 0.55). Two glaring departures from these patterns are positive AT-skews for monotremes and marsupials (mean = 0.05 and 0.07, respectively) (fig. 3.8).

Loop Size

Table 3.3 shows mean loop sizes and standard errors for the Dihydrouridine (D) loop, variable (V) loop, and T-Pseudouridine (T) loop. Overall, the D-loop ranges from 3–12 nt ($\bar{x}$ = 5.6), the V-loop ranges from 2–6 nt ($\bar{x}$ = 4.1), and the T-loop ranges from 3–10 ($\bar{x}$ = 6.4). Again, I stress that some stems (following standard tRNA models) were truncated one base-pair to provide a stable hairpin of three nucleotides.

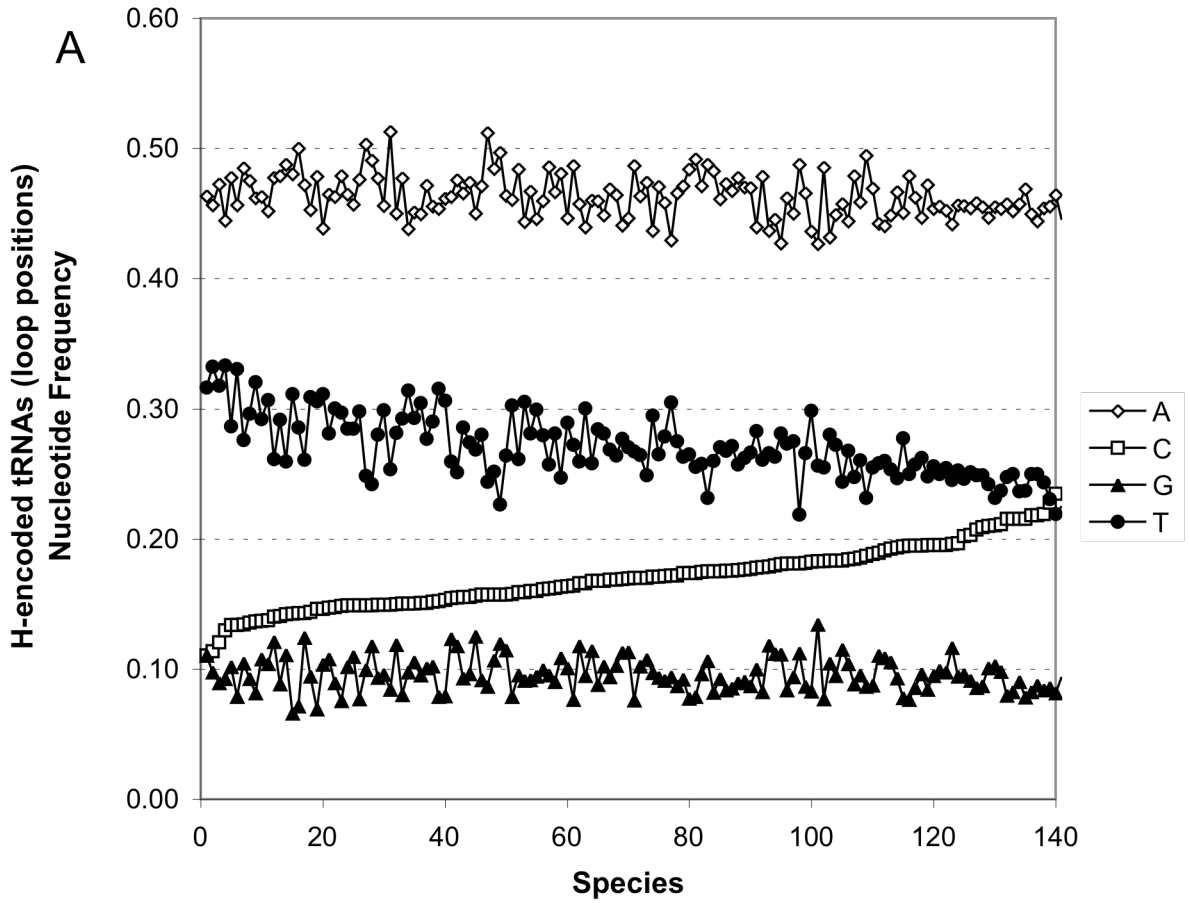


FIG. 3.7. —Observed nucleotide composition among all loop positions for each of the 141 taxa in this study. The species are ordered by increasing percentage of C: (*A*) H-encoded tRNAs; (*B*) L-encoded tRNAs.

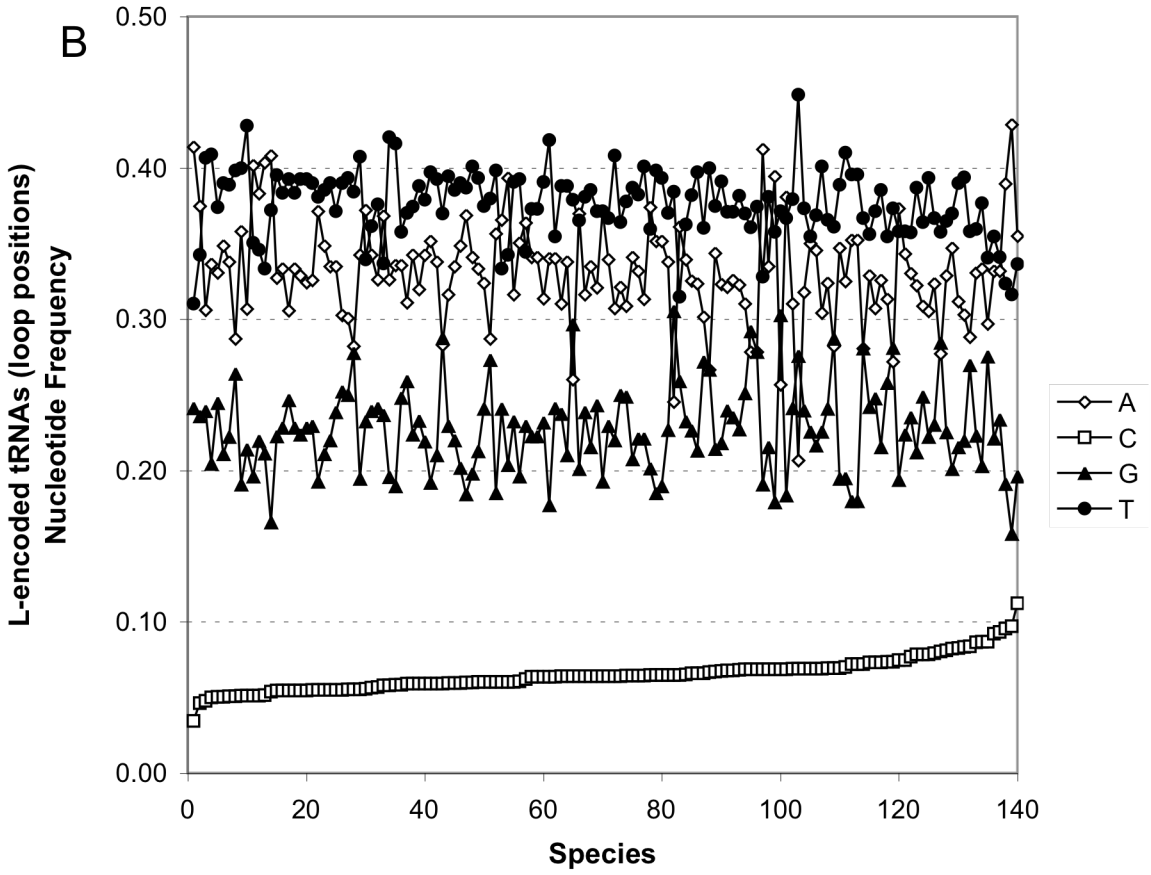


FIG. 3.7. —continued.

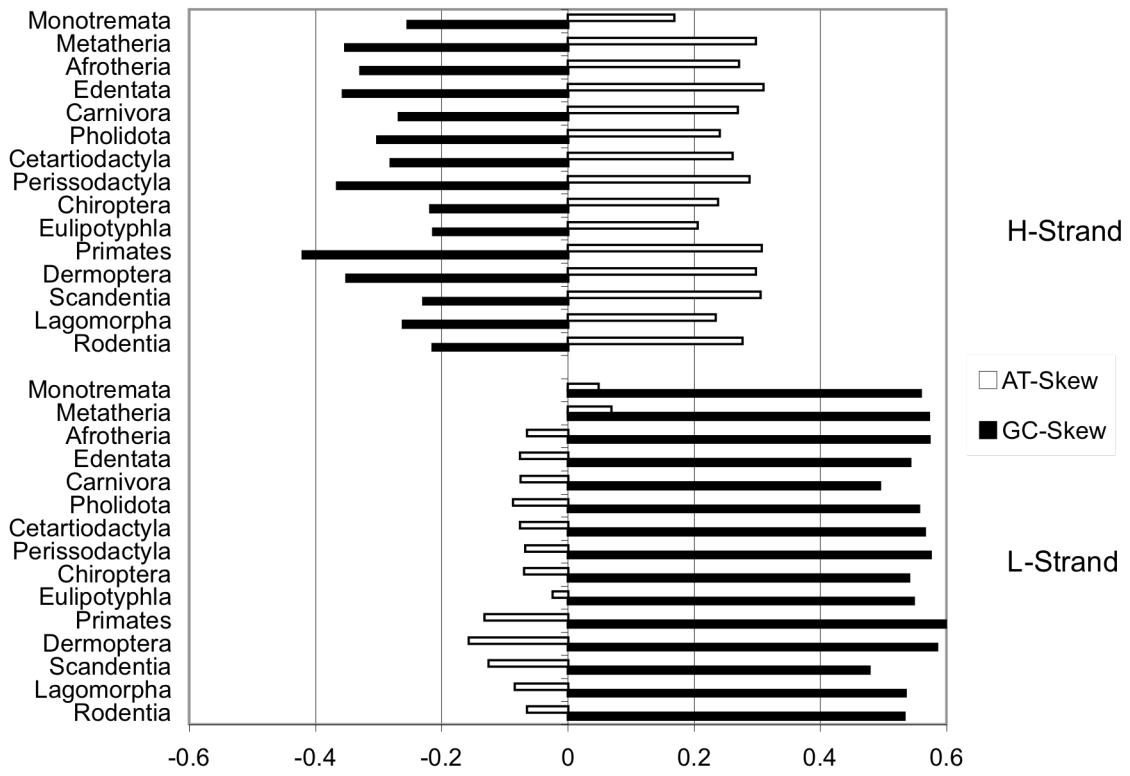


FIG. 3.8. —GC- and AT-skew observed among loop positions for the tRNA genes of each DNA strand for several mammalian orders. Values for H-encoded tRNA genes above, L-encoded tRNA genes below. The superorder Afrotheria includes the following 6 orders: Afrosoricida, Hyracoidea, Macroscelidea, Proboscidea, Sirenia, and Tubulidentata.

Genome Position and Asymmetrical Mutation Bias

To examine the potential effect of single-stranded duration during DNA replication on the tRNA genes, I used the following formulae of Reyes et al. (1998) to calculate the D_{ssH} for the H-encoded tRNAs:

$$D_{ssH} = \frac{L - 2(\bar{x} - O_l)}{L}$$

for the Phe, Val, Leu^{UUR}, Ile, Met, and Trp tRNA genes, and

$$D_{ssH} = \frac{2(O_L - \bar{x})}{L}$$

for the Asp, Lys, Gly, Arg, His, Ser^{AGY}, Leu^{CUN}, and Thr tRNA genes, where L is the total length of the genome, O_L is the position of the L strand origin of replication, and $\bar{x}$ is the middle of the anti-codon loop of the tRNA gene. I determined the following order from lowest to highest D_{ssH}: Asp, Lys, Phe, Val, Gly, Arg, Leu^{UUR}, His, Ile, Ser^{AGY}, Leu^{CUN}, Met, Trp, and Thr. Only three L-encoded tRNA genes (Gln, Ala, and Asn) remain single-stranded during replication of the genome and were not considered here. Figure 3.9 shows the percentage of C and T in H-encoded tRNA stems in order of increasing D_{ssH} for several major mammalian lineages. No obvious pattern is observed in the frequency of either C or T with respect to increasing D_{SSH}. In addition, none of the base-pair types (Watson-Crick, GU/UG, or mismatch) show any relationship to increasing D_{SSH} in H-encoded tRNA genes (data not shown).

