

TESTING MACROEVOLUTIONARY HYPOTHESES: DIVERSIFICATION AND  
PHYLOGENETIC IMPLICATIONS

by

HUGO ALAMILLO

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY  
School of Biological Sciences

December 2010

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of HUGO ALAMILLO find it satisfactory and recommend that it be accepted.

---

Michael E. Alfaro, Ph.D., Chair

---

Luke J. Harmon, Ph.D.

---

Eric Roalson, Ph.D.

---

Gary Thorgaard, Ph.D.

## ACKNOWLEDGMENT

Thank you to my major advisor, Michael Alfaro, for many productive conversations and for the many hours he spent helping me focus my thoughts and words. I am also indebted to Luke Harmon and Eric Roalson for all the theoretical, psychological and practical advice that they gave me. Gary Thorgaard was always a source of academic information as I witnessed first-hand from him how a departmental chair handles a biology department. Jack Sullivan served in my committee for several years and although he did not serve through the end purely for practical reasons, he was an excellent source of theoretical knowledge and a great model of confidence.

I am also extremely indebted to my parents, Andrés and Eugenia Alamillo, for struggling emotionally and financially to migrate from Mexico and invest in my education. Without them I likely would not have known the amazing opportunities that being in The United States has afforded me.

Individuals that I am thankful for also having shaped my professional perspective and fomented my success during my doctorate include, Chad Brock, Devin Drown, John Schenk, Anne Maglia, John Simmons, Christopher Sheil, Linda Trueb, Bill Duellman and Ed Myers.

Lastly, I thank my wife and colleague, Barb Banbury. From our early days as undergraduates she has been a major source of ideas and a pillar of emotional stability.

TESTING MACROEVOLUTIONARY HYPOTHESES: DIVERSIFICATION AND  
PHYLOGENETIC IMPLICATIONS

Abstract

by Hugo Alamillo, Ph.D.  
Washington State University  
December 2010

Chair: Michael E. Alfaro

This dissertation research tests macroevolutionary hypotheses in a large vertebrate group (Reptilia: Serpentes), and detects methodological problems of using molecular data to test macroevolutionary hypotheses. The first chapter examines lineage diversification of snakes. It is hypothesized that snake evolution underwent major lineage diversification after gaining the ability to swallow large prey and again in the Miocene during the so-called “age of snakes”. I use two new methods and a time-tree to test these hypotheses. I find no evidence for a diversification rate shift following the evolution of large gape size. In contrast, Neogene snake diversification was significantly higher than over any other geologic time interval in snake evolutionary history. This elevated diversification appears to be driven by the evolutionary radiation of the colubroids, which exhibit diversification rates 2.5 times greater than other related snake lineages. The second chapter examines mimicry of head shapes between slug-eating, false-viper, and viper snakes. Viper mimicry is hypothesized to have produced fascinating mimics, which might have led to increased diversification. I provide the first quantitative test of viper mimicry. I use traditional tests, the latest morphological model-fitting approaches and a new time tree to test this hypothesis. I find that head shape has evolved towards similar selective optima, when specific

viper and false-viper clades are included in the models. This suggests that the central adaptive hypothesis of mimicry between slug-eaters and viperids likely occurred. The third chapter surveys effects of using incomplete molecular data matrices to infer phylogenetic tree branches of a clade that had a fast molecular rate of evolution. I simulate phylogenies with increasing missing-data levels that are deleted throughout the tree randomly or taxonomically. I compare simulated-tree branch lengths to true-tree branch lengths using sum of branch lengths and a commonly used macroevolutionary measure that detects node distributions as a proxy for slowdowns in diversification rates. I find the amount of missing data has more impact in branch length estimation than presence of a fast radiating clade.

## TABLE OF CONTENTS

	Page
ABSTRACT .....	iv
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER 1	
ABSTRACT .....	2
INTRODUCTION .....	3
MATERIALS AND METHODS.....	5
RESULTS.....	9
DISCUSSION .....	12
ACKNOWLEDGMENTS.....	19
LITERATURE CITED.....	20
SUPPLEMENTARY MATERIALS.....	26
CHAPTER 2	
ABSTRACT .....	30
INTRODUCTION .....	31
MATERIALS AND METHODS.....	34
RESULTS.....	46
DISCUSSION .....	51

ACKNOWLEDGMENTS.....	57
LITERATURE CITED.....	58
SUPPLEMENTARY MATERIALS.....	62

## CHAPTER 3

ABSTRACT.....	71
INTRODUCTION .....	72
MATERIALS AND METHODS.....	75
RESULTS.....	79
DISCUSSION .....	86
ACKNOWLEDGMENTS.....	95
LITERATURE CITED.....	96
SUPPLEMENTARY MATERIALS.....	99

## LIST OF TABLES

<b>CHAPTER 1</b>		<b>page</b>
Table 1. Fossil calibrations used in this study .....	6	
Table 2. Taxonomic Richness .....	8	
Table 3. Crown group ages .....	13	
Table 4. The tempo of snake diversification .....	16	
<b>CHAPTER 2</b>		
Table 1. Fossil Calibrations used to build the colubroid chronogram.....	36	
Table 2. Model fitting results for Geiger and OUCH methods for both PC1 and PC2 data....	53	
<b>CHAPTER 3</b>		
Table 1. Comparisons of branch lengths estimated.....	84	
Table 2. Comparisons of gamma values estimated.....	90	

## LIST OF FIGURES

CHAPTER 1	page
Figure 1. GTI test.....	10
Figure 2. Chronogram of the major groups of snakes .....	11
Figure 3. Plot of extant diversity of snake groups sampled in this study as a function of stem age .....	15
CHAPTER 2	
Figure 1. Snake mimicry.....	32
Figure 2. Groups included in this study .....	33
Figure 3. Head morphological measurements used in this study .....	38
Figure 4. Brownian motion and Single Optimum model used in <i>OUCH</i> .....	42
Figure 5. Viper-Colubrid model used in <i>OUCH</i> .....	43
Figure 6. Multiple Optima model used in <i>OUCH</i> .....	44
Figure 7. Multiple-DTT model used in <i>OUCH</i> .....	45
Figure 8. Chronogram used for the morphological analyses in this study.....	47
Figure 9. Plot of loadings from PC1 and PC2.....	49
Figure 10. Disparity through time plot .....	50

## CHAPTER 3

Figure 1. Flow chart of steps and methods used in this study .....	77
Figure 2. Pruning Strategy vs. Branch Length.....	82
Figure 3. Proportion Pruned vs. Branch Lengths.....	83
Figure 4. Molecular Evolution vs. Branch Lengths.....	85
Figure 5. Pruning Strategy vs. Estimated Gamma.....	87
Figure 6. Proportion Pruned vs. Estimated Gamma.....	88
Figure 7. Molecular Evolution vs. Estimated Gamma.....	89

## CHAPTER 1

# RED QUEENS, COURT JESTERS, AND SNAKE BIODIVERSITY: TESTS OF OPHIDIAN MACROEVOLUTION

Hugo Alamillo<sup>1\*</sup>, Chad D. Brock<sup>1</sup>, Luke J. Harmon<sup>2</sup>, Michael E. Alfaro<sup>3</sup>

<sup>1</sup> School of Biological Sciences  
Washington State University  
Pullman, WA 99164

<sup>2</sup> Department of Biological Sciences  
Campus Box 443051  
University of Idaho  
Moscow, ID 83844

<sup>3</sup> Department of Ecology and Evolutionary Biology  
University of California, Los Angeles  
621 Charles E. Young Drive South  
CA 90095

\* Author for correspondence ([halamillo@wsu.edu](mailto:halamillo@wsu.edu))

Running title: Snake lineage diversification

**Keywords:** colubroid; diversification; Neogene; snakes; macroevolution;

## **Abstract**

There are two major and largely untested hypotheses for the extraordinary species diversity exhibited by snakes. First, large gape-size might have acted as a key ecological innovation which enabled rapid speciation as snakes exploited a range of new trophic niches. Second, grassland habitat expansion during the Miocene might have triggered rapid diversification during the so-called “age of snakes”. Here we use two newly developed methods (MEDUSA and Geologic Time Interval Test) along with a time-calibrated phylogeny to test these hypotheses. We find no evidence for a shift in diversification rate following the evolution of large gape size in macrostomatans. In contrast, snake diversification over the Neogene was significantly higher than over any other geologic time interval in snake evolutionary history. This elevated diversification in the Neogene appears to be driven by the evolutionary radiation of the colubroids, which exhibit diversification rates up to 2.5 times greater than the background rate of diversification in other related snake lineages. Our findings implicate processes acting during the last 23 MY, such as the expansion of the grasslands or the diversification of rodents, in producing the extraordinary diversity of modern snakes.

## 1. INTRODUCTION

Historically, two kinds of explanations have been offered to describe patterns of species diversification across clades and through time. The Red Queen (RQ) hypothesis emphasizes the importance of biotic interactions that underlie shifts in rates of diversification (Van Valen 1973). The name refers to the Red Queen in Lewis Carroll's *Through the Looking-Glass* who tells Alice "Now, here, you see, it takes all the running you can do, to keep in the same place" (Carroll 1872). RQ serves as an umbrella concept encompassing various types of micro- and macroevolutionary hypotheses. In a microevolutionary perspective RQ has been used to explain frequency-dependent selection and the maintenance of genetic variation (Lively 1996). However, in the macroevolutionary world it is typically used to explain differential richness patterns among clades. As in, the evolution of key characters that can provide a clade with increased ecological opportunity, leading to greater species richness (Simpson, 1953; Ree & Smith 2008; Schlüter 2000). The Court Jester (CJ) explanation focuses instead on abiotic events that affect multiple clades across the same time interval (Barnosky 2001). Models similar in spirit to the CJ include the *Traffic Light Model* (Vrba 1995), *Relay Model* (Vrba 1995), and *Tiers of Time Model* (Gould 1985), among others, where large-scale climatic shifts leave a signature across many groups within a given time interval (Benton 2009).

Both RQ and CJ explanations have been invoked to explain patterns of species richness in snakes. With nearly 3,000 species occupying every continent except Antarctica (Greene 1997; Uetz & Etzold 2007) and tremendous ecological diversity, snakes represent one of the most conspicuous radiations of reptiles. Red Queen-type explanations link differential patterns of species richness within snakes to clade-specific key traits (Greene 1997; Vidal 2002). The most striking of these, large gape size, has been hypothesized as the primary biotic factor driving the

diversification of one particular clade of snakes, the species-rich alethinophidians. Containing nearly 90% of the total species richness of snakes, alethinophidians are characterized by exceptionally large gapes (Zug et al. 2001). Innovation in cranial function, including the evolution of flexible articulation of the maxillae, nasal, and palatine bones, is hypothesized to have enhanced gape size in the lineage leading to the ancestral alethinophidian (Gans, 1961), increasing the range of available prey (Green, 1983; Arnold, 1993; Rodriguez-Robles et al. 1999; Cundall & Greene 2000; Vincent et al., 2006). However the hypothesis that such a key innovation enabled the subsequent diversification of alethinophidians (Gans 1961), has not been statistically tested.

In contrast to the large-gape hypothesis, paleontologists have attributed snake species richness to large-scale ecological changes. One of the most well-known of the CJ-type explanations is that climatic changes during the Miocene triggered a cascade of diversification that included grasslands, rodents, and finally snakes (Steppan et al. 2004; Rage 1987; Stanley 1979; Greene 1997). Though this scenario includes both abiotic and biotic interactions, we classify it as a CJ hypothesis because the primary cause of diversification is the abiotic shifting of climate. The sudden appearance of diverse fossil species during this time has led some paleontologists to dub the Neogene as the ‘Age of Snakes’ (Stanley 1979), though no formal tests of this pattern have even been performed.

Here we adopt a comparative phylogenetic approach to test macroevolutionary hypotheses of snake diversification. We apply a recently developed method to test for elevated diversification rates at the onset of large-gape size (RQ-scenario) using combined phylogenetic and taxonomic data. We also apply a newly developed comparative method for elevated

diversification rates during a specified time period (CJ-scenario) to test, for the first time, whether the Neogene was the ‘Age of Snakes’.

## 2. MATERIAL AND METHODS

### (a) *Divergence-time estimation*

We constructed a molecular phylogeny by downloading DNA sequences for 232 species from GenBank for two genes, cytochrome-b (cyt-b) and Oocyte maturation factor (c-mos). We used Clustal to align the sequences from both loci. Our total sequence alignment included 892 bp (cyt-b: 1–444; c-mos: 445–892 ). To infer both divergence times and the phylogeny we used the uncorrelated lognormal relaxed-clock model implemented in BEAST v1.4.6 (Drummond et al. 2006). We used *Shinisaurus crocodilurus* as the outgroup for our phylogenetic analysis (Townsend et al. 2004). We used ten probabilistic calibration prior distributions taken from described fossils (Burbrink & Lawson 2006; Scanlon et al. 2003; Noonan & Chippindale 2006; Table 1) to incorporate calibration uncertainty and avoid problems with truncating the posterior distribution (Yang & Rannala 2006). We specified a Yule prior on rates of cladogenesis. The model of evolution for this analysis was set to HKY+I+Γ (ModelTest 3.7) and run for 30 million generations. We analyzed the output using visual inspection of runs and calculated effective sample sizes for all parameters using TRACER 1.3.