As mutations in mt tRNAs are linked to at least 20 human pathologies, I examined whether there were any correlations between the tRNAs known to be

Table 3.3 Mean loop size (in nucleotides) of mammalian mt tRNAs

tRNA	D-loop	V-loop	T-loop
H-strand Encoded			
Phe	5.67 (0.24)	4.00 (0.00)	4.00 (0.29)
	8.53 (0.19)	4.00 (0.00)	4.61 (0.11)
Val	6.44 (0.47)	4.00 (0.00)	4.89 (0.31)
	5.56 (0.07)	4.00 (0.00)	5.13 (0.09)
Leu^{CUN}	6.95 (0.04)	3.08 (0.04)	6.97 (0.03)[c]
	7.04 (0.02)	3.24 (0.04)	7.00 (0.0)
Ile	3.64 (0.08)	5.00 (0.00)	6.87 (0.07)
	4.01 (0.02)	5.00 (0.00)	7.00 (0.03)
Met	**5.18 (0.07)**	4.00 (0.00)	**6.92 (0.04)**
	5.09 (0.03)	4.00 (0.00)	6.85 (0.04)
Trp	5.22 (0.08)	4.00 (0.00)	**5.50 (0.23)***
	6.13 (0.10)	4.00 (0.00)	5.07 (0.08)
Asp	**5.00 (0.17)***	4.00 (0.00)	6.67 (0.29)
	4.37 (0.11)	4.00 (0.00)	6.87 (0.04)
Lys	3.33 (0.24)	**4.89 (0.11)**	5.89 (0.39)
	3.44 (0.07)	4.76 (0.07)	6.90 (0.14)
	7.44 (0.69)[a]		
Gly	5.44 (0.34)	4.00 (0.00)	**6.33 (0.17)**
	5.77 (0.09)	4.00 (0.00)	6.22 (0.06)
Arg	**5.78 (0.28)**	4.00 (0.00)	5.22 (0.28)
	5.59 (0.06)	4.00 (0.00)	5.66 (0.11)
His	**5.08 (0.04)**	4.00 (0.00)	**7.00 (0.14)**
	4.96 (0.03)	4.00 (0.00)	6.90 (0.05)
Ser^{AGY}	3.92 (0.07)[b]	**3.24 (0.07)***	**8.16 (0.06)**
	4.10 (0.08)	3.02 (0.04)	8.11 (0.05)
Leu^{UUR}	**9.89 (0.11)**	5.00 (0.00)	**7.00 (0.00)**
	9.86 (0.03)	5.02 (0.02)[c]	6.99 (0.01)[c]
Thr	**5.67 (0.71)**	4.00 (0.00)	5.78 (1.09)
	5.52 (0.70)	4.00 (0.00)	6.41 (1.92)

Table 3.3 continued

tRNA	D-loop	V-loop	T-loop
L-strand Encoded			
Pro	5.22 (0.22)	4.00 (0.00)	5.11 (0.26)
	5.24 (0.05)	4.01 (0.01)[c]	4.69 (0.09)
Glu	**5.11 (0.11)**	4.00 (0.00)	6.78 (0.22)
	5.00 (0.02)	4.00 (0.00)	6.97 (0.03)
Ser^{UCN}	5.00 (0.00)	4.00 (0.00)	6.89 (0.11)[c]
	5.00 (0.00)	4.02 (0.01)[c]	7.00 (0.00)
Tyr	**3.89 (0.92)***	4.00 (0.00)	5.50 (1.03)
	3.61 (0.78)	4.00 (0.00)	6.29 (1.01)
Cys	**3.86 (0.54)**	4.00 (0.00)	5.53 (1.11)
	3.80 (0.80)	4.00 (0.00)	6.18 (0.77)
Asn	7.44 (1.11)	4.97 (0.17)	6.81 (1.24)
	7.89 (0.45)	5.02 (0.14)	7.18 (0.44)
Ala	5.03 (0.17)	4.00 (0.00)	**7.28 (0.66)***
	5.09 (0.29)	4.00 (0.00)	6.91 (0.59)
Gln	7.51 (0.76)	3.97 (0.16)[c]	7.00 (0.00)
	8.57 (1.00)	4.00 (0.00)	7.00 (0.00)

For each tRNA, the values for rodents are above, non-rodent mammals below. Standard errors are given in parentheses. [a] the D-replacement loop of the metatherian tRNA Lys homolog; [b] the D-replacement loop of Ser^{AGY}; [c] loops where one group (rodents or non-rodents) is fixed, while the other shows size variation. Bold values indicate loops where rodents show greater than non-rodent mammals and asterisks indicate significance at p = 0.05. For the range values of the loops for each tRNA, see Figure 3.1.

associated with disease and a number of molecular features of the tRNAs (Table 3.4). Briefly, I found no correlations between the tRNA (ranked by the number of disease-related mutation identified) and any of the following: strandedness; overall TS:TV ratio; stem TS:TV ratio; loop TS:TV ratio; genome position or D_{SSH} (for H-encoded tRNAs).

Variation Among Closely Related Species

Recently, Reyes et al. (2004) noted a substantial amount of sequence variation in the H-strand protein-coding genes between the two published *Cynocephalus variegatus* mitochondrial genomes (AJ428849 and AF460846: Schmitz et al. 2002). To determine if this variation extended to the tRNA genes, I estimated the sequence divergence between pairs of conspecifics (*Capromys piloroides*, *Cynocephalus variegatus*, *Jaculus jaculus*, *Pongo pygmaeus*, and *Thryonomys swinderianus)* and other closely related species (*Cavia porcellus* and *C. aperea*, and *Bos taurus* and *B. indicus*). The results were mixed with measurable mean divergences from 0.3% (*Capromys* and *Jaculus*) to 6.3% (*Cynocephalus*). The mean difference between the domestic (*Cavia porcellus*) and wild (*C. aperea*) guinea pigs was 2.1%. The cow (*Bos taurus*) and zebu (*B. indicus*) were only 0.8% divergent. The two samples of *Thryonomys* were invariant.

DISCUSSION

Predicted secondary structures of mammalian mt tRNA molecules presented here are in general agreement with those presented by Helm et al. (2000) with a few exceptions. These deviations result from the dissolution of some apical base-pairs of

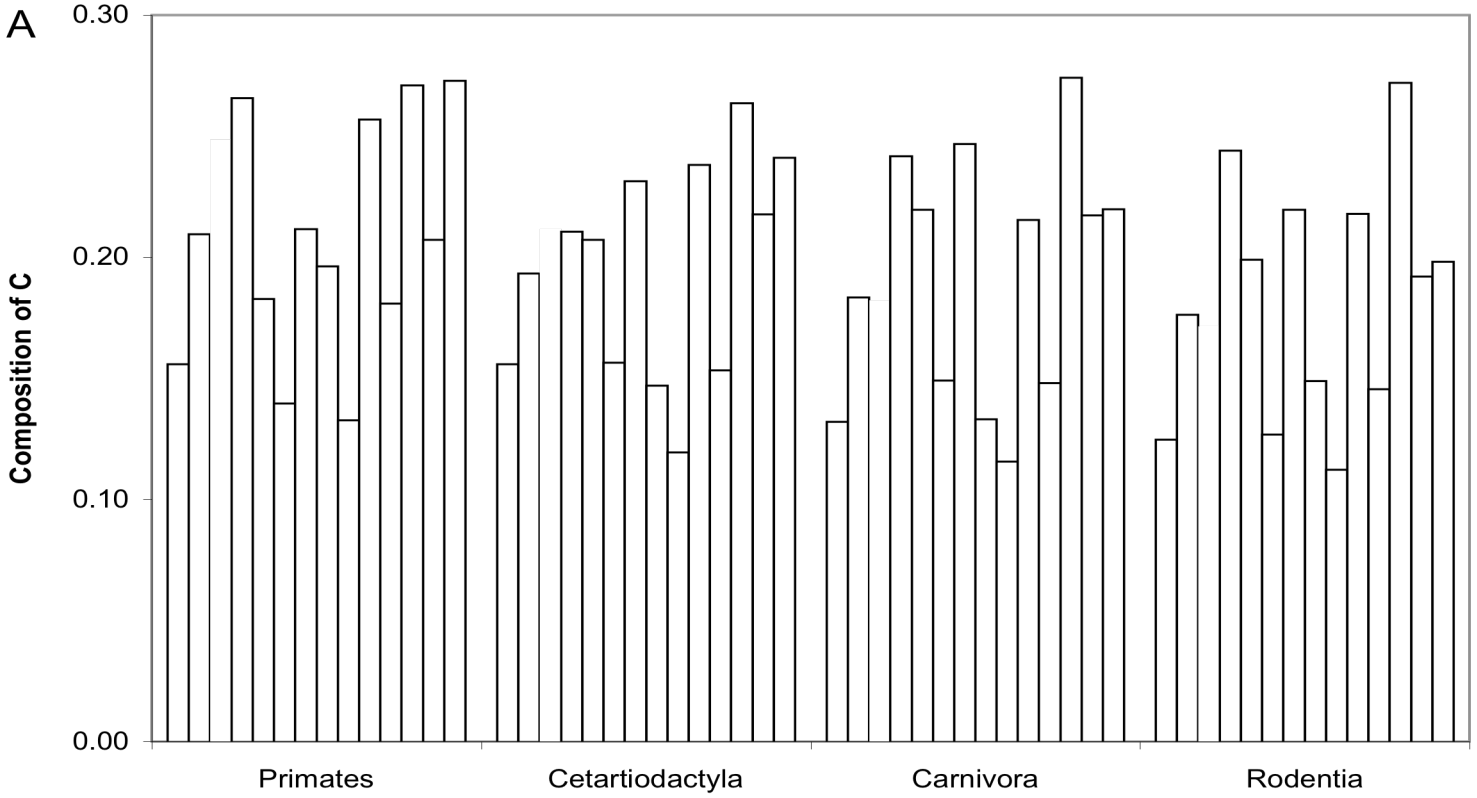


FIG. 3.9. — Observed frequencies of C (*A*) and T (*B*) nucleotides across all stem positions for four major mammalian lineages. The H-encoded tRNA genes are in order of increasing duration of single-strandedness (D_{SSH}) predicted during asymmetric strand replication defined by Reyes et al. (1998); tRNAs, left to right: asparagine, lysine, phenyalanine, valine, glycine, arginine, leucine^{UUR}, histidine, isoleucine, serine^{AGY}, leucine^{CUN}, methionine, tryptophan, and threonine.

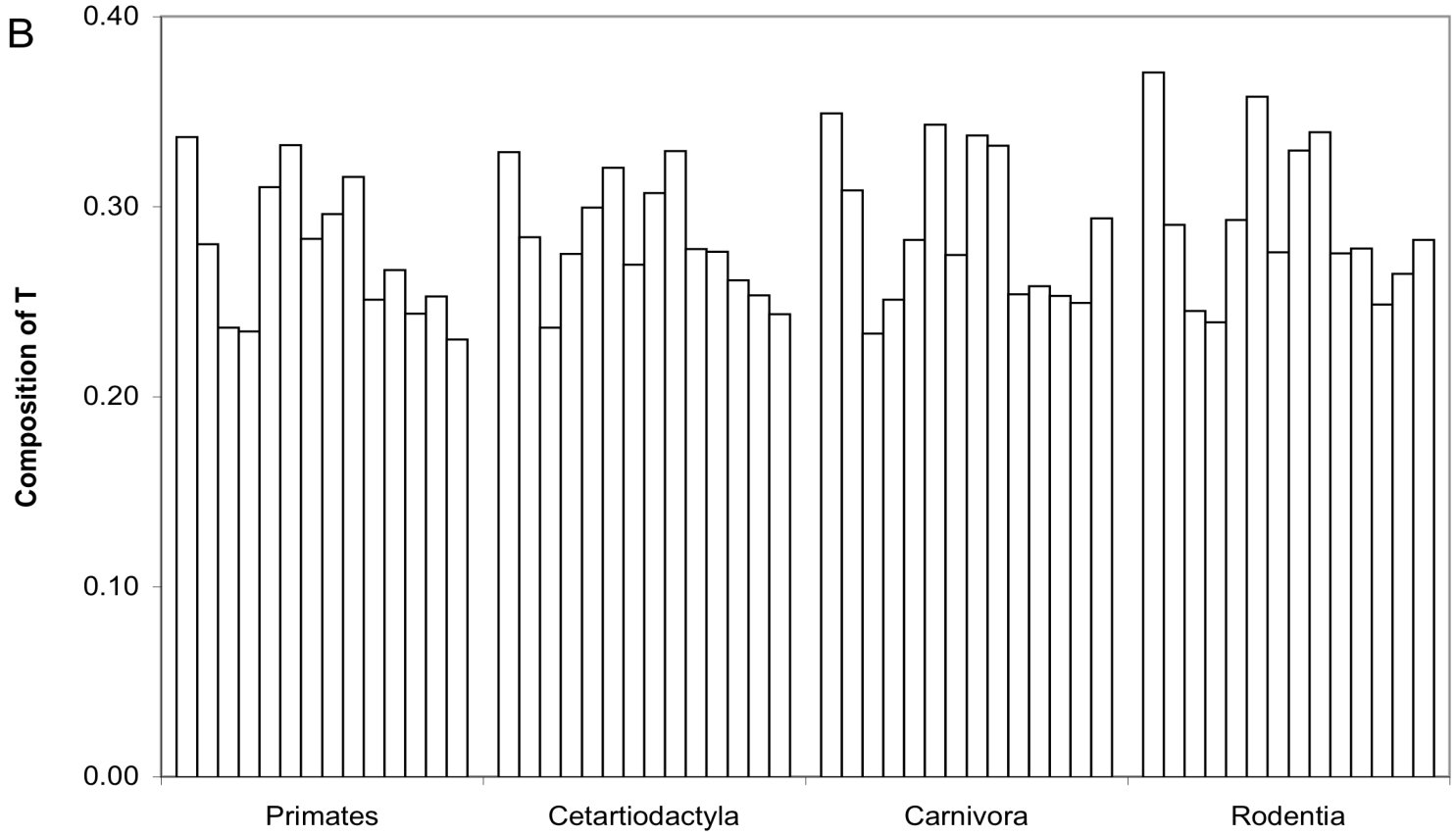


FIG. 3.9. —continued.

Table 3.4 Number of mutations associated with human pathologies identified per tRNA gene

D_{SSH}	tRNA gene	number of mutations
1	Asp	1
2	Lys	13
3	Phe	6
4	Val	4
5	Gly	5
6	Arg	0
7	Leu^{UUR}	24
8	His	3
9	Ile	13
10	Ser^{AGY}	2
11	Leu^{CUN}	6
12	Met	2
13	Trp	5
14	Thr	4
mean		6.3
-	Cys	1
-	Tyr	2
-	Ser^{UCN}	6
-	Glu	1
-	Pro	2
-	Gln	3
-	Ala	1
-	Asn	3
mean		2.4

H-strand encoded tRNA genes are ranked by increasing predicted duration of single-strandedness (D_{SSH}) during genome replication. - refers to L-encoded tRNA genes, ranked in increasing distance from origin of light replication (O_L) along molecule. Numbers of identified disease-related mutations from MITOMAP (2005), available at http://www.mitomap.org.

certain helices due to a reduction in hairpin size. I adhered to the general principle that at least three nucleotides are required to form a stable terminal bulge (Groebe and Uhlenbeck 1988), and used this criterion to reject previously proposed base-pairings. Specifically, the $tRNA^{Cys}$ D-stem is reduced to three base-pairs in both monotremes, the armadillo, the pangolin, and a number of afrotherians. Likewise, the $tRNA^{Lys}$ D-stem is reduced to three base-pairs in a large number of taxa, including all lagomorphs, several cetartiodactyls, three bats, two hegdehogs, the pangolin, three perissodactyls, three primates, several carnivores, four muroid rodents, the colugo, and the tree shrew. Also, there are a number of notable apparent mutations not discussed in Helm *et al.* (2000). Two cetaceans (*Platanista minor* and *Physeter catodon*) show an unusual anti-codon in the $tRNA^{Cys}$ gene with CGA instead of GCA, while another cetacean, *Phocoena phocoena*, has an unusual anti-codon (UCA vs. UCG) in the $tRNA^{Arg}$ gene. Neither of these anti-codon variants, nor a possible means of post-transciptional conversion were presented by the original authors for these aberrant tRNA sequences (Arnason et al. 2000; Arnason, Gullberg, and Janke 2004). A unique bulged insertion (U69) present in the acceptor arm of $tRNA^{Arg}$ is a potential synapomorphy shared by the taxonomically problematic monito del monte (*Dromiciops gliroides*) and all but one of the Australian metatherians (the wallaroo: *Macropus robustus*) (Springer et al. 1998; Nilsson et al. 2003; Asher, Horovitz, and Sanchez-Villagra 2004; Nilsson et al. 2004). Two independent mutations in the $leucine^{UUR}$ tRNA gene suggest either loss or modification of tertiary interaction: a deletion at position U54 or U55 in *Lemur catta* and a shared deletion of G9 or G10 in two carnivores (*Canis familiaris* and *Odobenus rosmarus*).

While 24 separate point mutations in the leucine^{UUR} tRNA are linked to human pathologies, neither of the above mutations are known to be associated with human disease (Schon, Bonilla, and DiMauro 1997; Mitomap 2005).

Mean base composition or skew values for the complete tRNA genes do not appear to be related to either base composition or skew values from either stem or loop regions. In one case, monotremes show the highest %T and lowest AT-skew in all categories of H-stranded tRNAs (see Appendix), and show the highest %A in all categories of L-encoded tRNAs, but the lowest AT-skew in only the L-encoded loop regions. In other examples, marsupials show the lowest overall mean %C, but do not show the lowest %C in either stems or loops of the H-encoded tRNAs. The northern tree shrew (*Tupaia belangeri*: Scandentia) has the highest %C and the lowest %T for all sites and stem regions of the H-encoded tRNAs and the highest %G in the H-encoded stems and loops, but not the highest %G overall (see Appendix).

Nucleotide composition across all sites of H-encoded tRNAs is similar to values found by Nedbal (1995) for all sites of the 12S rRNA gene for a large sample of rodent taxa and nearly identical to that observed by Gibson et al. (2005) for rRNA loop regions, while the nucleotide composition of L-encoded tRNAs is reminiscent of the pattern observed in the rRNA stems (Gibson et al. 2005), but with markedly lower %C. This differs from the expected base frequencies over the entire mitochondrial DNA strands (Anderson et al. 1981; Reyes et al. 1998; Boore 1999). For the H-encoded tRNA genes, direction of skew and magnitude of positive AT-skew observed are consistent with that observed in the H-encoded protein-coding genes of mammalian and other vertebrate

mitochondrial genomes (Reyes et al. 1998; Saccone et al. 1999), but the magnitude of the negative GC-skew is much lower, with no observed values less than -0.15: $\bar{x}$ = -0.091 (fig. 3.4). Skew values calculated from the sense strand of L-encoded tRNAs are expected to be the additive inverse of values observed for the H-strand. These values were consistent with this expectation and the negative AT-skews are of comparable magnitude with those observed in the H-strand tRNAs ($\bar{x}$ = -0.10). In contrast, the positive GC-skew values observed in the L-encoded tRNAs are markedly higher in magnitude ($\bar{x}$ = 0.27) than those of the H-encoded tRNAs, and comparable but inverse to those observed for H-encoded mammalian protein-coding genes (Saccone et al. 1999). Observed TS:TV ratio for all unambiguous sites of either H- or L-encoded tRNAs are substantially higher than those observed for 12S rRNA of rodents (Nedbal, Honeycutt, and Schlitter 1996) or mammals in general (Springer, Hollar, and Burk 1995), 1.37 and 1.77, respectively. Nedbal, Honeycutt, and Schlitter (1996) observed that the low estimated TS:TV ratio for rodent 12S compared to the mammal data set was inconsistent with higher TS:TV ratios for closely-related taxa observed in primates by Hixson and Brown (1986).

Patterns of nucleotide frequency from either H- or L-encoded tRNA stem positions are inconsistent with those of mt rRNA stem regions observed by Gibson et al. (2005) and intermediate to values seen in rodent 12S stem positions by Nedbal (1995). Both the AT- and GC-skew observed in the stem positions of the H-encoded tRNA genes show no obvious pattern with no values significantly different from zero (fig. 3.6). This is in drastic contrast to the pattern observed in L-encoded tRNA stem positions,

where AT-skew ($\bar{x}$ = -0.13) has the same direction and magnitude as that observed for all L-strand tRNA sites, and GC-skew has the same direction, but lower magnitude ($\bar{x}$ = 0.15). In contrast to the %GC over all sites, %GC of the stem portions of H-strand tRNA genes are 5% lower than that for eukaryotic mitochondrial stem positions reported by Higgs (2000) (%GC = 0.45), while the %GC of L-strand tRNA genes are consistent with Higgs (2000). Mean TS:TV ratio of stem positions are higher in both H- and L-strand encoded stem positions (5.86 and 7.82, respectively) than those observed in 12S rRNA stems by Nedbal (1995) or Springer, Hollar, and Burk (1995) (3.7 and 5.13, respectively). Compensatory base changes (CBC) consist of either two transistions or two transversions, therefore the higher biases seen in tRNA stems suggest more stringent compensatory substitution in the mitochondrial tRNA genes compared to the mt rRNA genes.

At nearly all base-pair positions where there was a major shift in frequency between rodents and non-rodents, rodent tRNAs exhibit strict covariation between base-pairs (data not shown). Frequency of GU or UG base-pairs in H-encoded tRNAs of all mammals are consistent with those reported by Helm et al. (2000) for a smaller data set of mammalian mt tRNAs, but lower than those reported for all eukaryotic mt tRNAs by Higgs (2000). Frequency of GU/UG observed in L-strand stems of mammals are also consistent with those reported by Helm et al. (2000), but are significantly higher than those reported by Higgs (2000) for eukaryotic mt tRNAs. The frequencies of GU and UG of L-encoded tRNAs are substantially higher than the overall frequency of GU and UG base-pairs (Table 3.1), suggesting a more pronounced transitional function of GU or

UG in base-pairing of mammalian L-encoded tRNA stem regions. This is in agreement with the spontaneous deamination of C→U associated with single-strandedness of the H-strand during replication (Reyes et al. 1998). Overall mismatch frequencies are consistent between rodents and non-rodents, but different between H-encoded (6%) and L-encoded (3%) tRNAs (Table 3.1). These values reflect a difference in the tolerance of mismatches in tRNAs coded by each strand and are in agreement with those reported by Helm et al. (2000).

Reyes et al. (1998) explored the effect of single-stranded state duration of the H-strand (D_{ssH}) on the asymmetry in base composition for protein-coding genes and found strong evidence to support its effect on nucleotide frequency variation in mammalian mitochondrial genomes. Recently, the effect of the D_{ssH} of protein-coding genes on base composition was confirmed by Gibson et al. (2005) using a larger data set (69 mammalian mitochondrial genomes). Gibson et al. (2005) also showed that the trends of increased %C and decreased %T with longer D_{ssH} varied among taxonomic groups with Primates showing higher %C and lower %T than other taxonomic groups examined. This is likely because the primate mitochondrial genomes have a higher %C and lower %T than the other taxonomic groups examined by Gibson *et al.* (2005). Unlike the protein-coding genes, stem portions of H-encoded tRNAs do not show similar trends in the changes of %C or %T with increasing D_{ssH} or major differences among taxonomic groups. Instead, base composition appears to be primarily related to the tRNA family, with some exceptions such as %C in the rodent methionine (fig. 3.9*A*). Base-pair composition (Watson-Crick, GU/UG, or mismatch) fails to show any correlation with

increasing D_{ssH} in the H-encoded tRNAs, and is instead related only to strandedness (Tables 3.1, 3.2).

Variation in loop regions varies markedly between H- and L-encoded tRNAs. The pattern of variation in H-encoded loops is very consistent with those obtained for loop regions of mt rRNA by Gibson et al. (2005), Nedbal, Honeycutt, and Schlitter (1996), Springer, Hollar, and Burk (1995), and Springer and Douzery (1996) with very high %A, very low %G, and an inverse relationship between change in %C and %T. This bias is similar to that observed in fourfold degenerate sites of the protein-coding genes Variation in L-encoded loop positions is quite different from that observed in the H-encoded tRNAs (above) or rRNAs (Gibson et al. 2005) (fig. 3.7) with very low %C, and higher %G and %T, suggesting very different mutational pressure on L-encoded tRNA genes. The TS:TV ratios of unambiguous loop positions are comparable to those observed by Nedbal (1995), Springer, Hollar, and Burk (1995), or Springer and Douzery (1996) for the 12S rRNA (1.1, 1.16, and 1.2, respectively).

Lynch (1996, 1997) proposed that mammalian mt tRNA genes are accumulating an overwhelming number of mutations, resulting in a mutational meltdown known as Muller's ratchet. As the result of this meltdown, Lynch concluded that mammals should show an overall reduction in loop size among all animal mitochondrial tRNAs. The evidence from my increased taxon sampling suggests that mammals show a wide range of loop lengths, ranging from very small (3 nt) to large (12 nt). In addition, rodents, a group shown to exhibit high mutation rates in their mtDNA (*e.g.*, Nedbal *et al.* 1996;

Rowe and Honeycutt 2002), exhibit as broad a range of loop sizes as all of Mammalia (Table 3.2).

While I found no correlations between disease-related tRNA mutations and a number of molecular features, such as strandedness, TS:TV ratio, or genome position (as D_{SSH}), most L-encoded tRNAs show few of these mutations ($\bar{x}$ = 2.4), while some H-encoded tRNAs show very high numbers of these mutations (13 for tRNAs$^{Lys, Ile}$ and 24 for tRNA^{LeuUUR}) and a much higher mean (6.3) (Table 3.4). The three H-encoded tRNA genes with very high numbers of pathologic mutations are subjected to the effects of single-strandedness during replication of the genome, but for different durations (Table 3.4).

Among closely related taxa, the accumulation of transversions and all substitution classes in stem positions is proportional to sequence divergence. Divergences of less than 0.5% showed only loop transitions, *Bos* (at 0.8%) also showed C→T transitions in H-encoded stems, and *Cavia* (at 2.1%) showed additional A→G transitions in L-encoded stems. Only the most divergent comparisons (*Cynocephalus* and *Pongo*: 6.3 and 4.0%, respectively) showed evidence of CBCs. The two subspecies of orangutan showed single CBC events in the A-stem of tRNA^{Lys} and the T-stem of tRNA^{Thr}. Xu and Arnason (1996) estimated that *Pongo pygmaeus pygmaeus* and *P. p. abelii* have been separated for ~10 my based on mitochondrial protein-coding gene sequences. For *Cynocephalus*, my results from the tRNA genes support those of Reyes et al. (2004) for these two mitochondrial genomes. The two highly-divergent colugos show evidence of four separate compensatory substitution events: one in the T-stem of

$tRNA^{His}$, one in the A-stem of $tRNA^{Leu(CUN)}$, and two in the T-stem of $tRNA^{Thr}$. The Bayesian consensus tree of Reyes et al. (2004; their fig. 3.1) shows deep divergence between these two colugos (~ 8 - 10 mybp).

CONCLUSIONS

Based on fossil evidence, the split among the major extant mammal groups (Monotremata, Metatheria, and Eutheria) dates to at least 120 mybp (Novacek 2001). The data presented here contain representatives of most major lineages of extant mammals, at least to the familial level, thus providing a broad overview of patterns of nucleotide substitutions and the structure of the 22 tRNAs encoded within the mitochondrial genome. These data have been used to evaluate the extent to which tRNA gene position on the H- and L-strands influences overall base composition and substitution patterns seen in both stem and loop regions. Although loop size, a potential indicator of increased susceptibility to mutations, did not differ between H- and L-strands, tRNAs encoded on these two strands did show different patterns of base composition, as reflected by the patterns of AT- and GC-skew in stems and loops. In addition to comparisons of H- and L-strand encoded genes, comparisons of base-pairing in stems and variation in loop size were made across all major groups of mammals, with special emphasis on differences between rodents and non-rodents, as the protein-encoding and rRNA mitochondrial genes of rodents evolves considerably faster than in non-rodents. Although rodents and non-rodents demonstrated different patterns of substitutions, both groups have a relatively high frequency of invariant base-pairs in

stems and do not demonstrate an overall reduction in size of loops across divergent lineages. The high level of compensatory substitutions and changes that maintain pairing versus low levels of mispairing in stem regions of these two groups indicate that the structural integrity of most tRNAs throughout the mammalian radiations is maintained. The only exception to this rule involves the aberrant $tRNA^{Lys}$ of marsupials. Most phylogenetic comparisons of mammals based on mitochondrial sequences have focused on either rRNA genes or protein-encoded genes to the exclusion of tRNAs. Given the overall conservative nature of these genes, especially with respect to a decrease in indels relative to rRNAs, I feel that tRNAs provide a conservative set of markers that should be included in phylogenetic analyses. Nevertheless, the inclusion of such markers should take into consideration the overall patterns of substitution classes and base composition biases revealed through structural comparison, particularly with respect to strandedness, as demonstrated in this study.