### (b) *Patterns of diversification*

All the diversification analyses were done using R Statistical Package v. 2.7.1 with the following libraries: Ape (Paradis et al. 2004), Geiger (Harmon et al. 2008), and Laser (Rabosky 2006).

taxon	calibration fossil	min	max	source
Colubridae	<i>Lampropeltis</i>	15	19	[1]
Colubridae	<i>Pantherophis</i>	16	20	“
Colubridae	<i>Zamenis</i>	6	20	“
Colubridae	<i>Coluber, Masticophis</i>	6	15	“
Colubridae	<i>Col., Mas., Salvadora</i>	20	24	“
Cylindrophiid, Acrochordid. Uropeltid, split	<i>Nigerophis</i>	61	130	“
Aniliid, Tropidophid split	<i>Coniophis cosgriffi</i>	70.6	130	“
Elapidae	<i>Laticauda</i>	16.4	23.7	[2]
Boids	indet	65	130	[3]
Crown Group Snakes	<i>Lapparentophis defrennei</i>	130	180	[4]

Table 1. Fossil calibrations used in this study. Calibrations used in the preliminary study (Min/Max in millions of years). Source for calibrations: [1] Burbrink & Lawson 2006; [2] Scanlon, Lee & Archer, 2003; [3] Noonan & Chippindale 2006; [4] Rage & Richter 1994.

Snake groups with high rates of diversification, were identified with MEDUSA (Modeling Evolutionary Diversification Using Stepwise AIC, Alfaro et al. 2009a; Sup. Mat.). This approach allows the recognition of exceptional clades using an incompletely sampled phylogeny and incorporates species richness for designated taxonomic groups. We assigned total species richness of snake groups to representative lineages on the time tree. To do this we collapsed our snake tree to a set of monophyletic clades (Table 2), and added taxonomic richness for each clade from the TIGR reptile database (Uetz and Etzold 2007).

We then ran MEDUSA to calculate the background tempo of snake diversification ( $r$ ), the signature of extinction ( $\epsilon$ ), and to ask which lineages stand out from this background (see Alfaro et al. 2009a). The Akaike Information Criterion with the standard second order correction for small sample sizes with a significance cutoff of 4 units of difference, was used to differentiate between alternative models (i.e. AICc, McQuarrie & Tsai 1998; Table 4). If the evolution of large gape was responsible for the species richness of macrostomatans, we predicted that we would detect a substantial rate increase along the branch leading to their most recent common ancestor.

### **(c) Geologic Time Interval Test**

We developed a new comparative method to test the hypothesis that the Neogene was a time period with elevated rates of diversification for the colubroids (Stanley 1979). The Geological Time Interval (GTI) Test, relies on Kendall-Moran (KM) estimates of diversification rates (e.g. Baldwin & Sanderson 1998; Weir 2006) over specific time periods of our snake time tree (see above). The typical KM estimator is  $\hat{S} = (N-2)/B$ , where  $\hat{S}$  is an estimate of net diversification rate,  $B$  is the sum of branches that have descended from the most recent common ancestor of the

clade Name	species	representative species
Anomalepididae	16	<i>Liotyphlops albirostris</i>
Typhlopidae	233	<i>Ramphotyphlops braminus</i>
Leptotyphlopidae	93	<i>Leptotyphlops columbi</i>
Aniliidae	1	<i>Anilius scytale</i>
Tropidophiinae	23	<i>Tropidophis haetianus</i>
Acrochordidae	3	<i>Acrochordus granulatus</i>
<i>Calabaria</i>	1	<i>Calabaria reinhardtii</i>
<i>Charina</i>	2	<i>Charina bottae</i>
<i>Eryx</i>	11	<i>Eryx colubrinus</i>
<i>Epicrates</i>	11	<i>Epicrates striatus</i>
<i>Eunectes</i>	4	<i>Eunectes marinus</i>
<i>Sanzinia</i>	1	<i>Sanzinia madagascariensis</i>
<i>Acrantophis</i>	2	<i>Acrantophis madagascariensis</i>
<i>Candoia</i>	3	<i>Candoia carinata</i>
Pareatidae	15	<i>Pareas macularius</i>
Xenodermatinae	17	<i>Oxyrhabdium leporinum</i>
Calamariinae	90	<i>Calamaria pavimentata</i>
Homolopsinae	34	<i>Cerberus rynchops</i>
Boodontinae	74	<i>Mehelya unicolor</i>
Boodontinae-incertae_sedis	6	<i>Duberria lutrix</i>
Pseudoxyrhophiinae	75	<i>Pseudoxyrhopus ambreensis</i>
Colubrinae	571	<i>Masticophis flagellum</i>
Psammophiinae	39	<i>Psammophis lineolatus</i>
Natricinae	193	<i>Afronatrix anoscopus</i>
Natricinae-incertae_sedis	18	<i>Psammodynastes pulverulentus</i>
Dipsadinae	197	<i>Hypsiglena torquata</i>
Xenodontinae	250	<i>Heterodon simus</i>
Crotalinae	154	<i>Agkistrodon piscivorus</i>
Viperinae	85	<i>Bitis nasicornis</i>
Azemiopinae	1	<i>Azemiops feae</i>
Atractaspinae	21	<i>Atractaspis bibronii</i>
Aparallactinae	48	<i>Aparallactus werneri</i>
Elapinae	141	<i>Micrurus fulvius</i>
Hydrophiinae	177	<i>Notechis ater</i>
Laticaudinae	8	<i>Laticauda colubrina</i>
Xenopeltidae	2	<i>Xenopeltis unicolor</i>
Bolyeridae	2	<i>Casarea dussumieri</i>
Pythoninae	36	<i>Python molurus</i>

Table 2. Taxonomic Richness. Species richness assigned to each clades in this study. All clades were assigned species numbers according to TIGR Reptile Database.

species in the clade and  $N$  represents the number of extant lineages (Kendall 1949; Moran 1951). The KM estimator uses  $N-2$  because we used crown ages and therefore there are two taxa at the basal split. However, over a time-slice,  $N$  is based on the species present at the beginning of the time period. To implement this test, we first calculated the background diversification rate of all snakes (Magallón & Sanderson 2001),  $r_G$  ( $r_G = [\ln(n) - \ln(2)]/\tau$ ), where  $\tau$  is the total clade time (Fig. 1). We then calculated the Neogene-specific rate,  $r_i$ , as  $r_i = [N_{\text{end}} - N_{\text{start}}]/B_i$  where  $B_i$  is the sum of branch lengths for that interval. To test whether the interval-specific rate was significantly different from the background rate we constructed a null distribution for the expected rate over the interval given the overall snake rate. To do this, we simulated trees under the global snake rate until they reached 3000 species, the estimated living diversity of snakes (TIGR Reptile Database: <http://www.reptile-database.org/>). We then randomly pruned these trees to 281 species to match our sampled diversity. Given that the simulated trees did not all have equal tree-lengths, we rescaled the branch lengths to find the time segment of interest in all simulated trees, and thereby calculate interval-specific diversification rate,  $r^{\circ}_i$  for each replicate. If diversification over the Neogene was exceptionally rapid, we expected that  $r_i$  would fall in the upper tail of the distribution of  $r^{\circ}_i$ s.

### 3. RESULTS

#### (a) Divergence time estimation

We inferred phylogenetic relationships among the major groups of snake that were largely concordant with recent studies of snake relationships (Lawson et al. 2005; Lee et al. 2007) (Fig. 2). Effective Sample Sizes (ESS) were greater than 200 for all parameter traces and all parameter sample distributions were unimodal which is indicative of convergence. We estimated the age of

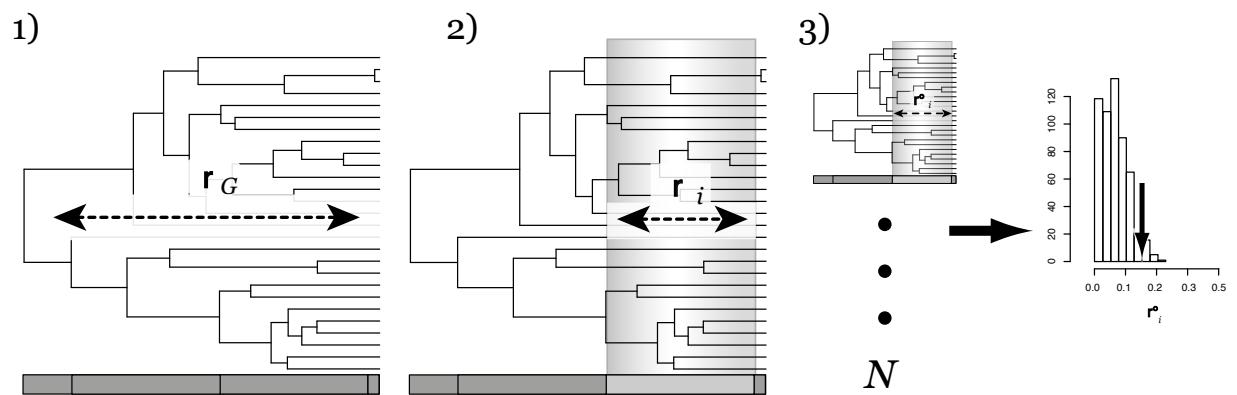


Figure 1. GTI test. (a) Calculate overall rate,  $r_G$ . (b) Calculate interval rate,  $r_i$ . (c) Simulate under  $r_G$  and calculate  $r_i^*$  for each replicate. If  $r_i$  falls in tail of  $r_i^*$  reject  $r_G$  for that interval.

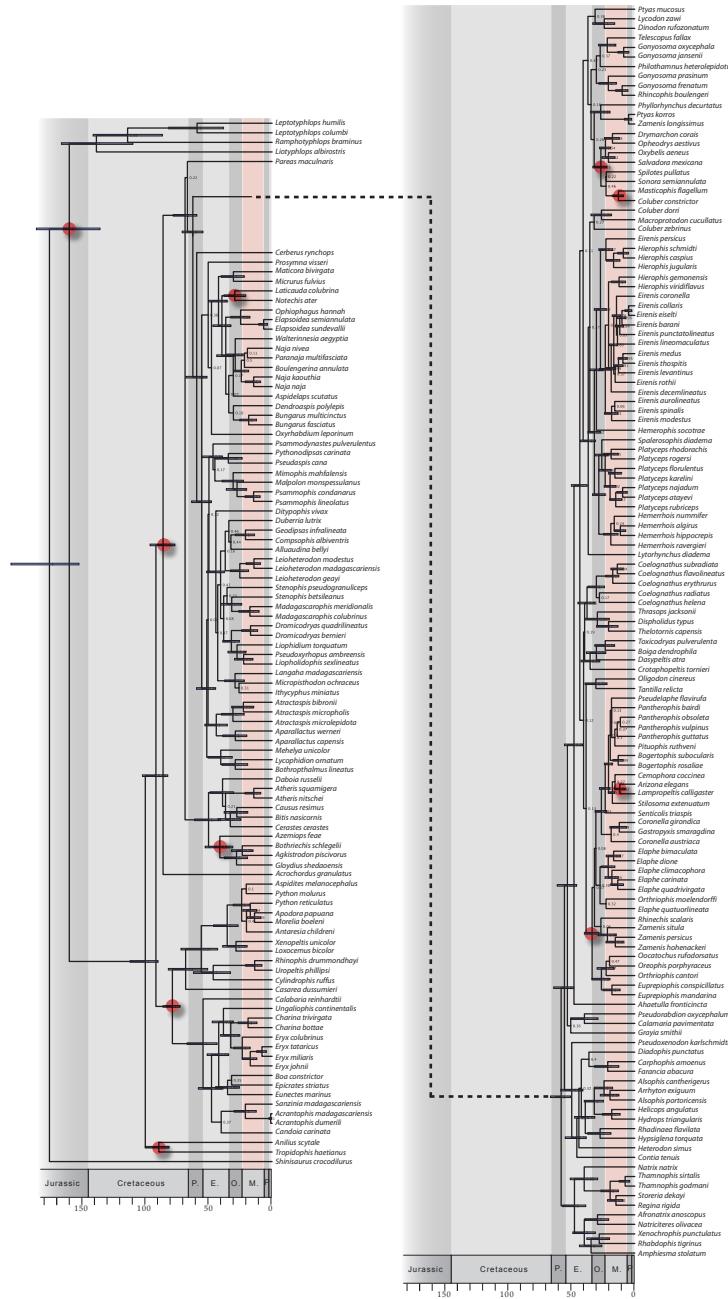


Figure 2. Chronogram of the major groups of snakes. Molecular time-scale for the snake taxa sampled in this study. Pink region represents the specific time interval we tested for higher than expected rates of diversification, the Neogene. Time-scale in millions of years from present time. Red circles on nodes represent calibration points.

the most recent common ancestor (MRCA) of all snakes to have lived in the Late Jurassic (156 mya, 95% high posterior density, HPD: 132–181 mya), concordant with available fossil data (Rage 1994). We also estimated ages for major clades within the tree that are generally older than documented fossils. For example, we calculated the age of the MRCA of Colubridae to be from the Paleocene (57 mya, [49, 65 95% HPD]), while the oldest documented fossil from this group is from the Oligocene (33.9–23.03 mya; see Table 3 for crown group ages recovered in this study).