CHAPTER IV

SUMMARY

In Chapter I, I described the historic classification schemes put forth for Rodentia. Many major rodent lineages arose rapidly in the Eocene (65–54 MYA) and of the 31+ extant families were well established by the Oligocene (54–38 MYA). Rapid radiations, followed by extensive evolutionary change within lineages, can make the interpretation of relationships among crown taxa difficult. Collectively, the rodents show an enormous number of parallelisms and convergences in a variety of morphological characters, including enamel microstructure (von Koenigswald 1985), cranial anatomy (Vianey-Liaud 1985), and numerous adaptations to a fossorial habit (Eisenberg 1981) that add to the complication. Two major classification schemes have persisted for over a century. Brandt's (1855) classification divides rodents into 3 suborders based on the zygomasseteric apparatus (an anatomical complex between the masseter muscles and zygomatic arch and infraorbital foramen of the skull adapted for gnawing): Hystricomorpha, Myomorpha, and Sciuromorpha. Tullberg's (1899) scheme separates rodents into 2 suborders based on the plane of the angular process of the mandible: Hystricognathi and Sciurognathi. To emphasize the difficulties in determining higher-level relationships among rodents, I presented a collection of published variants on these classifications. Many of the difficulties in higher-level rodent classifications have involved the placement of a number of enigmatic rodent families and superfamilies: Anomaluridae (Central African scaly-tailed squirrels),

Aplodontidae (mountain beavers), Castoridae (beavers), Ctenodactylidae (gundis), Erethizontidae (New World porcupines), Geomyoidea (gophers, kangaroo rats, and pocket mice), Gliridae (dormice), Hystricidae (Old World porcupines), and Pedetidae (springhares). To add to the controversy, a number of molecular studies have questioned the monophyly of Rodentia over the past decade (Graur et al. 1991; D'Erchia et al. 1996; Reyes et al. 2000). With rodents comprising nearly half of all extant eutherian mammals species, establishing well-supported and accepted phylogenetic classification is essential in our understanding of the Tree of Life.

Many of the studies that refute rodent monphyly have analyzed complete mitochondrial genome sequences to make their claims (D'Erchia et al. 1996; Reyes et al. 2000; Mouchaty et al. 2001; Reyes et al. 2004), but the limited number of available rodent mtDNA genomes continues to plague these studies. In Chapter II, I presented the complete mitochondrial genome sequences for four enigmatic rodent taxa: *Aplodontia rufa* (mountain beaver: Aplodontidae), *Cratogeomys castanops* (yellow-faced pocket gopher: Geomyoidea, Geomyidae), *Erethizon dorsatum* (North American porcupine: Erethizontidae), and *Hystrix africaeaustralis* (African porcupine: Hystricidae). Structurally, the new genomes are unremarkable, with the typical vertebrate gene complement and mammalian gene order, including the positions of all 22 tRNA genes. A multiple sequence alignment of the protein-coding genes (13 available rodents, 3 lagomorphs (rabbits, hares, and pikas), 6 primates, 1 colugo (or flying lemur), 1 tree shrew, and a perissodactyl outgroup) was analyzed under 3 optimality criteria: maximum-parsimony (MP), maximum-likelihood (ML), and Bayesian inference (BI: a

Metropolis-coupled Markov chain Monte Carlo sampling). Different analyses under ML and BI converged on a single topology with weak support for a monophyletic Rodentia that was significantly better than those from MP.

A concatenated alignment of the 22 tRNA genes for the same taxon set was analyzed under MP and BI (DNA and RNA models). Both MP and BI (DNA model) analyses failed to recover a monophyletic Rodentia and both topologies were significantly worse than the ML tree for the protein-coding genes. Bayesian analysis of the tRNA dataset under the parameter-rich compensatory RNA-based models was problematic, failing to converge on a stationary likelihood distribution after 2×10^7 generations. Parallel (cluster-based) computing may be necessary to fully analyze this dataset.

With a single exception (the position of the Gliridae), the relationships recovered in the ML tree are supported by other data. The sister relationship between Aplodontidae and Sciuridae is supported by numerous morphological and molecular data (Wood 1955; Lavocat and Parent 1985; Sarich 1985; Vianey-Liand 1985; Nedbal, Honeycutt, and Schlitter 1996; Adkins et al. 2001; DeBry and Sagel 2001; Huchon et al. 2002; Montgelard et al. 2002). The affinity of Geomyoidea to the myomorph rodents has been proposed for decades (Hill 1937; Wilson 1949; Wood 1955; Wood 1959; Wahlert 1978) and the two lineages are linked by numerous fetal membrane characters (Luckett 1985) and the carotid arterial patterns (Bugge 1985). The Hystricognathi were well supported as monophyletic with Hystricidae sister to the Caviomorpha (New World hystricognaths) separate from the other Old World hystricognaths (Phiomorpha). This

relationship was proposed by Wood (1965) and is supported by some molecular evidence (Rowe 2002; but see Adkins et al. 2001; Huchon et al. 2002). Due to insufficient taxon sampling, the placement of Erethizontidae within the Caviomorpha can not be established with the present dataset. In the ML tree, Gliridae is sister to the Hysticognathi. While this may seem dubious, this may simply be due to a lack of important taxa (Anomaluridae, Ctenodactylidae, and Pedetidae) in the present dataset.

As the number of complete mitochondrial (mt) genome sequences has grown, few authors have examined mt tRNAs, instead focusing on protein-coding and/or rRNA genes. In Chapter III, I present revised consensus secondary structure models for all 22 mitochondrial (mt) tRNA genes based on comparisons of 109 mammalian mt genomes. Additionally, 11 tRNAs were sequenced for another 31 rodent taxa, and these data were used in combination to examine the molecular evolution of tRNAs based on comparisons of rodent and non-rodent taxa as well as a detailed examination of variation among genes encoded by light (L) and heavy (H) strands. Unlike the protein-coding genes, the proportion of C or T is not correlated with the position of the tRNA gene along the genome, but rather is attributed to either the specific tRNA species or coding strand. Overall, both H- and L-encoded tRNAs are AT-rich with markedly different %G and GC-skew. There is a striking difference in observed skew between the stem regions of H- and L-encoded tRNAs. The proportion of Watson-Crick base-pairs is also higher in H-encoded tRNAs, with a higher proportion of GU/UG pairs in L-encoded tRNAs. This suggests increased level of mismatch compensation in L-strand tRNAs. Among rodents, the number of variable stem base-pairs was nearly 75% of that observed across

all other orders combined. Among closely related taxa, compensatory base changes (CBC) were present only at divergences of 4% or greater. Nucleotide frequencies in loop regions are also markedly different between the two strands with %A of H-encoded tRNAs of ~ 50%. For 10 tRNAs, rodents match or exceed the range in loop size observed across all other mammals. Neither a reduction of loop size nor an accumulation of deleterious mutations, both presumably suggestive of a mutational meltdown (Muller's ratchet), was observed. Identified mutations associated with human pathologies appear to be correlated only with the coding strand, with H-strand tRNAs being linked to substantially more of these mutations.

LITERATURE CITED

Adkins, R. M., E. L. Gelke, D. Rowe, and R. L. Honeycutt. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. Molecular Biology and Evolution **18**:777–791.

Adkins, R. M., and R. L. Honeycutt. 1991. Molecular phylogeny of the superorder Archonta. Proc. Natl. Acad. Sci. USA **88**:10317-10321.

Adkins, R. M., R. L. Honeycutt, and T. R. Disotell. 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.

Allard, M. W., R. L. Honeycutt, and M. J. Novacek, 1999. Advances in higher level mammalian relationships. Cladistics **15**:213–219.

Allard, M. W., M. M. Miyamoto, and R. L. Honeycutt, 1991. Tests for rodent polyphyly. Nature **353**:610–611.

Alvarez, W., E. G. Kauffman, F. Surlyk, L. W. Alvarez, F. Asaro, and H. V. Michel. 1984. Impact theory of mass extinctions and the invertebrate fossil record. Science **223**:1135–1141.

Amrine-Madsen, H., K. P. Koepfli, R. K. Wayne, and M. S. Springer. 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. Mol. Phylogenet. Evol. **28**:225–240.

Anderson, S, A. T. Bankier, B. G. Barrell, M. H. de Bruijn, A. R. Coulson et al. (9 co-authors). 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

Arnason, U., J. A. Adegoke, K. Bodin, E. W. Born, Y. B. Esa, A. Gullberg, M. Nilsson, R. V. Short, X. Xu, and A. Janke. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA **99**:8151–8156.

Arnason, U., A. Gullberg, S. Gretarsdottir, B. Ursing, and A. Janke. 2000. The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. J. Mol. Evol. **50**:569–578.

Arnason, U., A. Gullberg, and A. Janke. 1997. Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and Ferungulates. Mol. Biol. Evol. **14**:762–768.

Arnason, U., A. Gullberg, and A. Janke. 1999. The mitochondrial DNA molecule of the aardvark, *Orycteropus afer*, and the position of the Tubulidentata in the eutherian tree. Proc. R. Soc. Lond. B. Bio Sci. **266**:339–345.

Arnason, U., A. Gullberg, and A. Janke. 2004. Mitogenomic analyses provide new insights into cetacean origin and evolution. Gene **333:**27–34.

Arnason, U. and A. Janke. 2002. Mitogenomic analyses of eutherian relationships. Cytogenet. Genome Res. **96**:20–32.

Asher, R. J., I. Horovitz, and M. R. Sanchez-Villagra. 2004. First combined cladistic analysis of marsupial mammal interrelationships. Mol. Phylogenet. Evol. **33**:240–250.

Baker, R.H., and R. DeSalle. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. Syst. Biol. **46**:654–673.

Baker, R. H., X. B. Yu, and R. DeSalle. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. **9**:427–436.

Bielawski, J. P. and J. R. Gold. 2002. Mutation patterns of mitochondrial H- and L-strand DNA in closely related Cyprinid fishes. Genetics **161**:1589–1597.

Bientema, J. J., K. Rodewald, G. Braunitzer, J. Czelusniak, and M. Goodman. 1991. Studies on the phylogenetic position of the Ctenodactylidae (Rodentia). Mol. Biol. Evol. **8**:151–154.

de Blainville, H. M. D. 1816. Prodrome d'une nouvelle distribution systématique du règne animal. Bull. Soc. Philomath. Paris Sér. 3, **3**:105–124.

Boore J. L. 1999. Animal mitochondrial genomes. Nuc. Acids Res. **27**:1767–1780.

Brandt, J. F. 1855. Beitrage zür nahern Kenntniss der Saugerthiere Russlands. Mem. Acad. Imp. Sci. Petersbourg **6-9**:1–375.

Bruno, W. J., and A. L. Halpern. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. **16**:564–566.

Bugge, J. J. 1971. The cephalic arterial system in mole-rats (Spalacidae), bamboo rats (Rhizomyidae), jumping mice and jerboas (Dipodoidea), and dormice (Gliroidea) with special reference to the systematic classification of rodents. Acta Anat. **79**:165–180.

Bugge, J. J. 1974. The cephalic arterial system in insectivores, primates, rodents, and lagomorphs, with special reference to the systematic classification. Acta Anat. **87**(Supp 62):1–160.

Bugge J. J. 1985. Systematic value of the carotid arterial pattern in rodents. Pp. 355–379 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92. Plenum, New York.

Butler, P. M. 1988. Phylogeny of the insectivores. Pp. 117–141 *in* M. J. Benton, ed. The Phylogeny and Classification of the Tetrapods, Volume 2: Mammals, Clarendon Press, Oxford, UK.

Cannone, J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. DeSouza et al. (7 co-authors). 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics **3**:2.

Cao, Y., J. Adachi, and M. Hasegawa. 1994. Eutherian phylogeny as inferred from mitochondrial DNA sequence data. Jap. J. Gen. **69**:455–472.

Cao, Y., J. Adachi, T. Yano, and M. Hasegawa. 1994. Phylogenetic place of guinea pigs: no support of the rodent-polyphyly hypothesis from maximum-likelihood analyses of multiple protein sequences. Mol. Biol. Evol. **11**:593–604.

Cao, Y., M. Fujiwara, M. Nikaido, N. Okada, and M..Hasegawa. 2000. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. Gene **259**:149–158.

Cao, Y., A. Janke, P. J. Waddell, M. Westerman, O. Takenaka, S. Murata, N. Okada, S. Pääbo, and M. Hasegawa. 1998. Conflict among individual mitochondrial

proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. **47**:307–322.

Cao, Y., N. Okada, and M. Hasegawa. 1997. Phylogenetic position of guinea pigs revisited. Mol. Biol. Evol. **14**:461–464.

Cheng, S., S.-Y. Chang, P. Gravitt, and R. Repress. 1994. Long PCR. Nature **369**:684–685.

Clayton, D. A. 1982. Replication of animal mitochondrial DNA. Cell **28**:693–705.

Cuvier, G. 1817. La Règne Animal Distribué d'Après Son Organization, Pour Servir de Base à l'Histoire Naturelle des Animaux et d'Introduction à l'Anatomie Comparée, Vol. 1, Carpenter and Westwood, London, UK (1859 Reprint).

Czelusniak, J., M. Goodman, B. F. Koop, D. A. Tagle, J. Shoshani, G. Braunitzer, T. Kleinschmidt, W. W. de Jong, and G. Matsuda. 1990. Perspectives from amino acid and nucleotide sequences on cladistic relationships among higher taxa of Eutheria. Pp. 545–572 *in* H. H. Genoways, ed. Current Mammalogy, Vol. 2, Plenum, New York.