### **(b) *RQ* diversification**

Under a constant rates model, our data strongly reject a pure-birth model for snake evolution in favor of a birth-death model with high turnover rates (pure birth,  $r_{\text{snakes}} = 0.09$ ,  $\ln L = -347.28$ ; birth-death,  $r = 0.04$ ,  $\varepsilon = 0.92$ ,  $\ln L = -333.29$ ; Likelihood-ratio test,  $\chi^2_1 = 27.97$ ,  $p < 0.0005$ ). Our MEDUSA analysis supports the flexible rate model where both speciation and extinction rates change along one branch in the tree, the branch leading to most recent common ancestor of the colubroids ( $\Delta \text{AICc} = 47.81$ ,  $r_{\text{non-colubroids}} = 0.03$ ,  $r_{\text{colubroids}} = 0.1$ ).

### **(c) “Age of snakes” hypothesis**

Our simulations of expected diversity revealed that colubroid richness was significantly higher than expected over the Neogene. The GTI test indicates that colubroids have radiated faster than expected in the Neogene ( $p = 0.038$  from 1000 simulations) given the background rate of diversification estimated.

## **4. DISCUSSION**

group	mean age (my)	95% HPD
Serpentes	156.05	132.4, 181.3
Scolecophdia	135.1	107.1, 162.08
Anilioiida	87.2	79.3, 97.6
Booidea	76.7	70.8, 83.9
Alethinophidia	83.9	74.8, 93.9
Colubroidea	67.0	57.9, 76.1
Viperidae	48.9	39.7, 59.6
Elapidae	41.3	34.7, 48.8
Colubridae	56.7	49.2, 64.8
Pseudoxyrophinae	43.3	36.7, 50.3
Natricinae	46.5	37.8, 55.5

Table 3. Crown group ages. Age estimates for various groups in the study. MY = million years.

### *Red Queens and snake diversification.*

We found little evidence to support the hypothesis that large gape size acted as a key innovation to drive alethinophidian diversification (Fig. 3). Instead we found evidence for a change in diversification rate approximately 25 MY later in the ancestor to colubroids (elapids, viperids, and colubrids). This clade, which contains >2500 species today, expanded at ~ 2.5 times the net background snake diversification rate (Table 4). Traditionally, key innovation hypotheses have not been invoked to explain species richness within this group. However, our results are consistent with the hypothesis that venom evolved in the ancestral colubroid instead of in viperids, and that this trait contributed to the evolutionary success of the group (Vidal 2002).

Our scan for diversification rate shifts across the tree also fails to support many previously proposed hypotheses for differential diversification within snakes. These include mimicry of noxious invertebrates (e.g. millipedes) by early alethinophidians (*Anomochilus*, *Cylindrophis*, *Anilius*, etc., Greene 1997) and live birth in vipers (Lynch 2009). Despite a nine-fold difference in species richness between blindsnakes and all other snakes, our analysis does not place a rate shift on the basal ophidian split. And, somewhat surprisingly, we find only equivocal support for a rapid radiation of hydrophiine snakes ( $\Delta\text{AICc} = 2.48$  Shift 2, Fig. 3), despite recent studies which have suggested that they explosively diversified (Kelly et al. 2009; Lukoschek 2006).

These findings are consistent with a recent study of diversification in other vertebrates which similarly found a lack of a tight relationship between the inferred evolution of a key trait and a rate shift in the clade (Alfaro et al. 2009 a,b). The lack of correlation suggests that other,

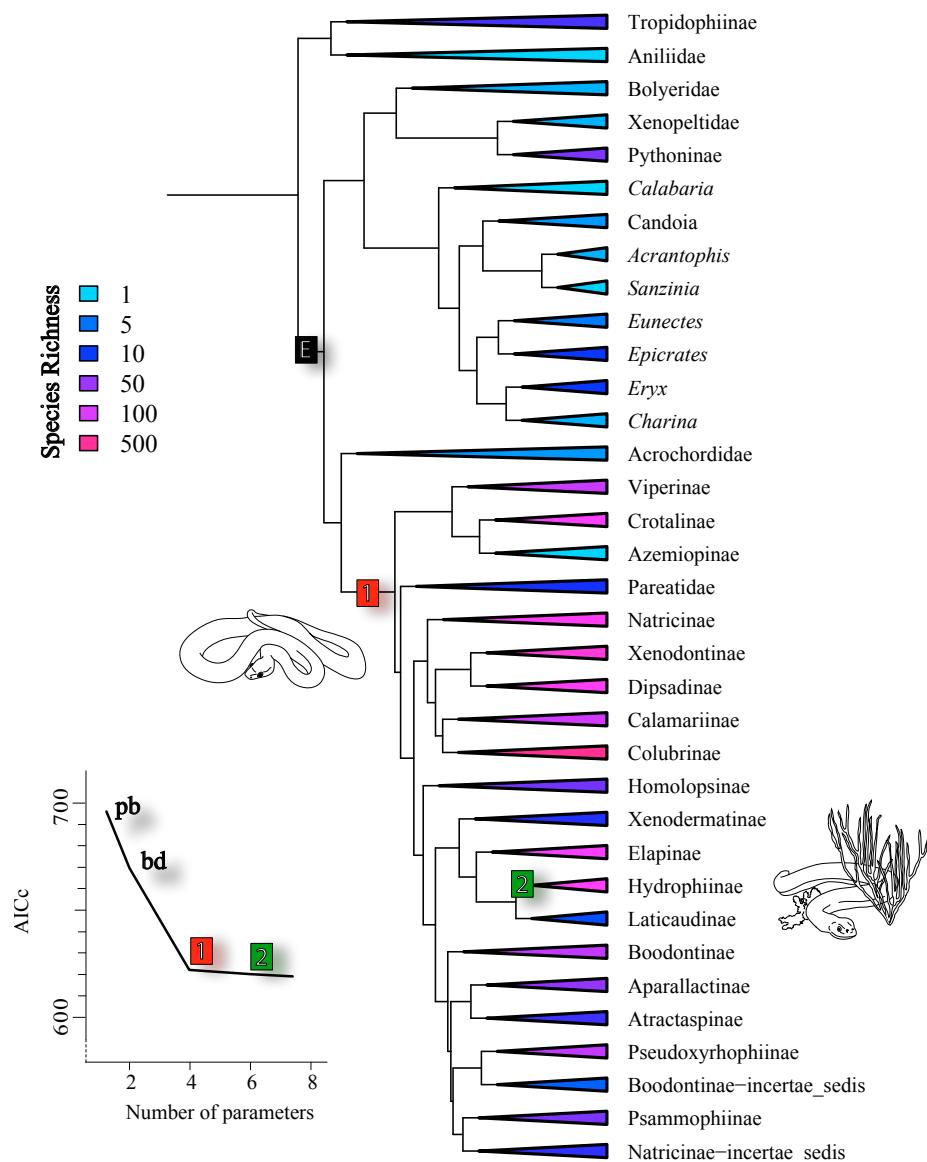


Figure 3. Plot of extant diversity of snake groups sampled in this study as a function of stem age. Pruned snake chronogram showing diversity of the groups tested in this study. Cooler colors represent clades with lower species numbers, while warmer colors depict clades with more species. Numbered clades represent shifts of diversification, with red boxes representing significant increases and green boxes representing nearly significant shifts. Insert graph depicts the improvement of AICc score as MEDUSA traverses the tree fitting new models of b and d. Black box with "E" represents lineage where "Large gape size" would have been gained.

<b># of shifts</b>	<b>clade</b>	<b>r</b>	<b><math>\varepsilon</math></b>	<b>AICc</b>	<b><math>\Delta\text{AICc}</math></b>
0 (whole-tree birth-death model)	whole tree	0.041	0.917	670.75	--
1	Colubroids	0.01	2.50E-06	622.94	47.81
2	Hydrophiines	0.148	0.603	620.47	2.47
--	background	0.033	0.606	--	--

Table 4. The tempo of snake diversification. Clade number refers to rate shifts identified in Fig. 3. r is the net diversification rate,  $\varepsilon$  is the extinction fraction (d/b). AICc and  $\Delta\text{AICc}$  show improvement of AICc score over a constant rates birth-death model as clades are allowed to change rates. Background shows background rates of other snake clades under the two-rate model.

likely clade-specific factors underlie macrostomatan diversification. Furthermore, the general lack of relationship between vertebrate key traits and diversification rate shifts raises doubts about the utility of the idea of key innovations as a general explanation of species richness. It is possible that key traits act in subtle and context dependent ways such that there is a gradual preamble to a radiation (Simpson 1953; Donoghue 2005). However, if this is the general mode of action for key traits, they will be very difficult to test with comparative methods alone unless there are multiple originations of the trait itself (Maddison et al. 2007).

#### *Miocene Court Jesters.*

Our study provides the first quantitative support for the hypothesis that Neogene events fueled snake diversification. Though the specific effects on fauna and flora of global climate remodeling during the Neogene are widespread, the evidence of a transition from a greenhouse Cenozoic world to the icehouse of the Pleistocene is well documented (Zachos et al. 2001). The elevation of several mountain ranges (Himalayas and Tibetan Plateau [Raymo et al. 1998], the Andes [Gregory-Wodzicki 2000], and the California Sierras [Mulch et al. 2009]) is one possible mechanism that drove diversification. Silicate weathering of the new peaks during the Miocene decreased atmospheric CO<sub>2</sub> levels and increased light regimes (François et al. 2006), while monsoonal changes that modified eco-regions across continents (Fluteau et al. 1999) also had an impact as a result of cordilleran uprise. Specifically, reduction of desert areas and expansion of tropical seasonal forests were key results of reshuffling of vegetation distributions and biosphere carbon stock during this period.

Biotic reorganization because of climate change might have led to concerted evolution or offset diversification of groups during the Neogene. Current models of C4 grassland expansions suggest that fire regimes, driven by Late Miocene monsoonal changes, ignited grass diversification (Keeley & Rundel 2005). A scenario of woodland or C3 plants replacement by grassland in a short period of time, at a global scale, but with different rates at each region (Latorre et al. 1997), might have facilitated ecological specialization and led to the morphological and phenotypic diversity we see within the snakes. This would have been enabled by the Miocene diversification of one of the main prey-types of colubroids, rodents, which diversified 21–27 mya (Hugall 2007; Pereira & Baker 2006; Penny et al. 1999; Steppan et al. 2004). This hypothesis of codiversification between grasses, rodents, and snakes could be tested by comparing diversification rates from phylogenies for these three groups (e.g. Bokma 2003).

Testing hypotheses about the underlying causes of major species radiations is central to understanding the patterns of ecological and phenotypic diversity on Earth (Simpson 1953; Schulter 2000). Our analysis implicates both CJ and RQ processes in shaping species richness patterns in ophidians, consistent with multilevel models of clade diversification (Benton 2009). Future work examining both the influence of envenomation and the influence of Miocene processes on diversity patterns within snake subclades is needed to validate the generality of these macroevolutionary explanations. Taxon-specific models are sorely needed to gain a better understanding of fine-scale processes that lead to rapid lineage and morphological diversification (Gavrillets & Losos 2009) across snakes.

#### **4. Acknowledgements**

We would like to thank Barb Banbury, Alex Dornburg, and Dr. Ken Kardong for discussions and detailed comments about this study. Drs. Matthew Settles and Matthew King helped immensely with computer cluster technical support. This work was supported by NSF grants awarded to M. E. Alfaro, DEB 0918748 and IOS 0819009

## REFERENCES

- Alfaro, M. E., Santini, F., Brock, C. D., Alamillo, H., Dornburg, A., Carnevale, G., Rabosky, D. & Harmon, L. J. 2009a Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *P. Natl. A. Sci.* 106, 13410–13414.
- Alfaro M. E., Brock, C., Banbury, B. L. & Wainwright, P. 2009b Does evolutionary innovation in pharyngeal jaws lead to adaptive radiation? Evidence from analysis of diversification in labrids and parrotfishes. *BMC Evol. Biol.* 9, 1–14.
- Arnold, S. J. 1993 Foraging theory and prey-size-predator-size relations in snakes, p. 87–115. In *Snakes, Ecology and evolutionary biology* (ed. R. A. Seigel, J. T. Collins & S. S. Novak), McGraw-Hill Publishing.
- Baldwin, B. G. & Sanderson M. J. 1998 Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *P. Natl. A. Sci.* 95, 9402–9406.
- Barnosky, A. D. 2001 Distinguishing the effects of the Red Queen and Court Jester on Miocene Mammal evolution in the northern Rocky Mountains. *J. Vert. Paleont.* 21, 172–185.
- Benton, M. J. 2009 The red queen and the court jester: Species diversity and the role of biotic and abiotic factors through time. *Science* 323, 728–732
- Bokma, F. 2003 Testing for equal rates of cladogenesis in diverse taxa. *Evolution*. 57, 2469–2474.
- Burbrink, F. T. & Lawson R. 2006 How and when did Old World Ratsnakes disperse into the New World? *Mol. Phyl. & Evol.* 43, 173–189.
- Carroll, L. 1872 Through the looking-glass, and what Alice found there. Macmillan & Co., London.
- Cundall, D. & Greene, H. W. 2000 Feeding in snakes. In *Feeding: Form, Function, and Evolution in Tetrapod Vertebrates* (ed. K. Schwenk), pp. 293–333. San Diego: Academic Press.

- Donoghue, M. J. 2005 Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. *Paleobiology* 31, 77–93.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4, 0699–0700.
- Fluteau, F., Ramstein, G. & Besse, J., 1999 Simulating the evolution of the Asian and African monsoons during the past 30Myr using an atmospheric general circulation model. *J. Geophys. Res.* 104, 11995–12018.
- François, L., Ghislain, M., Otto, D. & Micheels, A. 2006 Late Miocene vegetation reconstruction with the CARAIB model. *Palaeogeogr. Palaeocl.* 238, 302–350
- Gans, C. 1961 The feeding mechanism of snakes and its possible evolution. *Am. Zool.* 1, 217–227.
- Gavrillets, S. & Losos, J. B. 2009 Adaptive radiation: contrasting theory with data. *Science*. 323, 732–737
- Gould, S. J. 1985 The paradox of the 1st tier: an agenda for paleobiology. *Paleobiology* 11, 2–12.
- Greene, H. W. 1983 Dietary correlates of the origin and radiation of snakes. *Am. Zool.* 23, 431–441.
- Greene, H. W. 1997 Snakes, the evolution of mystery in nature. University of California Press.
- Gregory-Wodzicki, K. M. 2000 Uplift history of the central and northern Andes: A review. *Geol. Soc. Am. Bull.* 112, 1091–1105.
- Harmon, L. J., Weir, J., Brock, C., Glor, R. E. & Challenger, W. 2008 GEIGER: Investigating evolutionary radiations. *Bioinformatics*. 24, 129–131.

Hugall, A. F., Foster, R. & Lee, M. S. 2007 Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst Biol* 56, 543–563.

Keeley J. E. & Rundel, P. W. 2005 Fire and the Miocene expansion of the C4 grasslands. *Ecol. Lett.* 8, 636–690.

Kelly, C. M. R., Barker, N. P., Villet, M. H. & Broadly, D. G.. 2009 Phylogeny, biogeography, and classification of the snake superfamily Elapoidea, a rapid radiation in the late Eocene. *Cladistics* 25, 38–63.

Kendall, D. G. 1949 Stochastic processes and population growth. *J. R. Statist. Soc. B.* 11, 230–264.

Latorre, C., Quade, J. & McIntosh, W. C. 1997 The expansion of C4 grasses and global change in the late Miocene: stable isotope evidence from the Americas. *Earth Planet Sci. Lett.*, 146, 83–96.

Lawson, R., Slowinski, J. B., Crother, B. I. & Burbrink, F. T. 2005 Phylogeny of the Colubroidea (Serpentes), New evidence from mitochondrial and nuclear genes. *Mol. Phyl. & Evol.* 37, 581–601.

Lee, M. S. Y., Hugall, A. F., Lawson, R. & Scanlon, J. D. 2007 Phylogeny of snakes (Serpentes), combining morphological and molecular data in likelihood, Bayesian and parsimony analyses. *Syst. & Biodivers.* 5, 371–389.

Lively, C. M. 1996 Host-Parasite coevolution and sex. *BioScience* 46, 107–114.

Lukoschek, V. & Keogh, J. S. 2006 Molecular phylogeny of sea snakes reveals a rapidly diverged adaptive radiation. *Biol. J. Linn. Soc.* 89, 523–539.

- Lynch, V. 2009 Live-birth in vipers (Viperidae) is a key innovation and adaptation to global cooling during the Cenozoic. *Evolution* 63, 2457–2465
- Maddison, W. P., Midford, P. E. & Otto, S. P. 2007 Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56, 701–710.
- Magallón, S. & Sanderson, M. J. 2001 Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780.
- McQuarrie, A. D. R. & Tsai, C.-L. 1998. Regression and Time Series Model Selection. Singapore: World Scientific.
- Moran, P. A. P. 1951 Estimation methods for evolutive processes. *J. R. Statist. Soc. B.* 51, 141–146.
- Mulch, A., Sarna-Wojcicki, A. M., Perkins, M. E. & Chamberlain, C. P. 2008 A Miocene to Pleistocene climate and elevation record of the Sierra Nevada (California). *P. Natl. A. Sci.* 105, 6819–6824.
- Noonan, B. P. & Chippindale P. T. 2006 Dispersal and vicariance, The complex evolutionary history of boid snakes. *Mol. Phyl. & Evol.* 40, 347–358.
- Paradis, E., Claude J. & Strimmer, K. 2004 APE, Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289–90.
- Penny, D., Hasegawa, M., Waddell, P. J. & Hendy, M. D. 1999 Mammalian Evolution, Timing and Implications from Using the LogDeterminant Transform for Proteins of Differing Amino Acid Composition. *Syst. Biol.* 48, 76–93.
- Pereira, S. & Baker, A. 2006 A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol. Biol. Evol.* 23, 1731–1740.

- Rabosky, D. L. 2006 LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evol. Bioinform. Online* 2, 257–260.
- Rage, J-C. 1984 Handbuch der Paläoherpetologie, Part 11. Serpentes. New York: Gustav Fischer Verlag.,
- Rage, J-C. 1987 Fossil Record, p. 51–76. In Snakes, Ecology and evolutionary biology (ed. R. A. Seigel, J. T. Collins & S. S. Novak), McGraw-Hill Publishing.
- Rage, J-C & Richter, A. 1994 A snake from the Lower Cretaceous (Barremian) of Spain, The Oldest known snake. *Jarbuch für Geologie und Paläontologie, Monatshefte*, Stuttgart, fasc H.9, 561–565.
- Ree, R. H. & Smith, S. A. 2008 Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14. (doi, 10.1080/10635150701883881.)
- Rodriguez-Robles, J. A., Bell, C. J. & Greene, H. W. 1999 Gape size and evolution of diet in snakes: feeding ecology of erycine boas. *J. Zool. Lond.* 248, 49–58.
- Scanlon, J. D., Lee M. S. Y. & Archer, M. 2003 Mid-Tertiary elapid snakes (Squamata, Colubroidea) from Riversleigh, northern Australia, early steps in a continent-wide adaptive radiation. *Geobios* 36, 573–601.
- Schluter, D. 2000 The ecology of adaptive radiation. Oxford: Oxford University Press.
- Simpson, G. G. 1953 The major features of evolution. New York: Columbia Univ. Press.
- Stanley, S. M. 1979 Macroevolution, patterns and process. San Francisco: W. H. Freeman Press.
- Steppan, S. J., Adkins, R. M. & Anderson J. 2004 Phylogeny and divergence-date estimation of rapid radiations in Murid rodents based on multiple nuclear genes. *Syst. Biol.* 53, 533–544.

- Townsend, T. M., Larson, A., Louis, E. & Macey, J. R. 2004 Molecular phylogenetics of Squamata, The position of Snakes, Amphisibaenians, and Dibamids, and the root of the squamate tree. *Syst. Biol.* 53, 735–757.
- Uetz, P. & Etzold, T. 2007 The Reptile Database, <http://www.reptile-database.org>.
- Van Valen, L. M. 1973 A new evolutionary law. *Evol. Theory* 1, 1–30.
- Vincent, S. E., Dang, P. D., Herrel, A. & Kley, N. J. 2006 Morphological integration and adaptation in the snake feeding system, a comparative phylogenetic study. *J. Evol. Biol.* 19, 1545–1554.
- Vrba, E. S. 1995 On the connection between paleoclimate and evolution, p. 24–45. In *Paleoclimate and evolution, with emphasis on human origins.* (eds. E. S. Vrba, G. H. Denton, T. C. Partridge, & L. H. Burckle), New Haven: Yale University Press.
- Weir, J. T. 2006 Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution.* 60, 842–855.
- Yang, Z. & Rannala, B. 2006 Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* 23, 212–226.
- Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. 2001 Trends, rhythms, and aberrations in global climate 65Ma to present. *Science* 292, 686–693.
- Zug, G. R., Laurie, J. V. & Caldwell, P. C. 2001 *Herpetology: An introductory biology of amphibians and reptiles*, 2nd ed. Academic Press, San Diego.

## Supplementary Materials

**Use of MEDUSA.** We fitted simple models with MEDUSA where birth and death rates were allowed to vary across the tree. The first model fitted was a Pure Birth model (PB) with one parameter, b, that represents the per-lineage rate of speciation. A second model, the Birth Death model (BD) had two parameters, b and d, which represent per-lineage rates of speciation and extinction. By using equations from Rabosky (2006) we fitted these simple models to the combined phylogenetic and taxonomic data set. Given the particular values of b and d, we found the likelihood of obtaining a phylogeny with the indicated taxonomic data (e.g. group age and richness). To assess the fit of all models, their individual AIC scores were calculated ( $AIC = 2k - 2\ln L$ , where k is the number of parameter that describe the model). Our stepwise AIC approach first found the ML value of b and d and used those ML values as the starting point for the AIC algorithm. The algorithm then fit alternate models that each have increasing complexity and stopped when AIC improvement was less than 4, which was the threshold for significance.

To illustrate this approach we explain here an iteration of the algorithm. The algorithm first broke the chronogram into two sections, by splitting a branch that has a high likelihood of seeing a rate shift. This model now had a clade descended from the break branch with its own b and d rates (b<sub>1</sub> and d<sub>1</sub>), and the remaining part of the tree had different rates (b<sub>2</sub> and d<sub>2</sub>). If in previous steps a subclade chosen already included other clades with their own b and d values, those already chosen clades kept their assigned rate values. The algorithm then found the ML estimate for all b and d parameters in each branch of the tree. For selection of a branch that minimized this score, it calculated the AIC score. This new tree now represented a new candidate for the next model to be kept. If the AIC score was 4 or more units lower in this new model, then

it was retained. The procedures stopped when the AIC score for the brake point with the best score was not less than four units smaller than the previous model.

**Table S1| The shifts in snake diversification.** Break refers to rate shifts identified by MEDUSA.

# breaks	clade	InL	parameters	AIC	AICc
0	Whole Tree	-333.29	2	670.58	670.75
1	Colubroidea	-306.03	5	622.07	622.94
2	Hydrophiinae	-301.14	8	618.28	620.46
3	Pythonidae	-296.52	11	615.05	619.24
4	Azemiopinae	-292.42	14	612.84	619.84

**List S1| GenBank sequence numbers used in this study.**

CytB:

AF039260; AF217812; AF217827-8; AF217832; AF217834-8; AF217840-2; AF241385;  
 AF241405; AF283599; AF337090; AF337107; AF337111; AF337113; AF337157; AF420073;  
 AF420135; AF420193; AF471028-48; AF471050-59; AF471061; AF471063; AF471064-5;  
 AF471067; AF471071-86; AF471088-90; AF471092-3; AF544669; AF544671-3; AJ275684;  
 AJ749344; AJ749349; AY058965; AY058967-8; AY099983; AY099985-6; AY099988;  
 AY099990-1; AY099994-5; AY121369; AY188005-6; AY188008-14; AY188016-7;  
 AY188019-27; AY188030; AY188032-8; AY188040-3; AY223555; AY223557; AY223566;  
 AY223590; AY352747; AY376739; AY376741-2; AY376744; AY376747-9; AY376751;  
 AY376755; AY376757-62; AY376764; AY376766-7; AY376766-7; AY486911-2; AY486915-23;  
 AY486925-7; AY486929-32; AY486934; AY586243; AY612006; AY713376; DQ112075;  
 DQ112077; DQ538341; DQ902102-10; DQ902112-21; DQ902123; DQ902125-8; DQ902133;

DQ902135; DQ902137; DQ979989; EF078542; EU180477; EU193721; FJ267679; FJ267686;  
U69735-6; U69738; U69740-1; U69754-5; U69799; U69809; U69812; U69823; U69826;  
U69829; U69832; U69845-6; U69859; U69866; U69869-70;

Cmos:

AF435016; AF435018-21; AF471096-7; AF471100-13; AF471116-20; AF471123; AF471125-8;  
AF471132; AF471134; AF471136-7; AF471141-2; AF471144-54; AF471156-65; AF544674-83;  
AF544687; AF544689; AF544694-98; AF544704-6; AF544717-20; AF544722-3; AF544727-8;  
AF544731; AF544734; AY058924; AY058933-6; AY058938-41; AY058943; AY099962;  
AY099964; AY099966-7; AY099970; AY099975-6; AY099979; AY187966-8; AY187970-4;  
AY187976-7; AY187981; AY187982-5; AY187988; AY187991-2; AY187994; AY187996-8;  
AY188001-4; AY376797; A376799-800; AY376804-5; AY376807-8; AY376813; AY376815-21;  
AY376824-5; AY444035; AY486935-6; AY486938-43; AY486944-51; AY486953-6; AY486958;  
AY586226; AY611932; AY611956; DQ112078-84; DQ465557; DQ465568; DQ486161;  
DQ486167; DQ486172; DQ486195 DQ902058-69; DQ902071-4; DQ902076-9; DQ902081;  
DQ902083-4; DQ902086; DQ902089-91; DQ902093-4; DQ902096; DQ902098-100;  
EF203999; EU366446; EU403578; EU403583-4; EU546944; FJ387197; FJ387199; FJ387209;  
FJ387212; FJ627792; FJ627804-5;