D'Erchia, A. M., C. Gissi, G. Pesole, C. Saccone, and U. Arnason. 1996. The guinea-pig is not a rodent. Nature **381**:597–600.

de Bruijn, M. H. L., and A. Klug. 1983. A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire 'dihydrouridine' loop and stem. EMBO J. **2**:1309–1321.

de Jong, W. W., M. A. M. van Dijk, C. Poux, G. Kappé, T. van Rheede, and O. Madsen. 2003. Indels in protein-coding sequences of Euarchontoglires constrain the rooting of the eutherian tree. Mol. Phylogenet. Evol. **28**:328–340.

DeBry, R. W., and R. M. Sagel. 2001. Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. Mol. Phylogenet. Evol. **19**:290–301.

Dörner, M., M. Altmann, S. Pääbo, and M. Mörl. 2001. Evidence for import of a lysyl-tRNA into marsupial mitochondria. Mol. Biol. Cell **12**:2688–2698.

Douzery, E., and F. M. Catzeflis. 1995. Molecular evolution of the mitochondrial 12S rRNA in Ungulata (Mammalia). J. Mol. Evol. **41**:622–636.

Eisenberg, J. F. 1981. The Mammalian Radiations: an Analysis of Trends in Evolution, Adaptation, and Behavior. Univ. of Chicago Press, Chicago, IL.

Eizirik, E., W. J. Murphy, and S. J. O'Brien. 2001. Molecular dating and biogeography of the early placental mammal radiation. J. Hered. **92**:212–219.

Engel, S. R., K. M. Hogan, J. F. Taylor, and S. K. Davis. 1998. Molecular systematics and paleobiogeography of the South American sigmodontine rodents. Mol. Biol. Evol. **15**:35–49.

Erixon, E., B. Svennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. **52**:665–673.

Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998. Basecalling of automated sequencer traces using phred. I. accuracy assessment. Genet. Res. **8**:175–185.

Faith, J. J., and D. D. Pollock. 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. Genetics **165**:735–745.

Fahlbusch, V. 1985. Origins and evolutionary relationships among geomyoids. Pp. 617–630 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92. Plenum, New York.

Fischer, T. V., and H. W. Mossman. 1969. The fetal membranes of *Pedetes capensis* and their taxonomic significance. Amer. J. Anat. **124**:89–116.

Florentz, C., and M. Sissler. 2001. Disease-related versus polymorphic mutations in human mitochondrial tRNAs: where is the difference? EMBO Rep. **2**:481–486.

Flores-Villela, O. K. M. Kjer, M. Benabib, and J. W. Sites, Jr. 2000. Multiple data sets, congruence, and hypothesis testing for the phylogeny of basal groups of the lizard genus *Sceloporus* (Squamata, Phrynosomatidae). Syst. Biol. **49**:713–739.

Flynn, L. J., L. L. Jacobs, and E. H. Lindsay. 1985 Problems in muroid phylogeny: relationship to other rodents and origin of major groups. Pp. 589–616 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92. Plenum, New York.

Frabotta, L. J. 2002. Phylogeography of the sagebrush lizard, *Sceloporus graciosus* (Phrynosomatidae) in California: an analysis of a mitochondrial DNA data set. Unpublished Master's thesis. California State University, Long Beach, CA.

Frye, M. S., and S. B. Hedges. 1995. Monophyly of the order Rodentia inferred from mitochondrial DNA sequences of the 12S rRNA, 16S rRNA, and tRNA-valine. Mol. Biol. Evol. **12**:168–176.

Gaut, B. S., and P. O. Lewis. 1995. Success of maximum likelihood in the four-taxon case. Mol. Biol. Evol. **12**:152–162.

George, W., 1985. Reproductive and chromosomal characters of ctenodactylids as a key to their evolutionary relationships. Pp. 453–474 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92. Plenum, New York.

Gibson, A., V. Gowri-Shankar, P. G. Higgs, and M. Rattray, M. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol. Biol. Evol. **22**:251–264.

Gidley, J. W. 1912. The lagomorphs an independent order. Science **36**:285–286.

Gillespie, J. J. 2004. Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. Mol. Phylogenet. Evol. **33**:936–943.

Gillespie, J. J. , J. J. Cannone, R. R. Gutell, and A. I. Cognato. 2004. A secondary structural model of the 28S rRNA expansion segments D2 and D3 from rootworms and related leaf beetles (Coleoptera: Chrysomelidae: Galerucinae). Insect Mol. Biol. 13: 495–518.

Gissi, C., A. Reyes, G. Pesole, and C. Saccone. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Mol. Biol. Evol. **17**:1022–1031.

Goldman, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. **37**:50–61.

Goto, Y.-I., I. Nonaka, and S. Horai, S. 1990. A mutation in the $tRNA^{Leu(UUR)}$ gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. Nature **348**:651–653.

Graur, D., W. A. Hide, and W. H. Li. 1991. Is the guinea-pig a rodent? Nature **351**:649–652.

Graur, D., W. A. Hide, A. Zharkikh, and W. H. Li. 1992. The biochemical phylogeny of guinea-pigs and gundis, and the paraphyly of the order Rodentia. Comp. Biochem. Phys. B **101**:495–498.

Gregory, W. K. 1910. The orders of mammals. Bull. Am. Mus. Nat. Hist. **27**:1–525.

Groebe, D. R., and O. C. Uhlenbeck. 1988. Characterization of RNA hairpin loop stability. Nuc. Acids Res. **16**:11725–11735.

Gutell, R. R. 1996. Comparative sequence analysis and the structure of 16S and 23S rRNA. Pp. 111–128 *in* A. E. Dahlberg and R. A. Zimmerman, eds. Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Synthesis. CRC Press, Boca Raton, FL.

Hancock, J. M., and A. P. Vogler. 2000. How slippage-derived sequences are incorporated into rRNA variable-region secondary structure: implications for phylogeny reconstruction. Mol. Phylogenet. Evol. **14**:366–374.

Hartenberger, J.-L. 1985. The order Rodentia: major questions on their evolutionary origin, relationships, and suprafamilial systematics. Pp. 1-34 *in* W. P. Luckett

and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

Hartenberger, J.-L., 1998. Description de la radiation des Rodentia (Mammalia) du Paléocène supérieur au Miocène; incidences phylogénétiques. CR Acad. Sci. II **326**:439–444.

Hasegawa, M., Y. Cao, J. Adachi, and T. Yano. 1992. Rodent polyphyly? Nature **355**:595.

Helm, M., H. Brulé, D. Friede, R. Giegé, D. Pütz, and C. Florentz. 2000. Search for characteristic structural features of mammalian mitochondrial tRNAs. RNA **6**:1356–1379.

Higgs, P. G. 2000. RNA secondary structure: physical and computational aspects. Quart. Rev. of Biophysics **33**:199–253.

Hill, J. E. 1937. Morphology of the pocket gopher mammalian genus *Thomomys*. U. Ca. Pub. Zool. **42**:81–172.

Hixson, J. E., and W. M. Brown. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: structure, evolution, and phylogenetic implications. Mol. Biol. Evol. **3**: 1–18.

Hogan, K. M., S. K. Davis, and I. F. Greenbaum. 1997. Mitochondrial DNA analysis of the systematic relationships within the *Peromyscus maniculatus* species group. J. Mamm. **78**:733–743.

Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir. 1965. Structure of a ribonucleic acid. Science

147:1462–1465.

Honeycutt, R. L., M. A. Nedbal, R. M. Adkins, and L. L. Janecek. 1995. Mammalian mitochondrial DNA evolution: a comparison of the cytochrome *b* and cytochrome *c* oxidase II genes. J. Mol. Evol. **40**:260–272.

Huchon, D., F. M. Catzeflis, and J. P. E. Douzery. 1999. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. Mol. Biol. Evol. **16**:577–589.

Huchon, D., F. M. Catzeflis, and J. P. E. Douzery. 2000. Variance of molecular datings, evolution of rodents and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. Proc. R. Soc. Lond. B Biol. Sci. **267**:393–402.

Huchon, D. F., and J. P. E. Douzery. 2001. From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. Mol. Phylogenet. Evol. **20**:238–251.

Huchon, D. F., O. Madsen, M. J. J. B. Sibbalb, K. Ament, M. J. Stanhope, F. Catzeflis, W. W. de Jong, and E. J. P. Douzery. 2002. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. Mol. Biol. Evol. **19**:1053–1065.

Hudelot, C. V., V. Gowri-Shankar, H. Jow, M. Rattray, and P. G. Higgs. 2003. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. Mol. Phylogenet. Evol. **28**:241–252.

Huelsenbeck, J. P. 1998. Performance of phylogenetic methods in simulation. Syst. Biol. **44**:17–48.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics **17**:754–755.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294**:2310–2314.

Inoue, J. G., M. Miya, K. Tsukamoto, and M. Nishida. 2001. Complete mitochondrial DNA sequence of *Conger myriaster* (Teleostei: Anguilliformes): novel gene order for vertebrate mitochondrial genomes and the phylogenetic implications for anguilliform families. J. Mol. Biol. **52**:311–320.

Irwin, D. M., T. D. Kocher, and A. C. Wilson. 1991. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. **32**:128–144.

Jameson, D., A. P. Gibson, C. Hudelot, and P. G. Higgs. 2003. OGRe: a relational database for comparative analysis of mitochondrial genomes. Nuc. Acids Res. **31**:202–206. (http://drake.physics.mcmaster.ca/ogre/index.shtml).

Janke, A., D. Erpenbeck, M. Nilsson, and U. Arnason. 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. Proc. R. Soc. Lond. B Biol. Sci. **268**:623–631.

Janke, A., G. Feldmaier-Fuchs, W. K. Thomas, A. von Haeseler, and S. Pääbo. 1994. The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137**:243–256.

Janke, A., X. Xu, and U. Arnason. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among

Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA **94**:1276–1281.

Jow H., C. Hudelot, M. Rattay, and P. G. Higgs. 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol. Biol. Evol. **19**:1591–1601.

Jukes, T. H. 1995. A comparison of mitochondrial tRNAs in five vertebrates. J. Mol. Evol. **40**:537–540.

Kim, S. H., G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, and A. Rich. 1973. Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. Science **179**:285–288.

Kim, S. H., F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman, and A. Rich. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science **185**:435–440.

Kishino, H., and M. Hasegawa. 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.

Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. **4**:314–330.

Kjer, K. M. 1997. An alignment template for amphibian 12S rRNA, domain III: conserved primary and secondary structural motifs. J. Herpetology **31**:599–604.

Korth, W. W. 1984. Earliest Tertiary evolution and radiation of rodents in North America. Bull. Carnegie Mus. Nat. Hist. **49**:307–322.

Korth, W. W. 1994. The Tertiary Record of Rodents in North America. Topics in Geobiology: Vol 12, Plenum, New York.

Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature **392**:917–920.

Kumazawa, Y., H. Himeno, K. Miura, and K. Watanabe. 1991. Unilateral aminoacylation specificity between bovine mitochondria and eubacteria. J. Biochem. **109**:421–427.

Kumazawa, Y., and M. Nishida. 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. **37**:380–398.

Kumazawa, Y., and M. Nishida. 1995. Variations in mitochondrial tRNA gene organization and deep-branch animal phylogenetics. Mol. Biol. Evol. **12**:759–772.

Kumazawa, Y., and M. Nishida. 1999. Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. Mol. Biol. Evol. **16**:784–792.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Mol. Biol. Evol. **13**:1242–1254.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics **150**:313–329.

Kumazawa, Y., T. Yokogawa, E. Hasegawa, K. Miura, and K. Watanabe. 1989. The aminoacylation of structurally variant phenylalanine tRNAs from mitochondria and various nonmitochondrial sources by bovine phenyalanyl-tRNA synthetase. J. Biol. Chem. **264**:13005–13011.

Landry, S. O., 1957. The relationships of New and Old World hystricomorph rodents. Univ. of Calif. Pub. Zool. **56**:1–118.

Larsson, N.-G., and D. A. Clayton. 1995. Molecular genetic aspects of human mitochondrial disorders. Ann. Rev. Genetics **29**:151–178.

Lavocat, R. 1951. Révision de la faune des Mammiféres Oligocène d'Auvergne et du Valay. Editions Science et Avenir, Paris.

Lavocat, R. 1973. Les rongeurs du Miocene d'Afrique oriental: 1 Miocene inferieur. Ec. Pract. Hautes Études Inst. Montpellier Mem. Trav. **1**:1–284.

Lavocat, R., and J.-P. Parent, 1985. Phylogenetic analysis of middle ear features in fossils and living rodents. Pp. 333–354 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis Series A: Vol. 92, Plenum, New York.

Lemmon, A. R., and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. **53**:265–277.

Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

Li, C.-K., and S.-Y. Ting. 1985. Possible phylogenetic relationship of Asiatic eurymylids and rodents. Pp. 35–58 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis Series A: Vol. 92, Plenum, New York.

Li, W. H., W. A. Hide, and D. Graur. 1992. Origins of rodents and guinea-pigs. Nature **359**: 277–278.

Li, W. H., W. A. Hide, A. Zharkikh, D. P. Ma, and D. Graur. 1992. The molecular taxonomy and evolution of the guinea pig. J. Hered. **83**:174–181.

Lin, Y., P. Waddell, and D. Penny. 2002. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. Gene **294**:119–129.

Linnaeus, C. 1758. Systema Naturae Per Regna Tria Naturae, Secundum Classes, Ordines, Genera, Species Cum Characteribus, Diffentiis, Synonymis, Locis. Stockholm, Laurentii Salvi.