## CHAPTER 2

### CONVERGENT HEAD SHAPE IN THE DIPSADINE SNAKES AND A CENTRAL AMERICAN VIPER CLADE

Hugo Alamillo<sup>1\*</sup>, Luke J. Harmon<sup>2</sup>, Michael E. Alfaro<sup>3</sup>

<sup>1</sup> Biological Sciences  
Washington State University  
Pullman, WA 99164

<sup>2</sup> Department of Biological Sciences  
Campus Box 443051  
University of Idaho  
Moscow, ID 83844

<sup>3</sup> Department of Ecology and Evolutionary Biology  
University of California, Los Angeles  
621 Charles E. Young Drive South  
CA 90095

\* Author for correspondence (halamillo@wsu.edu)

Running title: Snake morphological mimicry

Key words: mimicry, macroevolution, colubroids, morphological modeling, dipsadines, viperids

## ABSTRACT

Dipsadine snakes are the third most species-rich snake subfamily. Their lineage diversity has been hypothesized to be, in part, the result of mimicry to vipers. This viper mimicry is hypothesized to have produced some of the most fascinating mimics on Earth, ranging from similar coloration/displays in the Gopher Snakes and the Western Rattlesnakes, to species that mimic both vipers and coral snakes at once. In this study we provide the first quantitative test that head shape in dipsadines reflects a history of mimetic evolution on vipers. We first use multivariate morphometric analysis to test if head shape in these snakes is convergent their putative viper models. We then infer a new time tree of these groups, based on mitochondrial data and three fossil calibrations. With this framework, we identify if clades within this tree are convergent in morphospace by using disparity-through-time plots. We also use model-fitting approaches to explain the evolutionary history of dipsadine and viperid head-shape similarities. The results of this study suggest that head shape in dipsadines and viperids evolved towards similar selective optima. Furthermore, mimicry in dipsadines seems to have members of this Central American viper clade as their model.

## INTRODUCTION

Viper mimicry is a striking phenomenon, but is less well-studied than coral and milk snake mimicry (Fig. 1). It is postulated to have produced mimics that range from similar coloration/displays in Gopher Snakes and Western Rattlesnakes (Kardong, 1980), to species that mimic both vipers and coral snakes at once (*Siphlophis compressus*; Savage, 1992). Viper mimicry appears pervasive within the colubroids (Fig. 2), notably in the sister clades false pit vipers (xenodontines) and slug-eating snakes (dipsadines). Mimicry is hypothesized to be a key process that led the two groups to diversify into the second and third most species-rich snake subfamilies (Green and McDiarmid, 2004) given their moderate species numbers (Dipsadinae = 197 spp., Xenodontinae = 250 spp.; Uetz and Etzold, 2007). Surprisingly, this hypothesis has never been rigorously tested .

Several convergent elements occur in mimetic snake morphology, where coloration is the most studied factor (Green and McDiarmid, 2004). Mimics will copy the coloration patterns of the model species such as the yellow, red and black displays of milk and coral snakes. Some xenodontines also mimic the aposematic coloration of coral snakes. In contrast mimicry in dipsadines is characterized by similarity in head shape rather than body coloration. This idea rests completely upon qualitative descriptions and has never been rigorously tested.

Testing head-shape mimicry hypotheses can be done in two ways. One way is to examine if dipsadines and vipers explore the same morphospace by using commonly used multivariate tools like principal components and discriminant function analyses. If mimetic evolution has occurred, mimics should overlap their modeled taxa in multivariate space more than what one



Figure 1. Snake mimicry. A) *Bothrops asper* (V) is mimicked by B) *Dipsas variegata* (NV); C) *Erythrolamprus prusbizona* (NV) mimics D) *Micrurus dumerilli* (V); V = Venomous, NV = Non-venomous. Image A) from [www.mexico-herps.com](http://www.mexico-herps.com); B) from Sebastian Lotzkat; C) and D) from Jaime Viana.

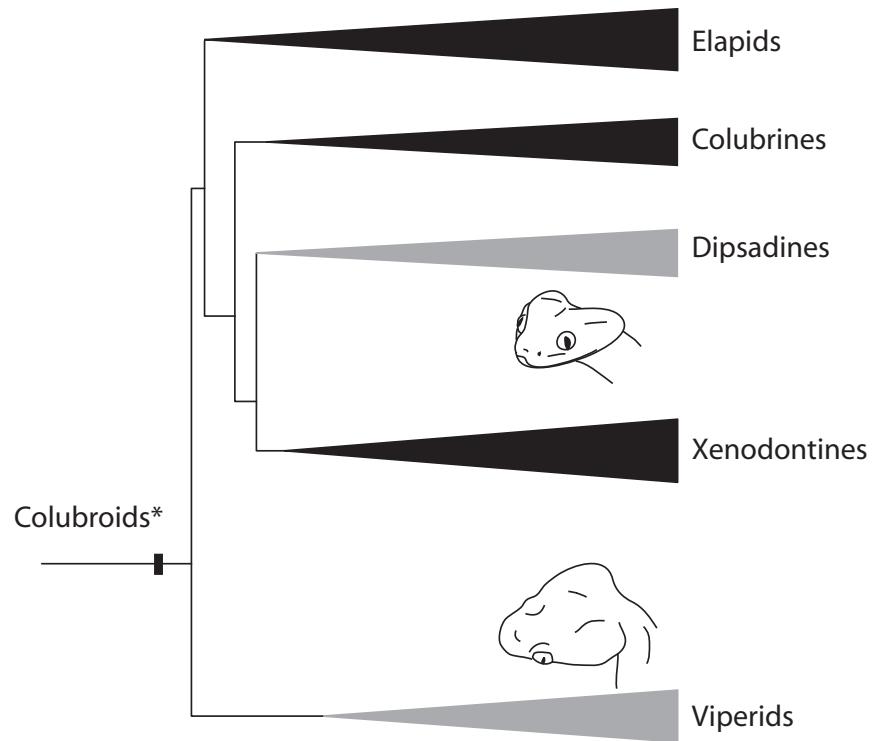


Figure 2. Groups included in this study depicting the clades (grey), dipsadines and viperids, that have been hypothesized to exploit similar head-shape morphospace. The \* indicates not all colubroid groups are included in the cladogram, atractaspids, homalopsids, natricines, and pseudoxyrophines are missing.

would expect under some random model. Another approach is to use presently available models of morphological evolution and test if similar selective peaks are being exploited by the two clades (Butler and King, 2004; O'Meara, 2006; Gavrillets and Losos, 2009; Harmon et al., 2010). If mimicry has occurred, then a single adaptive regime model is expected for both the mimic and model taxon.

In this study we provide the first quantitative test of viper mimicry using head-shape data from vipers, dipsadines, xenodontines, and elapids. We first infer a new time tree of these groups, based on mitochondrial data and three fossil calibrations. Using disparity-through-time plots we identify if clades within this tree are convergent in morphospace beyond the expectation of Brownian motion of evolution. We also use model-fitting approaches to explain the evolutionary history of dipsadine and viperid head-shape similarities.

## MATERIALS AND METHODS

*Divergence time estimation.*—We assembled a time-calibrated phylogenetic tree, chronogram, that comprised viperids, dipsadines, xenodontines and other colubroids using previously published sequences. We downloaded molecular sequences for 137 taxa from several recently published studies for two loci, 16S and 12S rRNA, for a total 754 base pairs (colubroids: Lawson et al., 2005; vipers: Castoe and Parkinson, 2006; xenodontines: Zaher et al., 2009; Supplementary Materials). These studies represent the most thorough phylogenetic analyses of dipsadine and xenodontine snakes to date. The two gene regions were aligned using MUSCLE

(Edgar, 2004) and concatenated using Mesquite v.2.73 (Madison and Madison, 2010). The molecular model of evolution was estimated for each locus by using likelihood ratio tests (LRT) in JModeltest v 0.1.1 (Posada, 2008).

We used BEAST v1.5.4 (Drummond et al., 2006) to sample the posterior probability distribution of phylogenetic trees and infer a colubroid chronogram. The MCMC chain was run for 50 million generations, after starting the analysis with a better-than-random tree that was inferred using MrBayes v.3.1 (Huelsenbeck and Ronquist, 2001). A Yule prior distribution on branch lengths was enforced. Three calibration fossils as lognormal prior distributions were used in the analysis (see Table 1 and Supplementary Materials for dates, distribution details, and justification of use). Convergence of the MCMC chain was checked with TRACER v 1.5, by checking for unimodal posterior parameter distributions and effective sample sizes (ESS > 200). The chain was sampled every 1000 generations. *Acrochordus javanicus* (Javan File Snake), a well-established sister clade to all colubroid snakes (*sensu* Lawson et al., 2007) was used as the outgroup. All phylogenetic analyses were performed using the Washington State University Biological Sciences computer cluster.

Although this study examined only 137 dipsadine and xenodontine species of their 447 total species, it captured all the lineages that are hypothesized to have experienced head-shape mimicry (i.e. *Dipsas*, *Sybinomorphus*, *Sibon*, etc). Furthermore, this taxonomic sampling included the major splits in the dipsadine evolutionary history and we expected the morphological disparity and model-fitting approaches would not be biased.

*Morphological characters sampled.*—To quantify head shape and infer morphological evolutionary patterns, we collected measurements for 135 species (1089 fluid-preserved

<b>Taxon Calibrated</b>	<b>Fossil</b>	<b>Min</b>	<b>Max</b>	<b>Source</b>
<i>Clelia</i> MRCA	<i>Clelia</i> indet.	0.8	16	Scanferla, 2006
<i>Heterodon</i> MRCA	<i>Paleoheterodon</i> sp.	23	65	Holman, 2000
<i>Natrix</i> MRCA	<i>Natrix natrix</i>	1.8	16	Szyndlar, 1991

Table 1. Fossil Calibrations used to build the colubroid chronogram.

specimens) spanning dipsadine, xenodontine, viper, and elapid clades (Supplementary Material Table 2). This data set included measurements for snout-vent length (SVL), head width (HW), intraocular distance (IOD), head length (HL), head height (HH), jaw length (JL), and sex (Fig. 3). SVL was measured by tracing a wet string from the first sublabial scale to the precloacal scale, then measuring the total string length on a meter stick to the nearest centimeter. HW, IOD, HL, HH, and JL were measured using Fowler digital calipers (Model EDP 13522) to the nearest millimeter. HW was measured at the widest point in the head, the quadratoarticular joint. IOD was measured at the widest point between the eyes on dorsum of head. HL was measured from the distal portion of the rostral scale to the groove leading down to the occipital condyles. HH was measured at the longest point between the dorsum portion of the head and the ventral part of the jaw. JL was measured from the quadratoarticular joint to the mandibular symphysis. The data were then log-transformed prior to the analyses to normalize the values' distribution. To identify independent axes of morphological variation we used a principal components analysis (PCA) and extracted several components that explained 95% of the variation. By using these morphometric measurements we expected to describe head variation in these snakes. The components found by the PCA were also used in all subsequent plots and tests.

*Disparity-through-time plots*.—We used disparity-through-time (DTT) plots to examine the time course of morphological diversification (Harmon et al., 2003). These plots identified partitioning of morphological disparity in the phylogenetic history of a clade.

This approach calculates average pair-wise distances between species and is a variance-related method for estimating point dispersion in multivariate space. Significance is calculated through simulations of morphological diversification from a null hypothesis of unconstrained

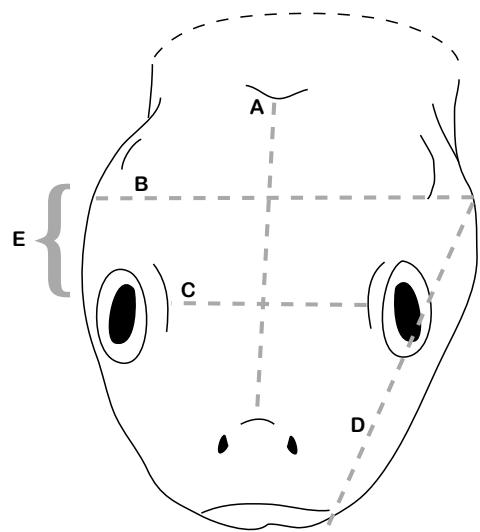


Figure 3. Head morphological measurements used in this study. A=head length, B=head width, C=intraocular distance, D=jaw length, and E=head height.

Brownian motion (BM) given the variance in the observed data set. If disparity peaks high above the null expectation, subclades are more diverse than predicted under BM and may even overlap in morphological diversity with other clades of that time period (i.e. convergent evolution). However, if disparity falls below the BM expectation we can anticipate most disparity to be among subclades of that age, which means each subclade diversifies little and does not overlap other subclades. If dipsadine head shape reflects a history of mimetic evolution on viper models, we predict that subclade disparity will be higher than the expected BM model.