Liu, F.-G. R., M. M. Miyamoto, N. P. Friere, P. Q. Ong, M. R. Tennent, T. S. Young, and K. F. Gugel. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. Science **291**:1786–1789.

Luckett, W. P. 1980. Monophyletic or diphyletic origin of Anthropoidea and Hystricognathi: evidence of the fetal membranes. Pp. 347–368 *in* R. L. Ciochon and B. Chiarelli. eds. Evolutionary Biology of the New World Monkeys and Continental Drift. Plenum, New York.

Luckett, W. P. 1985. Superordinal and intraordinal affinities of rodents: developmental evidence from the dentition and placentation. Pp. 227-276 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

Luckett, W. P., and J.-L. Hartenberger 1993. Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. J. Mamm. Evol. **1**:127–147.

Lynch, M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13**:209–220.

Lynch, M. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. Mol. Biol. Evol. **14**:914–925.

Ma, D.-P., A. Zharkikh, D. Graur, J. L. VandeBerg, and W. H. Li. 1993. Structure and evolution of opposum, guinea pig, and porcupine cytochrome b genes. J. Mol. Evol. **36**: 327–334.

Macey, J. R., A. Larson, N. B. Ananjeva, and T. J. Pappenfuss. 1997. Evolutionary shifts in three major structural features of the mitochondrial genome of iguanian lizards. J. Mol. Evol. **44**:660–674.

Macey, J. R., J. A. Schulte II, N. B. Ananjeva, A. Larson, N. Rastegar-Pouyani, S. M. Shammakov, and T. J. Pappenfuss. 1998. Phylogenetic relationships among agamid lizards of the *Laudakia caucasia* species group: testing hypotheses of biogeographic fragmentation and an area cladogram for the Iranian Plateau. Mol. Phylogenet. Evol. **10**:118–131.

Macey, J. R., J. A. Schulte II, and A. Larson. 2000. Evolution and phylogenetic information content of mitochondrial genomic structural features illustrated with acrodont lizards. Syst. Biol. **49**:257–277.

Macey, J. R., J. A. Schulte II, A. Larson, N. B. Ananjeva, Y. Wang, R. Pethiyagoda, N. Rastegar-Pouyani, and T. J. Pappenfuss. 2000a. Evaluating trans-Tethys migration: an example using acrodont lizard phylogenetics. Syst. Biol. **49**:233–256.

Macey, J. R., J. A. Schulte II, A. Larson, B. S. Tuniyev, N. Orlov, and T. J. Pappenfuss. 1999a. Molecular phylogenetics, tRNA evolution, and historical biogeography in anguid lizards and related taxonomic families. Mol. Phylogenet. Evol. **12**:250–272.

Macey, J. R., and A. Verma. 1997. Homology in phylogenetic analysis: alignment of transfer RNA genes and the phylogenetic position of snakes. Mol. Phylogenet. Evol. **7**:272–279.

Macey, J. R., Y. Wang, N. B. Ananjeva, A. Larson, and T. J. Pappenfuss,. 1999b. Vicariant patterns of fragmentation among gekkonid lizards of the genus *Teratoscincus* produced by the Indian collision: a molecular phylogenetic perspective and an area cladogram for central Asia. Mol. Phylogenet. Evol. **12**:320–332.

Maddison, D. R., and W. P. Maddison. 2002. MacClade 4: Analysis of Phylogeny and Character Evolution. Version 4.0. Sinauer Associates, Sunderland, MA.

Madsen, O., M. Scally, C. J. Douady, D. J. Kao, R. W. DeBry, R. Adkins, H. M. Amrine, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature **409**:610–614.

Marivaux, L., M. Vianey-Liaud, and J.- J. Jaeger. 2004. High-level phylogeny of early Tertiary rodents: dental evidence. Zool. J. Linn. Soc.-Lond. **142**:105–134.

Martin, T., 1994. African origin of caviomorph rodents is indicated by incisor enamel microstructure. Paleobiology 20:5–13.

Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288**:911–914.

Matthee, C. A., and T. J. Robinson. 1997. Molecular phylogeny of the springhare, *Pedetes capensis*, based on mitochondrial DNA sequences. Mol. Biol. Evol. **14**:20–29.

McKenna, M. 1975. Toward a phylogenetic classification of the Mammalia. Pp. 21–46. *in* W. P. Luckett and F. S. Szalay, eds. Phylogeny of the Primates, Plenum Press, New York, NY.

MITOMAP: A Human Mitochondrial Genome Database. 2005. (http://www.mitomap.org).

Miya, M., A. Kawaguchi, and M. Nishida, 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with

38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. **18**:1993–2009.

Miya, M., and M. Nishida. 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol. Phylogenet. Evol. **17**:437–455.

Miya, M., H. Takeshima, H. Endo, N. B. Ishiguro, J. G. Inoue et al. (7 co-authors). 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol. Phylogenet. Evol. **26**:121–138.

Montgelard, C., S. Bentz, C. Tiraud, O. Verneau, and F. M. Catzeflis. 2002. Molecular systematics of Sciurognathi (Rodentia): the mitochondrial cytochrome *b* and 12S rRNA genes support the Anomaluroidea (Pedetidae and Anomaluridae). Mol. Phylogenet. Evol. **22**:220–233.

Moriya, J., T. Yokogawa, K. Wakita, T. Ueda, K. Nishikawa et al. (8 co-authors). 1994. A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. Biochemistry **33**:2234–2239.

Mossman, H. W. 1987. Vertebrate Fetal Membranes. Rutgers University Press, New Brunswick, NJ.

Mouchaty, S. K., F. Catzeflis, A. Janke, and U. Arnason. 2001. Molecular evidence of an African Phiomorpha-South America Caviomorpha clade and support for Hystricognathi based on the complete mitochondrial genome of the cane rat (*Thryonomys swinderianus*). Mol. Phylogenet. Evol. **18**:127–135.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001*a*. Molecular phylogenetics and the origin of placental mammals. Nature **409**:614–618.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally et al. (6 co-authors). 2001*b*. Resolution of the early placental mammal radiation using Bayesian inference. Science **294**:2348–2351.

Nedbal, M. A. 1995. Higher level systematics of rodents: evidence from the mitochondrial 12S rRNA gene. Unpublished Ph.D. dissertation. Texas A&M University, College Station, TX.

Nedbal, M. A., M. W. Allard, and R. L. Honeycutt. 1994. Molecular systematics of hystricognath rodents: evidence from the mitochondrial 12S rRNA gene. Mol. Phylogenet. Evol. **3**:206–220.

Nedbal, M. A., R. L. Honeycutt, and D. A. Schlitter. 1996. Higher-level systematics of rodents (Mammalia, Rodentia): evidence from the mitochondrial 12S rRNA gene. J. Mamm. Evol. **3**:201–237.

Nilsson, M.A., U. Arnason, P. B. Spencer, and A. Janke. 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. Gene **340**:189–196.

Nilsson, M. A., A. Gullberg, A. E. Spotorno, U. Arnason, and A. Janke. 2003. Radiation of extant marsupials after the K/T boundary: evidence from complete mitochondrial genomes. J. Mol. Evol. **57**:S3–S12.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302**:205-217.

Novacek, M. J. 1985. Cranial evidence for rodent affinities. Pp 59–81 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

Novacek, M. J. 1992. Mammalian phylogeny: shaking the tree. Nature **356**:121–125.

Novacek, M. J. 2001. Mammalian phylogeny: genes and supertrees. Curr. Biol **11**:R573–R575.

Nowak, R. M., ed. 1999. Order Rodentia. Pp. 1243-1714 *in* Walker's Mammals of the World, Vol. II. Johns Hopkins University Press, Baltimore, MD.

Patterson, B., and A. E. Wood. 1982. Rodents from the Deseadan Oligocene of Bolivia and the relationships of the Caviomorpha. Bull. Mus. Comp. Zool. **149**:371–543.

Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.

Perna, N. T., and T. D. Kocher. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. **41**:353–358.

Philippe, H. 1997. Rodent monophyly: pitfalls of molecular phylogeny. J. Mol. Evol. **45**:712–715.

Poirot, O., K. Suhre, C. Abergel, E. O'Toole, and C. Notredame. 2004. 3-DCoffee@igs: a web server mixing sequences and structures into multiple sequence alignments. Nuc. Acids Res. **32**:W336-339.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of the Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. **53**:793–808.

Reyes, A., C. Gissi, F. Catzeflis, E. Nevo, G. Pesole, and C. Saccone. 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. Mol. Biol. Evol. **21**:397–403.

Reyes, A. C. Gissi, G. Pesole, F. M. Catzeflis, and C. Saccone. 2000. Where do rodents fit in? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. Mol. Biol. Evol. **17**:979–983.

Reyes, A., C. Gissi, G. Pesole, and C. Saccone. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. **15**:957–966.

Reyes, A., G. Pesole, and C. Saccone. 1998. Complete mitochondrial DNA sequence of the fat dormouse, *Glis glis*: further evidence of rodent paraphyly. Mol. Biol. Evol. **15**:499–505.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572–1574.

Rowe, D. L. 2002. Molecular phylogenetics and evolution of hystricognath rodents. Unpublished Ph.D. dissertation. Texas A&M University, College Station, TX.

Rowe, D. L., and R. L. Honeycutt. 2002. Phylogenetic relationships, ecological correlates, and molecular evolution within the Cavioidea (Mammalia, Rodentia). Mol. Biol. Evol. **19**:263–277.

Saccone, C., C. De Giorgi, C. Gissi, G. Pesole, and A. Reyes. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene **238**:195–209.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Sarich, V. M. 1985. Rodent macromolecular systematics. Pp. 423–452 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

Scally, M., O. Madsen, C. J. Douady, W. W. de Jong, M. J. Stanhope, and M. S. Springer. 2001. Molecular evidence for the major clades of placental mammals. J. Mamm. Evol. **8**:239–277.

Schmitz, J., M. Ohme, B. Suryobroto, and H. Zischler. 2002. The colugo (*Cynocephalus variegatus*, Dermoptera): the primates' gliding sister? Mol. Biol. Evol. **19**:2308-2312.

Schon, E. A., E. Bonilla, and S. DiMauro. 1997. Mitochondrial DNA mutations and pathogenesis. J. Bioenerg. Biomembr. **29**:131–149.

Schöniger, M., and A. von Haeseler. 1994. A stochastic model and the evolution of autocorrelated DNA sequences. Mol. Phylogenet. Evol. **3**:240–247.

Schultes, E. A., P. T. Hraber, and T. H. LaBean. 1999. Estimating the contributions of selection and self-organization in RNA secondary structure. J. Mol. Evol. **49**:76–83.

Shevchuk, N. A., and M. W. Allard. 2001. Sources of incongruence among mammalian mitochondrial sequences: COII, COIII, and ND6 genes are main contributors. Mol. Phylogenet. Evol. **21**:43–54.

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

Simpson, G. G. 1945. The principles of classification and the classification of mammals. B. Am. Mus. Nat. Hist. **85**:1–350.

Sissler, M., M. Helm, M. Frugier, R. Giegé, and C. Florentz. 2004. Aminoacylation properties of pathology-related human mitochondrial $tRNA^{Lys}$ variants. RNA **10**:841–853.

Sorenson, M.D. 1999. TreeRot, version 2. Boston University, Boston, MA.

Springer, M. S., and E. Douzery. 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. J. Mol. Evol. **43**:357–373.

Springer, M. S., L. J. Hollar, and A. Burk. 1995. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. Mol. Bio. Evol. **12**:1138–1150.

Springer, M. S., M. Westerman, J. R. Kavanagh, A. Burk, M. O. Woodburne, D. J. Kao, and C. Krajewski. 1998. The origin of the Australasian marsupial fauna and the phylogenetic affinities of the enigmatic monito del monte and marsupial mole. Proc. Biol. Sci. **265**:2381-2386.

Sprinzl, M., C. Horn, M. Brown, A. Loudovitch, and S. Steinberg. 1998. Compilation of tRNA sequences and sequences of tRNA genes. Nuc. Acids Res. **26**:148–153. (http://www.uni-bayreuth.de/departments/biochemie/sprinzl/trna/index.html).

Steinberg, S., D. Gautheret, and R. Cedergren. 1994. Fitting the structurally diverse animal mitochondrial $tRNA^{Ser}$ to common three-dimensional constraints. J. Mol. Biol. **236**:982–989.

Sullivan, J., and D. L. Swofford. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mamm. Evol. **4**:77–86.

Swofford, D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D. L., and D. P. Begle. 1993. Phylogenetic Analysis Using Parsimony (Version 3.1), User's Manual, Smithsonian Institute Laboratory of Molecular Systematics, Champaign, IL.

Szalay, F. S. 1977. Phylogenetic relationships and a classification of the eutherian Mammalia. Pp. 375–394 *in* M. K. Hecht, P. C. Goody, and B. M. Hecht, eds. Major Patterns of Vertebrate Evolution, Plenum Press, New York, NY.

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599–607.

Takemoto, C., T. Koike, T. Yokogawa, L. Benkowski, L. L. Spremulli, T. A. Ueda, K. Nishikawa, and K. Watanabe. 1995. The ability of mitochondrial transfer RNA^{Met} to decode AUG and AUA codons. Biochimie **77**:104–108.

Thaler, L. 1966. Les rongeurs fossils du Bas-Languedoc dans leurs rapports avec l'histoire des faunes et la stratigraphie du Tertiarie d'Europe. Mem. Mus. Natl. Hist. Paris Ser. C **17**:1–296.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weigh matrix choice. Nuc. Acids Res. **22**:4673–4680.

Tullberg, T. 1899. Uber das system der Nagertiere: eine phylogenetische studie. Nova Acta Reg. Soc. Sci. Uppsala **18**:1–514.

Vianey-Liaud, M., 1985. Possible evolutionary relationships among Eocene and Lower Oligocene rodents of Asia, Europe, and North America. Pp. 277–310 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

von Koenigswald, W. 1985. Evolutionary trends in the enamel of rodent incisors. Pp. 403–422 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York, NY.

Waddell, P. J., H. Kishino, and R. Ota. 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. **12**:141–154.

Waddell, P. J., and S. Shelley. 2003. Evaluating placental inter-ordinal phylogenies with novel sequences including RAG1, γ-fibrinogen, ND6, and mt-tRNA, plus

MCMC-driven nucleotide, amino acid, and codon models. Mol. Phylogenet. Evol. **28**:197–224.

Wahlert, J. H. 1978. Cranial foramina and relationships of Eomyoidea (Rodentia, Geomyoidea): skull and upper teeth of *Kansasimys*. Am. Mus. Novit. **2645**:1–16.

Wahlert, J. H. 1985. Cranial foramina of rodents. Pp. 311–332 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

Wallace, D. C. 1992. Diseases of the mitochondrial DNA. Ann. Rev. Biochem. **61**:1175–1212.

Wallace, D. C. 1999. Mitochondrial diseases in man and mouse. Science **283**:1482–1488.

Watanabe, Y.-I., H. Tsurui, T. Ueda, R. Furushima, S. Takamiya, K. Kita, K. Nishikawa, and K. Watanabe. 1994. Primary and higher order structures of nematode (*Ascaris suum*) mitochondrial tRNAs lacking either the T or D stem. J. Biol. Chem. **269**:22902–22906.

Weinreich, D. M. 2001. The rates of molecular evolution in rodent and primate mitochondrial DNA. J. Mol. Evol. **52**:40–50.

Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. **25**:361–371.

Wilson, A. C., R. L. Cann, S. M. Carr, W. George, U. B. Gyllensten et al. (6 co-authors). 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26**:375–400.

Wilson, D. E., and D. M. Reeder. 1993. Mammal Species of the World: a Taxonomic and Geographic Reference, Smithsonian Institution, Washington, D.C.

Wilson, R. W. 1949. Early Tertiary rodents of North America. Carnegie Inst. Wash. Pub. **584**:67–164.

Wilson, R. W. 1986. The Paleogene record of rodents: facts and interpretation. Pp. 163–176 *in* K. M. Flanagan and J. A. Lillegraven, eds. Vertebrates, Phylogeny, and Philosophy, Contributions in Geology, University of Wyoming, Special Paper 3, Laramie, WY.

Wolstenholme, D. R., J. L. MacFarlane, R. Okimoto, D. O. Clary, and J. A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA **84**:1324–1328.

Wood, A. E. 1937. The mammalian fauna of the White River Oligocene. Part II. Rodentia. Trans. Am. Philo. Soc. **28**:155–269.

Wood, A. E. 1955. A revised classification of rodents. J. Mamm. **36**:165–187.

Wood, A. E. 1959. Eocene radiation and phylogeny of the rodents. Evolution **13**:354–361.

Wood, A. E. 1965. Grades and clades among rodents. Evolution **19**:115–130.

Wood, A. E. 1980. The Oligocene rodents of North America. Trans. Am. Philo. Soc. **70**:1–68.

Wood, A. E. 1985. The relationships, origins, and dispersal of the hystricognathous rodents. Pp. 475–513 *in* W. P. Luckett and J.-L. Hartenberger, eds. Evolutionary Relationships among Rodents: a Multidisciplinary Analysis, Series A: Vol. 92, Plenum, New York.

Xu, X., and U. Arnason. 1994. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. Gene **148**:357–362.

Xu, X., and U. Arnason. 1996. The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. J. Mol. Evol. **43**: 431–437.

Yang, Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

Yang, Z. 1997. How often do wrong models produce better phylogenies? Mol. Biol. Evol. **14**:105–108.

Yokogawa, T., Y. Watanabe, Y. Kumazawa, T. Ueda, I. Hirao, K. Miura, and K. Watanabe. 1991. A novel cloverleaf structure found in mammalian $tRNA^{Ser}$(UCN). Nuc. Acids Res. **19**:6101–6105.

Zardoya, R., and A. Meyer. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. **13**:933–942.

Zittel, K. A. 1893. Handbuch der Paleoontologie, Sect. I, Paleozoologie, Vol. IV, Vertebrata (Mammalia), Oldenbourg, Munich.

Zuker, M. 2003. Mfold: web server for nucleic acid folding and hybridization prediction. Nuc. Acids Res. **31:**3406–3415.

Zuker, M., D. H. Mathews, and D. H. Turner. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. Pp. 11–43 *in* J. Barciszewski and B.F.C. Clark, eds. RNA Biochemistry and Biotechnology, NATO ASI Series. Kluwer Academic Pub., Boston, MA.

Zwickl, D. J., and M. T. Holder. 2004. Model parameterization, prior distributions and the general time-reversible model in Bayesian phylogenetics. Syst. Biol. **53**:877–888.

APPENDIX 1

Supplemental file available online. The multiple sequence alignment of the 12 H-strand encoded protein-coding mitochondrial genes analyzed in Chapter II. File is plain-text, NEXUS format; includes gene partitions, codon positions, and MrBayes block.

Filename: Frabotta_2005_Dissertation_mtDNA_protein-coding_Alignment.nex

APPENDIX 2

List of taxa included in this study with classification, species names, common names, OGRe database (Jameson et al., 2003) abbreviations, and GenBank accession numbers

Classification	Species	Common Name	OGRe abbreviation	GenBank Accession
Monotremata				
Ornithorhynchidae	*Ornithorhynchus anatinus*	Duckbill Platypus	ORNANA	X83427
Tachyglossidae	*Tachyglossus aculeatus*	Australian Echidna	TACACU	AJ303116
Metatheria (Marsupials)				
Didelphimorphia	*Didelphis virginiana*	North American Opossum	DIDVIR	Z29573
	Thylamys elegans	Elegant Fat-tailed Opossum	THYELE	AJ508401
Diprotodontia	*Trichosurus vulpecula*	Silver-gray Brushtail Possum	TRIVUL	AF357238
	Macropus robustus	Wallaroo	MACROB	Y10524
	Vombatus ursinus	Common Wombat	VOMURS	AJ304826
Microbiotheria	*Dromiciops gliroides*	Monito del Monte	DROGLI	AJ508402
Paucituberculata	*Rhyncholestes raphanurus*	Chilean Shrew Opossum	RHYRAP	AJ508399
	Caenolestes fuliginosus	Silky Shrew Opossum	CAEFUL	AJ508400
Peramelemorphia	*Isoodon macrourus*	Northern Brown Bandicoot	ISOMAC	AF358864
Eutheria				
Afrosoricida	*Chrysochloris asiatica*	Cape Golden Mole	CHRASI	AB096866
	Echinops telfairi	Small Malagasy Hedgehog Tenrec	ECHTEL	AJ400734
Edentata	*Dasypus novemcinctus*	Nine-banded Armadillo	DASNOV	Y11832
	Tamandua tetradactyla	Southern Tamandua	TAMTET	AJ421450
Hyracoidea	*Procavia capensis*	Cape Rock Hyrax	PROCAP	AB096865
Macroscelidea	*Elephantulu ssp. VB001*	Elephant Shrew	ELESPV	AB096867
	Macroscelides proboscideus	Short-eared Elephant Shrew	MACPRO	AJ421452
Proboscidea	*Elephas maximus*	Asiatic Elephant	ELEMAX	AJ428946*
	Loxodonta africana	African Elephant	LOXAFR	AJ224821
Sirenia	*Dugong dugon*	Dugong	DUGDUG	AJ421723
Tubulidentata	*Orycteropus afer*	Aardvark	ORYAFE	Y18475
Carnivora	*Ursus americanus*	American Black Bear	URSAME	AF303109
	Arctocephalus forsteri	NewZealand Fur Seal	ARCFOR	AF513820
	Canis familiaris	Dog	CANFAM	U96639
	Eumetopias jubatus	Steller Sea Lion	EUMJUB	AJ428578

Classification	Species	Common Name	OGRe abbreviation	GenBank Accession
	Odobenus rosmarus	Atlantic Walrus	ODOROS	AJ428576
	Acinonyx jubatus	Cheetah	ACIJUB	AF344830
	Phoca vitulina	Harbor Seal	PHOVIT	X63726
	Ursus maritimus	Polar Bear	URSMAR	AF303111
	Felis catus	Cat	FELCAT	U20753
	Ursus arctos	Brown Bear	URSARC	AF303110
	Halichoerus grypus	Gray Seal	HALGRY	X72004
Pholidota	*Manis tetradactyla*	Long-tailed Pangolin	MANTET	AJ421454
Cetartiodactyla	*Berardius bairdii*	Baird's Beaked Whale	BERBAI	AJ554057
	Balaenoptera musculus	Blue Whale	BALMUS	X72204
	Pontoporia blainvillei	Franciscana	PONBLA	AJ554060
	Capra hircus	Goat	CAPHIR	AF533441
	Monodon monoceros	Narwhal	MONMON	AJ554062
	Hyperoodon ampullatus	Northern Bottlenose Whale	HYPAMP	AJ554056
	Ovis aries	Sheep	OVIARI	AF010406
	Caperea marginata	Pygmy Right Whale	CAPMAR	AJ554052
	Balaenoptera acutorostrata	Minke Whale	BALACU	AJ554054
	Muntiacus muntjak	Muntjak	MUNMUN	AY225986
	Sus scrofa	Pig	SUSSCR	AF034253
	Phocoena phocoena	Harbor Porpoise	PHOPHO	AJ554063
	Kogia breviceps	Pygmy Sperm Whale	KOGBRE	AJ554055
	Muntiacus crinifrons	Black Muntjac	MUNCRI	AY239042
	Lama pacos	Alpaca	LAMPAC	Y19184
	Bos taurus	Cow	BOSTAU	V00654
	Hippopotamus amphibious	Hippopotamus	HIPAMP	AJ010957
	Lagenorhynchus albirostris	White-beaked Dolphin	LAGALB	AJ554061
	Balaena mysticetus	Bowhead Whale	BALMYS	AJ554051
	Inia geoffrensis	Boutu	INIGEO	AJ554059
	Eschrichtius robustus	Grey Whale	ESCROB	AJ554053
	Muntiacus reevesi	Chinese Muntjac	MUNREE	AF527537
	Platanista minor	Indus River Dolphin	PLAMIN	AJ554058
	Bos indicus	Zebu	BOSIND	AY126697
	Balaenoptera physalus	Finback Whale	BALPHY	X61145
	Physeter catodon	Sperm Whale	PHYCAT	AJ277029
Perissodactyla	*Equus asinus*	Ass	EQUASI	X97337
	Tapirus terrestris	Brazilian Tapir	TAPTER	AJ428947

Classification	Species	Common Name	OGRe abbreviation	GenBank Accession
	Equus caballus	Horse	EQUCAB	X79547
	Rhinoceros unicornis	Greater Indian Rhinoceros	RHIUNI	X97336
	Ceratotherium simum	White Rhinoceros	CERSIM	Y07726
Chiroptera	*Pteropus scapulatus*	Little Red Flying Fox	PTESCA	AF321050
	Pipistrellus abramus	Japanese House Bat	PIPABR	AB061528
	Pteropus dasymallus	Ryukyu Flying Fox	PTEDAS	AB042770
	Rhinolophus monoceros	Formosan Lesser Horseshoe Bat	RHIMON	AF406806
	Artibeus jamaicensis	Jamaican Fruit-eating Bat	ARTJAM	AF061340
	Rhinolophus pumilus	Okinawa Least Horseshoe Bat	RHIPUM	AB061526
	Chalinolobus tuberculatus	NewZealand Long-tailed Bat	CHATUB	AF321051
Eulipotyphla	*Urotrichus talpoides*	Japanese Shrew Mole	UROTAL	AB099483
	Hemiechinus auritus	Long-eared Hedgehog	HEMAUR	AB099481
	Mogera wogura	Japanese Mole	MOGWOG	AB099482
	Sorex unguiculatus	Long-clawed Shrew	SORUNG	AB061527
	Talpa europaea	European Mole	TALEUR	Y19192
	Echinosorex gymnura	Moonrat	ECHGYM	AF348079
	Erinaceus europaeus	Western European Hedgehog	ERIEUR	X88898
	Episoriculus fumidus	Taiwanese Brown-toothed Shrew	SORFUM	AF348081
Primates	*Cebus albifrons*	White-fronted Capuchin	CEBALB	AJ309866
	Nycticebus coucang	Slow Loris	NYCCOU	AJ309867
	Pongo abelii	Sumatran Orangutan	PONPY1	X97707
	Lemur catta	Ring-tailed Lemur	LEMCAT	AJ421451
	Papio hamadryas	Baboon	PAPHAM	Y18001
	Pongo pygmaeus	Orangutan	PONPYG	D38115
	Homo sapiens	Human	HOMSAP	AF347015
	Tarsius bancanus	Western Tarsier	TARBAN	AF348159
	Pan paniscus	Bonobo	PANPAN	D38116
	Macaca sylvanus	Barbary Ape	MACSYL	AJ309865
	Gorilla gorilla	Gorilla	GORGOR	D38114
	Pan troglodytes	Chimpanzee	PANTRO	D38113
	Hylobates lar	Common Gibbon	HYLLAR	X99256
	Macaca mulatta	Rhesus Monkey	MACMUL	AY612638
Dermoptera	*Cynocephalus variegatus*	Malayan Flying Lemur	CYNVAR	AJ428849
Scandentia	*Tupaia belangeri*	Northern Tree Shrew	TUPBEL	AF217811
Lagomorpha	*Oryctolagus cuniculus*	Rabbit	ORYCUN	AJ001588
	Ochotona collaris	Rock Rabbit	OCHCOL	AF348080
	Lepus europaeus	European Hare	LEPEUR	AJ421471
	Ochotona princeps	American Pika	OCHPRI	AJ537415