Typically for statistical purposes only the first two-thirds of the phylogeny are considered in DTT plots. This is done because as the disparity plot approaches the present the effects of missing taxa and the decreased probability that clades in the present are affected by extinction might bias the disparity estimates (Harmon et al., 2003). Measurement error can also bias recent portions of the DTT plots. However because of the recent divergence in these clades, we examined nearly the entire plot excluding only the very recent portion of the plot (ca. last 15%).

*Identifying clades involved in head-shape mimicry.*—We used the *Geiger* package (v.1.0, Harmon, et al., 2008) in R (v.2.11) to examine if our data fit evolutionary scenarios in-line with mimicry. The scenarios considered were:

*Brownian Motion (BM)*: this is a single-rate random walk model. This model is representative of genetic drift or directional selection with changing optima over time and does not support mimicry.

*Ornstein-Uhlenbeck (OU)*: a random walk model with a central tendency. This model is representative of stabilizing selection and might indicate the signature of selective processes, such as mimicry.

*White Noise (WN)*: a random walk model, however unlike the BM model the variance accumulated through time is not linear. This “no phylogeny” model indicates extreme selection across the whole tree that erases the pattern of species resembling each other due to relatedness, and does not support mimicry.

We used Akaike Information Criterion to determine the best fit model ( $AIC = 2k - 2\ln(L)$ ), where k is the number of parameters in the model and L is the maximum likelihood estimate of the model; Burnham and Anderson, 2004). A better model is one that has at least four fewer AIC units.

After fitting the *Geiger* models, we used *OUCH* v.2.7.1 in R (Butler and King, 2004) to test explicitly the hypothesis that dipsadine snakes are exploring similar morphospace as viperid snakes. *OUCH* is a likelihood-based method that fits BM and OU models of phenotypic evolution, but differs from *Geiger* in that it can model multiple selective optima. The stochastic differential equation describing the OU model used in *OUCH*, for the evolution of a trait ( $X$ ), is

$$dX = \alpha(\theta(t)dt - X(t))dt + \sigma dB(t),$$

where  $\alpha$  is the strength of selection (i.e. rubber band parameter),  $\theta(t)$  is the optimum value for the trait,  $\sigma$  is the variance accumulated through time, and  $\sigma dB(t)$  is the BM process. Several observations should be made about this formula and *OUCH* in general. First, if there is no selection (i.e.  $\alpha = 0$ ), the equation reduces to the null BM model. Second, the OU model describes fixed optima ( $\theta$ ) as opposed to moving optima (Hansen, 2008). Third, fixed optima are

“painted” on the phylogeny and represent the specific hypotheses tested by the method.

Discrimination between scenarios is done by AIC, where  $\Delta\text{AIC}$  is the difference between the candidate model and best-fitting model. Lastly, the starting parameters used to maximize the likelihood of  $\alpha$  and  $\sigma$  are lower triangular matrices that are user defined. Because of this, searches need to be done to infer these starting parameters (A. King, *pers. comm.*), as opposed to using bound optima for the starting parameters like in *Geiger* (Harmon et al., 2008). Instead of searching *a posteriori* for the best-fitting starting parameters of these matrices, we used the  $\alpha$  and  $\sigma$  parameter estimates from the *Geiger* method.

We tested several possible regimes of morphological evolution (Figs. 4–7) to model processes of drift and selection on head shape and size. A scenario where morphological evolution proceeded without an adaptive optimum (BM) or in a single tree-wide OU optimum was represented by the same tree (Model name: Single Optimum; Fig. 4), but are two different models. An alternate scenario, was one where morphology evolved according to two main colubroid groups of snake forms, “viperid” and “colubrid”, yielding two separate selective optima (Model name: Viper-Colubrid Optima; Fig. 5). Another scenario we tested was one where each of the four sampled clades within colubroids (colubrids, elapids, viperids, and a “miniviperid” paraphyletic regime) each evolved morphology according to distinct selective optima (Model name: Multiple Optima; Fig. 6). In this model the dipsadines mimicry of smaller vipers like the arboreal Palm Vipers (*Bothriechis*), is represented by the “miniviperid” regime. Lastly we fit a model of evolution based on the clades that displayed convergent evolution in the DTT plot (Model name: Multiple-DTT; Fig. 7). If the dipsadines were mimicking viperids the Multiple Optima or the Multiple-DTT models should fit the best (i.e. have  $\Delta\text{AIC}=0$ ).

BM and Single Optimum Model

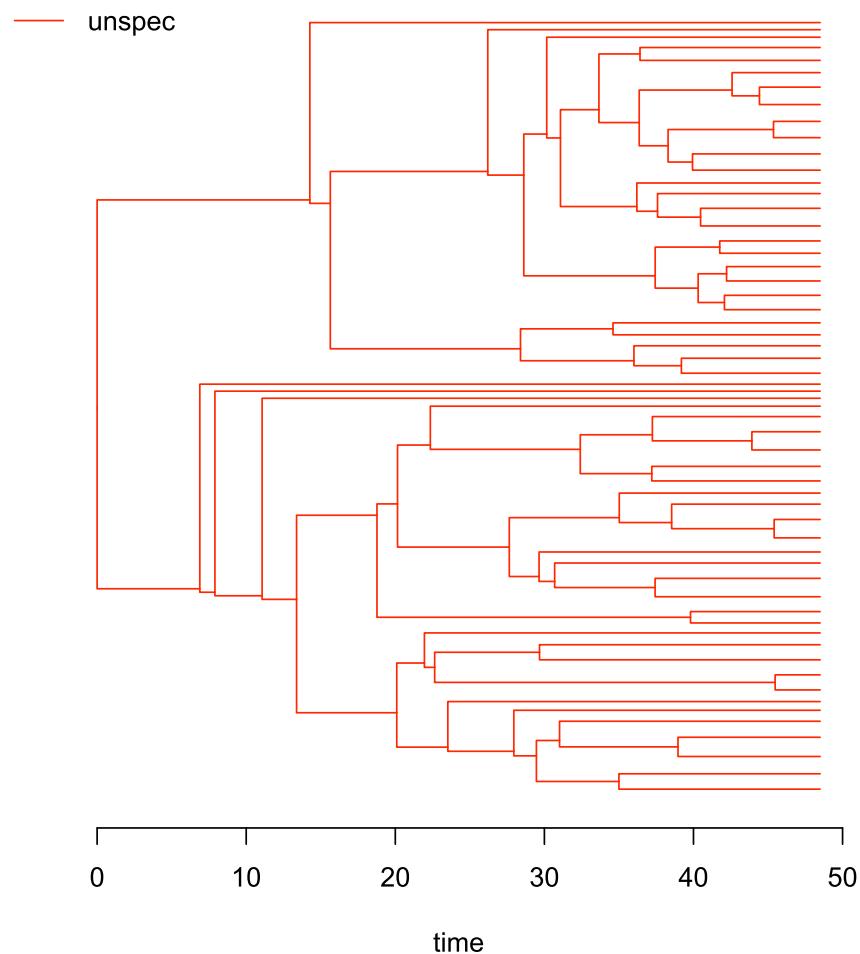


Figure 4. Brownian motion and Single Optimum model used in *OUCH*.

### Viper-Colubrid Model

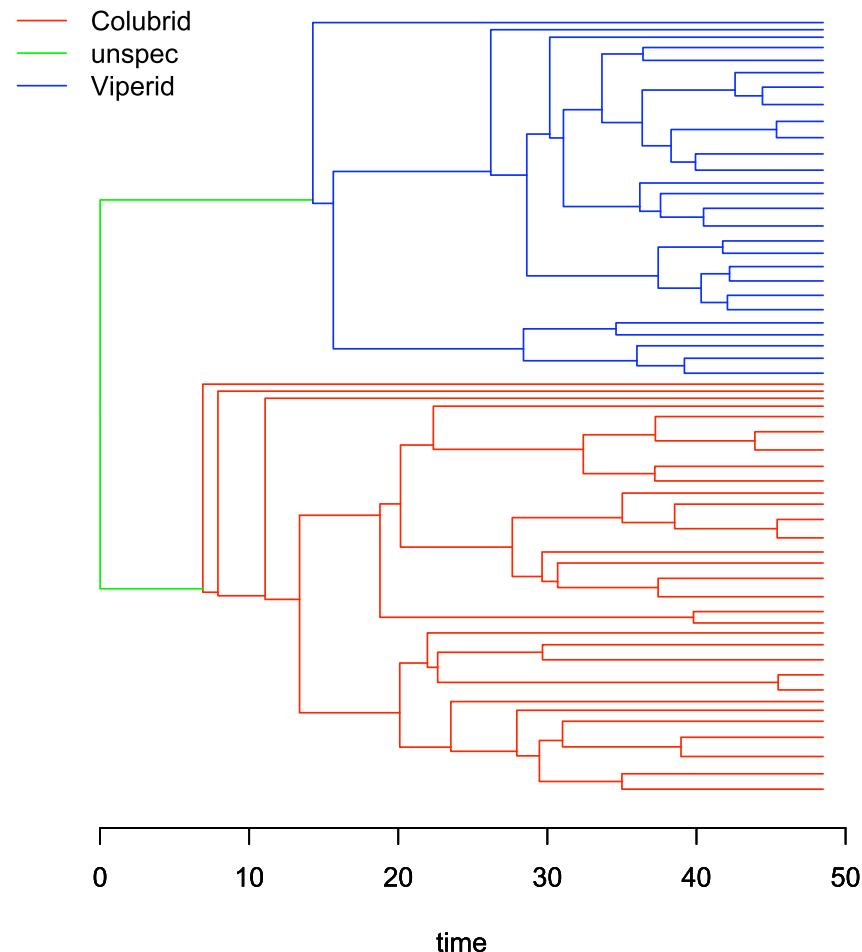


Figure 5. Viper-Colubrid model used in *OUCH* with three “painted” selective regimes, “colubrid”, “viperid”, and “unspecified” internal branches.

### Multiple Optima Model

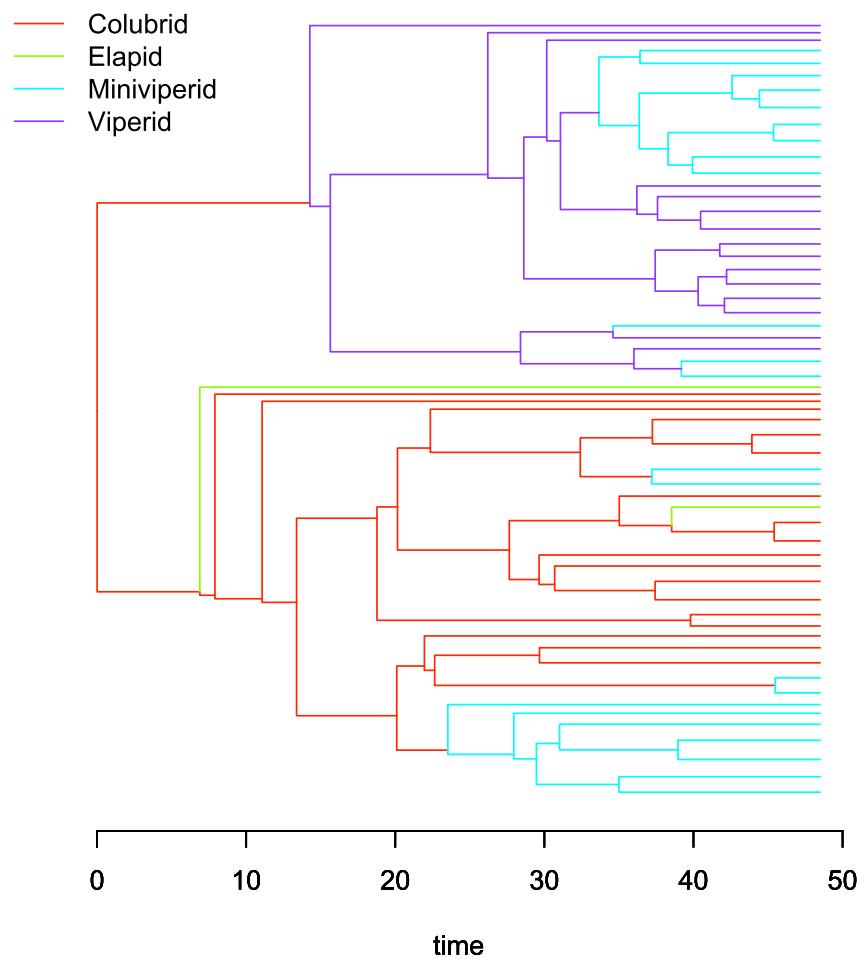


Figure 6. Multiple Optima model used in *OUCH* with four “painted” selective regimes, “colubrids”, “elapids”, “viperid”, and “miniviperid”.