Classification	Species	Common Name	OGRe abbreviation	GenBank Accession
Rodentia	*Sciurus vulgaris*	Eurasian Red Squirrel	SCIVUL	AJ238588
	Jaculus jaculus	Lesser Egyptian Jerboa	JACJAC	AJ416890
	Myoxus glis	Fat Dormouse	MYOGLI	AJ001562
	Mus musculus	House Mouse	MUSMUS	AY172335
	Nannospalax ehrenbergi	Ehrenberg's Blind Mole-rat	NANEHR	AJ416891
	Thryonomys swinderianus	Greater Cane Rat	THRSWI	AJ301644
	Volemys kikuchii	Taiwanese Vole	VOLKIK	AF348082
	Rattus norvegicus	Norway Rat	RATNOR	X14848
	Cavia porcellus	Domestic Guinea Pig (Cavy)	CAVPOR	AJ222767
New Rodents[1]				
	Abrocoma cinerea	Rat Chinchilla	ABRCIN	(NK30665)
	Agouti paca	Spotted Paca	AGOPAC	(K7)
	Aplodontia rufa	Sewellel (Mountain Beaver)	APLRUF	(H2370)
	Bathyergus janetta	Namaqua Dune Mole-rat	BATJAN	(M63565)
	Capromys pilorides	Cuban Hutia	CAPPIL	(H575)
	Capromys pilorides	Cuban Hutia	CAPPI2	(TK32007)
	Castor canadensis	North American Beaver	CASCAN	(H2205)
	Cavia aperea	Aperea (Wild Cavy)	CAVAPE	(H586/TK17830)
	Chinchilla lanigera	Chinchilla	CHILAN	(H647/K94)
	Cratogeomys castanops	Yellow-faced Pocket Gopher	CRACAS	(H110)
	Cryptomys hottentotus	Common Mole-rat	CRYHOT	(M63567)
	Ctenomys boliviensis	Tuco-Tuco	CTEBOL	(NK15277)
	Dasyprocta punctata	Agouti	DASPUN	(NK14094)
	Dinomys branickii	Pacarana	DINBRA	(K8)
	Dolichotis salinicola	Salt-Desert Cavy	DOLSAL	(AK14046)
	Erethizon dorsatum	North American Porcupine	EREDOR	(H648)
	Galea musteloides	Cuis	GALMUS	(AK13818)
	Gerbillurus vallinus	Brush-tailed Hairy-footed Gerbil	GERVAL	(H558/SP4232)
	Graphiurus murinus	African Dormouse	GRAMUR	(H687/SP5577)
	Heliophobius argenteocinereus	Silvery Mole-rat	HELARG	(M63562)
	Heterocephalus glaber	Naked Mole-rat	HETGLA	(H015)
	Hystrix africaeaustralis	South African Porcupine	HYSAFR	(SP7702)
	Jaculus jaculus	Lesser Egyptian Jerboa	JACJA2	(SP10206)
	Microcavia australis	Southern Mountain Cavy	MICAUS	(AK13309)

Classification	Species	Common Name	OGRe abbreviation	GenBank Accession
	Myocastor coypus	Coypu (Nutria)	MYOCOY	(H3919)
	Octodon degus	Degu	OCTDEG	(H645/K61)
	Octomys mimax	Viscacha Rat	OCTMIM	(AK13474)
	Pedetes capensis	Cape Springhaas (Springhare)	PEDCAP	(H551/SP6352)
	Perognathus flavus	Silky Pocket Mouse	PERFLA	(AK10368)
	Petromus typicus	Rock Rat	PETTYP	(H550/M63571)
	Proechimys longicaudatus	Long-tailed Spiny Rat	PROLON	(H582/TK22841)
	Thryonomys swinderianus	Greater Cane Rat	THRSW2	(H571/M63570)

* Record recently removed from GenBank at the submitter's request. 'New rodent taxa include collection information referenced as follows: AK (John Bickham, Wildlife & Fisheries Sciences, Texas A&M University); H (Rodney Honeycutt, Wildlife & Fisheries Sciences, Texas A&M University; K (William Kilpatrick, Zadock Thompson Natural History Collections, Biology, University of Vermont); M (Hennie Erasmus, MacGregor Museum, South Africa); NK (Terry Yates, Museum of Southwestern Biology, University of New Mexico); SP (Duane Schlitter and Sue MacLaren, Carnegie Museum of Natural History, Pittsburgh); TK (Robert Baker, Museum of Texas Tech University).

APPENDIX 3

Supplemental file available online. The multiple sequence alignment of the 22 mitochondrial tRNA genes analyzed in Chapter III. File is plain-text, NEXUS format; includes tRNA gene partitions and secondary structure mask suitable for use in PHASE (http://www.bioinf.man.ac.uk/resources/phase/).

Filename: Frabotta_2005_Dissertation_tRNA_Alignment.nex

APPENDIX 4

Supplemental file available online. Nexus tree file in NEWICK tree format used to estimate transition:transversion ratios in Chapter III. Appendix 3 required to estimate values. File is plain-text, NEXUS format.

Filename: Frabotta_2005_Dissertation_tRNA_TS_TV_Nexus_tree.tre

APPENDIX 5

Nucleotide Frequencies, %GC, and Skew Measures for mt tRNAs of Mammalian Orders*

tRNA	%A	%C	%G	%T	%GC	AT-Skew	GC-Skew
H-Strand Encoded							
Monotremata	0.353	0.157	0.180	**0.310**	0.337	**0.065**	-0.070
	0.300	**0.192**	0.197	**0.311**	0.389	**-0.017**	0.011
	0.434	0.161	0.096	**0.309**	0.257	**0.168**	-0.253
Metatheria	0.369	**0.146**	0.182	0.303	**0.328**	0.098	-0.108
	0.308	0.193	0.191	0.308	**0.384**	0.001	-0.005
	0.460	0.165	**0.078**	0.296	0.244	0.218	-0.352
Afrotheria*	0.364	0.156	0.192	0.288	0.348	0.118	-0.104
	0.297	0.203	0.200	0.300	0.403	-0.004	-0.009
	0.468	0.175	0.089	0.269	0.263	0.271	-0.329
Edentata	0.365	0.160	0.202	0.273	0.362	**0.145**	-0.118
	0.289	0.217	0.208	0.285	0.425	0.007	-0.020
	0.481	0.180	0.085	0.254	0.265	**0.309**	-0.357
Carnivora	0.358	0.164	0.191	0.287	0.355	0.111	-0.076
	0.293	0.201	0.205	0.301	0.406	-0.015	0.010
	0.458	0.176	0.101	0.265	0.277	0.268	-0.268
Pholidota	**0.372**	0.152	0.183	0.294	0.335	0.117	-0.094
	0.303	0.205	0.198	0.295	0.395	**0.014**	-0.017
	0.477	0.150	0.080	0.292	**0.231**	0.240	-0.302
Cetartiodactyla	0.357	0.164	0.196	0.282	0.360	0.118	-0.089
	0.290	0.211	0.208	0.291	0.419	0.000	-0.007
	0.459	0.174	0.097	0.270	0.272	0.260	-0.280

Perissodactyla	0.364	0.157	0.199	0.281	0.356	0.129	-0.116
	0.299	0.202	0.202	0.297	0.404	0.004	-0.002
	0.462	0.193	0.089	0.256	0.282	0.287	-0.366
Chiroptera	0.365	0.160	0.179	0.296	0.339	0.104	-0.055
	0.305	**0.192**	**0.198**	0.306	0.390	-0.002	**0.017**
	0.456	0.160	0.102	0.282	0.262	0.237	-0.218
Eulipotyphla	0.359	0.160	**0.178**	0.302	0.338	0.086	**-0.052**
	0.293	0.201	0.205	0.301	0.406	-0.014	0.009
	0.460	**0.143**	0.093	0.304	0.236	0.205	-0.213
Primates	0.359	0.155	0.210	0.276	0.365	0.132	**-0.149**
	0.293	0.210	0.201	0.295	0.411	-0.004	**-0.022**
	0.461	**0.210**	0.085	0.245	0.295	0.306	**-0.420**
Dermoptera	**0.347**	0.168	**0.214**	0.271	0.382	0.122	-0.121
	0.282	0.216	0.211	0.291	0.427	-0.015	-0.012
	0.447	**0.210**	0.101	**0.242**	**0.311**	0.297	-0.350
Scandentia	0.349	**0.178**	0.205	**0.268**	**0.383**	0.131	-0.069
	0.280	**0.218**	**0.218**	**0.284**	**0.436**	-0.006	0
	0.458	0.184	**0.115**	0.244	0.299	0.305	-0.229
Lagomorpha	0.349	0.167	0.198	0.286	0.365	0.099	-0.085
	0.288	0.211	0.208	0.293	0.418	-0.008	-0.007
	0.442	0.179	0.104	0.275	0.283	0.233	-0.261
Rodentia	0.368	0.161	0.180	0.291	0.341	0.117	-0.058
	0.300	0.196	0.199	0.306	0.394	-0.011	0.008
	0.473	0.157	0.102	0.269	0.259	0.276	**-0.214**
L-Strand Encoded							
Monotremata	0.302	0.243	0.139	0.317	0.382	**-0.024**	0.271
	0.253	0.189	0.259	0.299	0.448	**-0.084**	0.157

	0.379	0.061	0.217	0.344	0.278	0.048	0.559
Metatheria	**0.305**	0.236	**0.130**	0.329	**0.366**	-0.038	0.288
	0.254	**0.176**	**0.245**	**0.325**	**0.421**	-0.123	0.166
	0.384	0.060	0.221	**0.335**	0.281	**0.068**	0.572
Afrotheria*	0.270	0.251	0.145	0.334	0.396	-0.107	0.268
	0.227	0.200	0.271	0.303	0.470	-0.143	0.150
	0.336	0.060	0.221	0.383	0.281	-0.064	0.572
Edentata	0.260	0.263	0.146	0.331	0.409	-0.121	0.285
	0.220	0.196	0.280	0.304	0.476	-0.160	**0.176**
	0.320	0.070	0.236	0.373	0.307	-0.075	0.541
Carnivora	0.267	0.246	0.152	0.336	0.398	-0.115	**0.238**
	0.223	0.204	0.271	0.302	0.476	-0.151	0.142
	0.334	0.070	0.208	0.388	0.278	-0.075	0.493
Pholidota	0.267	0.240	0.145	**0.347**	0.385	-0.130	0.245
	0.219	0.204	0.269	0.308	0.473	**-0.170**	0.139
	0.343	**0.056**	**0.194**	**0.407**	**0.250**	-0.086	0.556
Cetartiodactyla	0.271	0.250	0.142	0.337	0.392	-0.108	0.276
	0.235	0.193	0.263	0.309	0.456	-0.135	0.154
	0.326	0.064	0.230	0.379	0.295	-0.076	0.565
Perissodactyla	0.268	0.256	0.145	0.331	0.401	-0.105	0.276
	0.225	0.200	0.277	0.298	0.477	-0.139	0.162
	0.335	0.060	0.223	0.383	0.283	-0.067	0.574
Chiroptera	0.268	0.250	0.148	0.334	0.398	-0.110	0.256
	0.221	0.204	0.277	0.298	0.481	-0.148	0.151
	0.340	0.062	0.209	0.390	0.271	-0.069	0.540
Eulipotyphla	0.294	**0.232**	0.136	0.338	0.368	-0.070	0.262
	0.249	0.187	0.255	0.310	0.442	-0.110	0.155
	0.364	0.057	0.197	0.382	0.254	-0.024	0.547

Primates	0.252	0.274	0.147	0.327	0.421	-0.130	**0.302**
	0.227	0.200	0.278	0.294	0.478	-0.129	0.163
	0.290	0.064	0.268	0.377	0.333	-0.132	**0.610**
Dermoptera	**0.228**	**0.295**	**0.168**	**0.309**	**0.463**	**-0.152**	0.273
	0.199	**0.229**	**0.304**	**0.268**	**0.533**	-0.147	0.140
	0.272	0.074	**0.281**	0.373	**0.355**	**-0.157**	0.584
Scandentia	0.257	0.250	0.161	0.332	0.411	-0.126	0.216
	0.226	0.214	0.268	0.292	0.482	-0.126	**0.111**
	0.306	**0.079**	0.222	0.394	0.301	-0.126	**0.477**
Lagomorpha	0.253	0.273	0.160	0.314	0.433	-0.107	0.260
	0.216	0.215	0.291	0.278	0.506	-0.127	0.150
	0.312	0.075	0.245	0.369	0.320	-0.084	0.535
Rodentia	0.279	0.241	0.143	0.336	0.384	-0.093	0.257
	0.246	0.191	0.251	0.311	0.442	-0.117	0.135
	0.330	0.069	0.227	0.375	0.295	-0.065	0.533

Values for each category are all sites (top), stem positions (middle), and loop positions (bottom). *The superorder Afrotheria includes members of the following 6 orders: Afrosoricida, Hyracoidea, Macroscelidea, Proboscidea, Sirenia, and Tubulidentata. Minimum and maximum values in each category are bolded.

VITA

Laurence John Frabotta

I. Personal Information

Address: c/o Dr. Rodney L. Honeycutt
Department of Wildlife & Fisheries Sciences
2258 TAMUS
College Station, TX 77843-2258

II. Education

M.S., Biology, California State University, Long Beach, January 2002
B.S., Zoology, California State University, Long Beach, August 1995

III. Awards & Fellowships

Kenneth L. Johnson Award for Outstanding Thesis in the Natural Sciences and Mathmatics 2002
Phi Kappa Phi Honor Society ..... Inducted 1999
Howard Hughes Fellow 1994, 1995
CSULB President's List ..... Spring 1994
CSULB Dean's List ..... 1993 - 1995
National Dean's List ..... 1994 - 1995