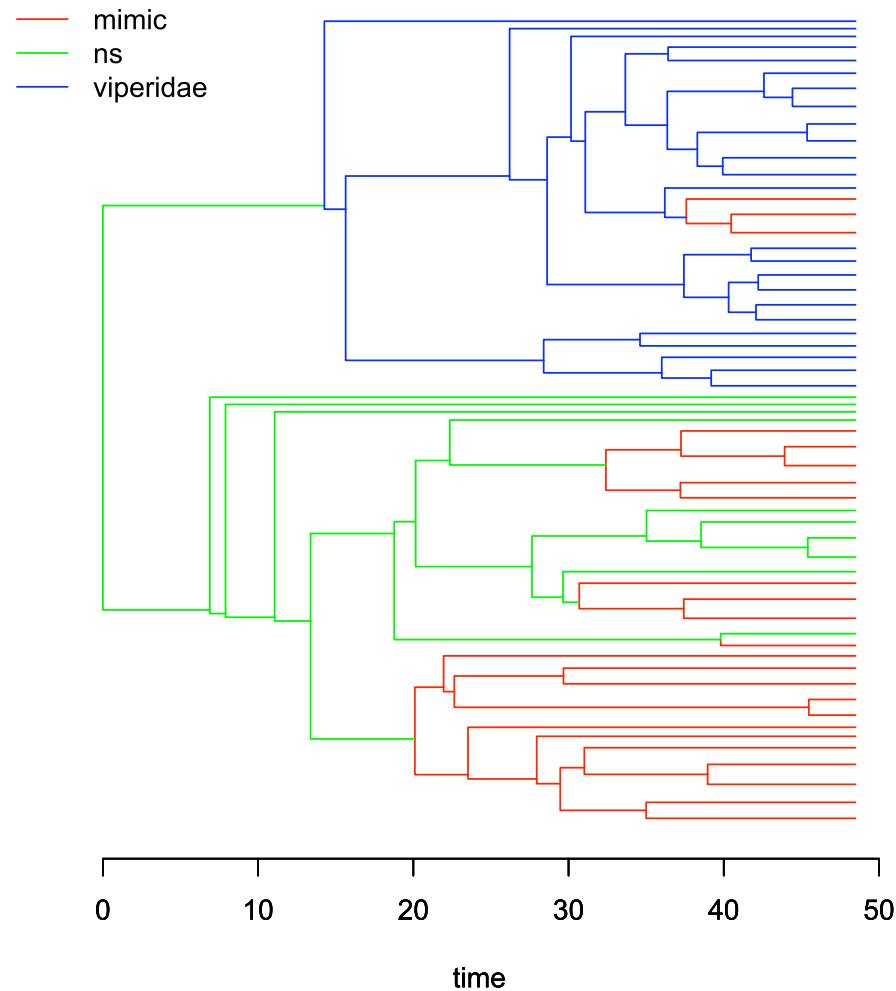


Figure 7. Multiple-DTT model used in *OUCH* after examining DTT plot with two “painted” selective regimes, “mimic” and “viperidae”.

## RESULTS

*Divergence time estimation.*—The TrN+I+Gamma model was chosen as the best model for both loci 12S and 16S alignments. All parameter traces were unimodal and had acceptable sampling (ESS > 200). After discarding the first 10% of the trees as burnin from the posterior sample, 45K trees were used to build a consensus tree for the morphological analyses (Fig. 8). Phylogenetic relationships were in concordance with other model-based hypotheses of colubroid evolution (Lawson et al., 2005; Castoe et al., 2006; Zaher et al., 2009).

*PCA results.*— PC analysis identified two axes of variation that explained 96% of the total variation (Supplementary Materials). PC1 is composed of the joint variation for Head Width, Intraocular Distance, Head Height, and Jaw length, making it a general head shape quantifier. PC2 was a representative of the variation for size (snout-vent length). Typically size loads as the first component, but because this data set had mostly head-shape variables, size loaded on the second component. Overlap of head-shape morphospace in the various groups was minimal, indicating viperids (red circles), dipsadines (green circles), elapid (yellow circle), and other colubrids (blue circles) are mostly segregating independently (Fig. 9). See Supplementary Materials for PCA loadings.

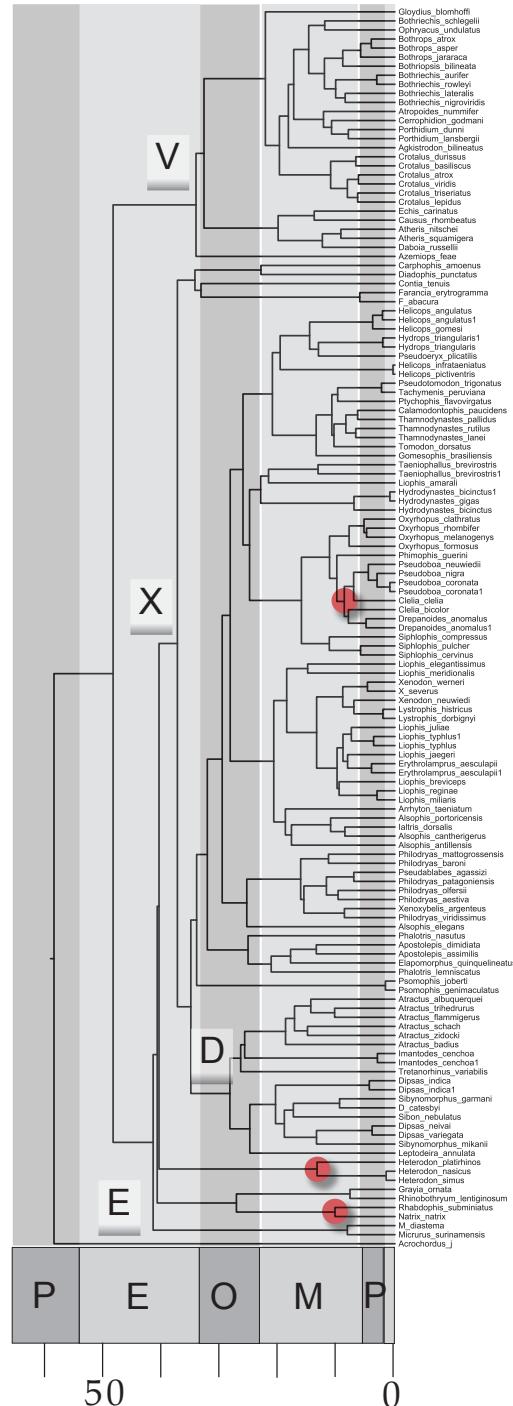


Figure 8. Chronogram used for the morphological analyses in this study. The red dots represent calibration fossils. "D" = dipsadines; "E" = elapids; "V" = Viperids; "X" = xenodontines. Time scale is in millions of years.

*Disparity-through-time plots.*—The DTT plot fell within the BM null expectation for most of the history of the colubroids (Fig. 10). This indicates that for most of the colubroid history variation has been partitioned among subclades rather than within subclades. A pronounced deviation from the null is present in the later third of the time tree (red peaks, Fig. 10). The observed disparity value was significantly different than the distribution of morphological simulations ( $n=1000$ ;  $p\text{-val}=0.008$ ), which rejects the null hypothesis of unconstrained BM. The first peak arises at 11 mya; the clades that contributes to this higher relative subclade disparity include, a *Cerrophidion* + *Porthidium* viper clade, and three South American xenodontine clades known for mimicry of vipers and coral snakes, *Alsophis* and the Pseudoboinii.

*Identifying clades involved in head-shape mimicry.*—The overall fit of the *OUCH* models and parameter estimations are summarized in Tables 2 and 3. Out of all the *OUCH* models fit to the head-shape component (PC1) the Multiple-DTT model fit the best ( $\Delta\text{AIC}=0$ , red-background cell Table 2). The next best model was BM ( $\Delta\text{AIC}=4.86$ ) closely followed by the Viper-Colubrid models ( $\Delta\text{AIC}=5.27$ ). A model of BM fit the size component (PC2) the best ( $\Delta\text{AIC}=0$ , blue background cell Table 2).

Sigma estimates for the best-fitting head-shape model (PC1 for Multiple-DTT,  $\sigma=0.007$ ) were an order of magnitude larger than the estimate for the best fitting size component model (PC2 for BM,  $\sigma=0.00035$ ). These were consistently derived for both *Geiger* and *OUCH* methods. The strength of selection estimate for the Multiple-DTT head shape component ( $\alpha=0.077$ ) was higher than the estimate for the next best fitting model with an optimum, the Viper-Colubrid model ( $\alpha=0.056$ ).

Multiple-DTT estimates place the viperid optima in the middle of the observed viperid

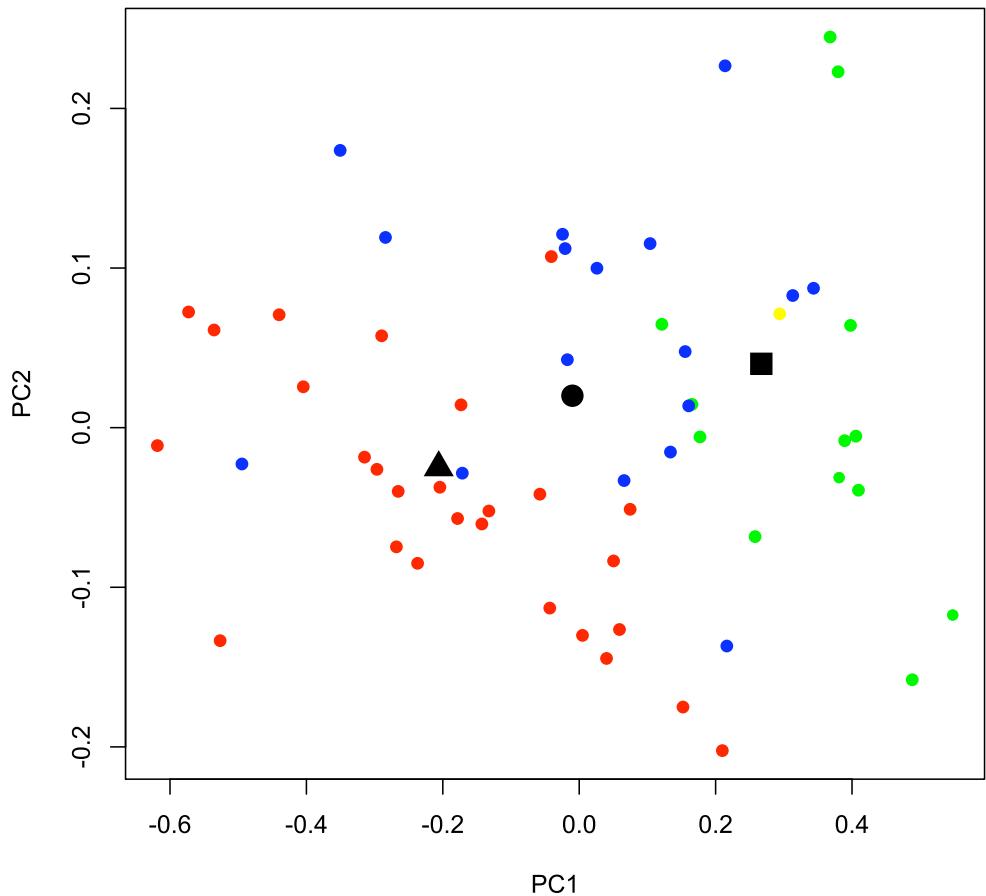


Figure 9. Plot of loadings from PC1 and PC2 in colors, and *OUCH* Multiple-DTT model estimated selective optima, in black. Red = vipers, green = dipsadines, blue = other colubrids, yellow = elapids, black circle=BM optima, black triangle=viperidae optima, black square = mimic optima

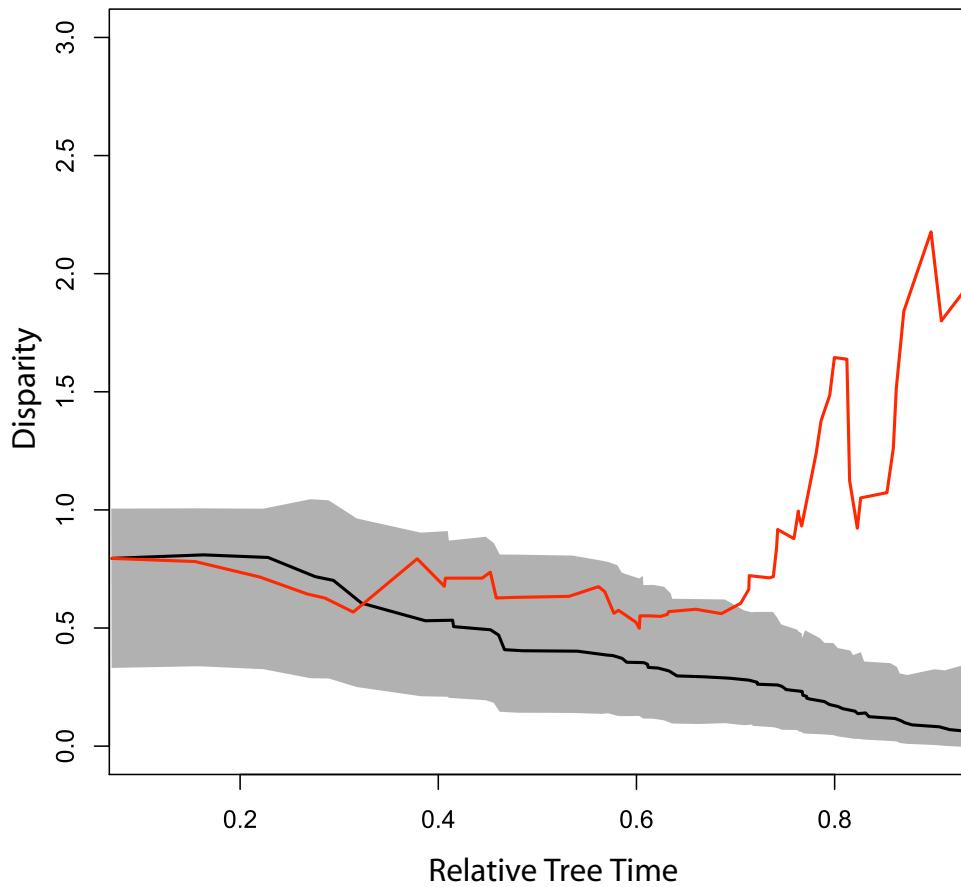


Figure 10. Disparity through time plot. Black and grey shade plot the expected morphological disparity under a BM model of evolution. The black line plots the median value of expected disparity for 1000 simulations and the grey shade represents the 95% confidence interval. The red line is the observed disparity of trait values across the phylogeny. Notice the Relative Tree Time at the origin is the root time of the time tree.

loadings (Fig. 9, black triangle), whereas the optima estimated for the mimetic clades were placed closer to other-colubrids and dipsidines (Fig. 9, black square). Ancestral optima estimated by the BM were placed closer to the other-colubrids, than the viper observed loadings (Fig. 9, black circle).

## DISCUSSION

Testing adaptive hypotheses by examining if species trait values fit the predictions from an optimality model is a standard approach of comparative biology (Hansen, 2008). Optimality models of morphological evolution fit in this study suggest that the central adaptive hypothesis of mimicry between dipsadines and viperids is supported, only when certain vipers and xenodontines are included in the models.

*Divergence-time estimation details.*—Before discussing morphological findings, some interesting details about the chronogram are worth briefly outlining. We recovered few topological discrepancies to other study. One region of the tree that was discordant with previously published phylogenies was in the divergence of elapids, natricines and *Heterodon*. This is a region with short branches and for at least two nodes, moderate support (posterior probability of 1 and 0.91, natricine and elapid clades, respectively). We also recovered divergence-time estimate disagreements to other studies. The ages recovered for the colubroids tended to be younger than those from other studies (Alamillo et al., in prep; Wiens et al., 2006)

and was likely a function of using two fast evolving mitochondrial genes. Where the Paleocene mean crown age for the caenophidia (colubroids + *Acrochordus*, 58 my, 35–68 95%HPD) is not unusual, the Eocene mean crown age of the vipers (34 my, 20–41 95%HPD) might be an underestimate. Other interesting ages include the crown age for dipsadines which we recovered in the mid Oligocene (28.4 my, 14.8–30 95%HPD). This was a time when South America was isolated from other continents and Central American highlands that are now inhabited by some dipsadines were islands (Iturralde-Vincent and MacPhee, 1999).

*Evolution of head-shape mimicry.*—Observed DTT values peaked high above the BM 95% interval in the more recent parts of the plot (Fig. 10). When focusing on this region of the plot, clades were more convergent with each other than predicted by BM (Harmon et al., 2003). An interesting pattern emerged for the first peak in the DTT plot at around 11 my (i.e. 0.8 relative tree time). During this time period *Cerrophidion* and *Porthidium*, two Mesoamerican and South American viper clades, arose. *Cerrophidion* and *Porthidium* form a monophyletic group of montane and lowland vipers that are relatively small in size (0.5–1 m SVL) in comparison to the other vipers (Campbell and Lamar, 2004; Castoe and Parkinson, 2006). Along with the viper clades, xenodontine clades also arose in this part of the time tree, which include viper-like clades (i.e. *Thamnodynastes* and *Tomodon* spp.) and enigmatic viper-like snakes with coral-like coloration (i.e. *Siphlophis* and *Oxyrhopus*).

By using the DTT plot as a tool that identifies convergent evolution we could then fit one selective optimum to the clades mentioned above. Upon fitting these clades with the same optimum in *OUCH* there was an improvement of fit over the BM model for PC1, the head-shape component. However the best fit resulted when the dipsadines were added to form the Multiple-DTT model ( $\Delta AIC=0$ ). This indicates that a single selective optimum best explains the

Geiger Models		AIC	$\Delta AIC$	lnL	sigma	alpha	OUCH Models	AIC	$\Delta AIC$	lnL	sigma	alpha	optima
PC 1	BM	2.13	0.74	0.93	0.0028	-	BM	2.13	4.86	0.93	0.0028	-	-0.018
	OU	1.39	-	2.30	0.0041	0.022	Single	4.15	6.88	1.92	0.0043	0.025	ns: 0.018 Unsp: 0.0076
	WN	26.81	25.42	-11.40	-	-	Viper-Colubrid	2.54	5.27	3.72	0.0066	0.056	Col: 0.166 Vip: -0.17
							Multiple	5.49	8.22	4.25	0.0048	0.036	Col: 0.026 Vip: -0.21 mini: 0.23 Elapid: 0.35
PC 2							Multiple-DTT	-2.73	-	7.16	0.007	0.077	Mimic: 0.267, Viperidae: -0.206
	BM	-117.56	0.27	60.78	0.00035	-	BM	-117.57	-	60.78	0.00035	-	0.020
	OU	-117.29	-	61.64	0.00053	0.022	Single	-114.60	2.97	61.30	0.00059	0.028	ns: 0.016 Unsp: 0.0055
	WN	-98.58	18.71	51.29	-	-	Viper-Colubrid	-114.58	2.99	62.29	0.00075	0.045	Col: 0.044 Vip: -0.046
							Multiple	-110.75	6.82	62.37	0.00069	0.040	Col: 0.046 mini: -0.033 Elapid: 0.081
							Multiple-DTT	-111.17	6.4	61.58	0.00065	0.035	Mimic: 0.04, Viperidae: -0.025

Table 2. Model fitting results for *Geiger* and *OUCH* methods for both PC1 and PC2 data. BM=Brownian Motion, OU=Ornstein-Uhlenbeck, WN=White Noise, AIC=Akaike Information Criterion,  $\Delta AIC$  = difference between the candidate model and best-fitting model. lnL=log likelihood.

morphological evolution of these clades, which is a signal representative of mimicry to *Cerrophidion* and *Porthidium*. For the size component (PC2), BM fit the best ( $\Delta AIC=0$ ). This is not unexpected since BM fits a scenario where randomly fluctuating directional selection has occurred on a trait, like it seems to be the case here for snout-vent length (Felsenstein, 1988).

Several inferences can be highlighted from the parameter estimates of the best-fitting *OUCH* model, the Multiple-DTT. Interestingly, expected variation in trait means per million years (the  $\sigma$  parameters) were an order of magnitude higher for the head-shape component (PC1) than the size component (PC2). Because the principal components analysis identified most variation to be located in the head-shape variables, the  $\sigma$  parameter for PC1 is larger. However, it is interesting that size within these snakes was not recovered as explaining much of the variation and might indicate a snake-specific phenomenon. This could mean that the rate of change for head shape in these colubroids was faster than body size. Such a decoupling of rate changes between head morphology and body size might be explained by ecological and speciation patterns. In terms of the ecology of the dipsadines, many of the species prefer arboreal microhabitats and nocturnal foraging strategies (Martin et al., 2009). If these arboreal species encounter predators from many more directions than ground dwelling snakes (Senter, 1999), defensive displays such as mimicry might be selected over modifications to body size. Arboreality might have also placed a constraint on body size evolution to be small and slender. Furthermore, if this was the case, slender body form could have decreased turnover of species in the dipsadines (i.e. high speciation, high extinction rates). These small-body effects have been shown in smaller bodied mammals where a group grows in lineages because of low turnover (Monroe and Bokma, 2009). Therefore, if high turnover did not occur in the dipsadines because of being small, this could have allowed the group to become the third largest clade of colubroids.

The strength of constraint on head-shape selection for the Multiple-DTT model was larger ( $\alpha=0.077$ , Table 2), but similar to the estimates of the next best-fitting selection model (Viper-Colubrid model,  $\alpha=0.056$ , Table 2). Since the best-fitting model had the highest value of  $\alpha$ , it indicates the presence of stronger stabilizing selection for the clades that were fit with the “mimic” regime. Because  $\alpha$  values are unit dependent, it was difficult to establish what strong selection was in comparison to values recovered from other studies of selective optima (Butler and King, 2004; Harmon et al., 2010). Regardless, our result supports the hypothesis of mimicry.

Examining the second-best selective model, the Viper-Colubrid model ( $\Delta\text{AIC} = 5.27$ ), is also interesting given its similar AIC value to the BM model. This model predicts that viperids and colubrids exploited separate optima (Fig. 4), which could reflect the distinct head morphologies associated with the venom delivery apparatus of viperids. This apparatus requires different osteological suspensorium arrangements giving them larger heads and longer, more fragile mandibles than the colubrid-like feeding apparatus (Cundal and Deufel, 2006). In comparison to other snakes, viperids consume prey with one-fourth the number of maxillary protractions used by other non-vipers (Pough and Groves, 1983), yet colubrids have more efficient intraoral prey transport performance (Cundal and Deufel, 2006). This exploitation of different selective optima is probably the result of viperids being “released” from the selective force that is driving the colubrids to have better intraoral ingestion in the absence of the viperid venom apparatus (Cundal and Deufel, 2006).

This study supports the long-standing hypothesis that clades of dipsadines and xenodontine snakes are mimicking Central and South American viperid snakes (Greene and McDiarmid, 2005). Future research can offshoot from this finding to tease apart how this

mimicry affected lineage diversification. Tests of lineage diversification and fitting of dispersal models for the paleohistory of dipsadines might be a worthwhile future endeavour. Furthermore, testing mimicry in some snake groups as a key innovation with available statistical comparative methods (e.g. BiSSE; Maddison et al., 2007) might explain other large-scale patterns in snakes. As more data sets become available and new methods are developed, we can continue to disentangle the evolutionary history of snakes.

## ACKNOWLEDGMENTS

This study would not have been possible without the accommodations made by Curators and Collections Managers at the following Natural History Museums: California Academy of Sciences (Dr. Robert Drewes, Jens Vindum, and Dr. Jeffery Wilkinson), The Carnegie Museum (Stephen P. Rogers), The Field Museum (Dr. Harold Voris and Alan Resetar), The University of Kansas Natural History Museum and Biodiversity Research Center (Drs. Linda Trueb and Rafe Brown, and Jonathan Campbell). We would also like to thank Drs. Harry Greene and John Cadle for interesting comments about snake mimicry and dipsadine biology. Drs. Eric Roalson and Barb Banbury, provided helpful comments on previous drafts of this paper. Funding for this project was partly provided by the Elling Award at Washington State University.

## LITERATURE CITED

- Alamillo, H., C. D. Brock, L. J. Harmon, and M. E. Alfaro. in prep. Red Queens, Court Jesters, and Snake Biodiversity: Tests of Ophidian Macroevolution.
- Burbrink, F. T and R. Alexander Pyron. 2010. How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in New World ratsnakes (Tribe Lampropeltini)? *Evolution* 64: 934–943
- Burnham, K. P. and D. R. Anderson. 2004. Model selection and inference: a practical information—theoretic approach. Sec. Ed. Springer, New York
- Butler, M. A. and A .A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* 164:683–695.
- Campbell J. A. Lamar WW. 2004. The Venomous Reptiles of the Western Hemisphere. Comstock Publishing Associates, Ithaca and London. 870 pp. 1500 plates. ISBN 0-8014-4141-2.
- Castoe, T. A. and C. L. Parkinson. 2006. Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). *Molecular Phylogenetics and Evolution*. 39:91–110.
- Cundal and Deufel, 2006. Influence of the venom delivery system on intraoral prey transport in snakes. *Zoologischer Anseiger*. 245:193–210
- Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Felsenstein, J. 1988. Phylogenies and quantitative characters. *Ann. Rev. Ecol and Systemat.* 1944–471.

- Gans, C. 1961. Mimicry in procrytically colored snakes on the genus *Daysipeltis*. *Evolution*, 15:72–91
- Greene, H. W. 1997. Snakes: the evolution of mystery in nature. University of California Press. Berkeley and Los Angeles, CA.
- Hansen, T., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution*. 62:1965–1977
- Harmon, L. J., J. A. Schulte II, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in Iguanian Lizards. *Science*. 301:961–964
- Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H., Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. Schulte II, O. Seehausen, B. L. Sidlauskas., O. Torres-Carvajal, J. T. Weir, and A. Ø. Mooers.. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution*. 64:2385–2396
- Harmon, L. J., J. Weir, C. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: Investigating evolutionary radiations. *Bioinformatics* 24:129–131
- Holman, J. A. 2000. Fossil Snakes of North America: Origin, Evolution, Distribution, Paleoecology (Life of the Past). Indiana University Press, Bloomington, IN.
- Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
- Iturrealde-Vinent, M.A. and MacPhee, R.D.E. 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bull. Am. Mus. Nat. Hist.* 238:1–95.
- Jackson, K. 2003. The evolution of venom-delivery systems in snakes. *Zoological Journal of the Linnean Society*, 137:337–354.

- Kley, N. J. and Brainerd, E. L. 1999. Feeding by mandibular raking in a snake. *Nature*. 402:369–370.
- Lawson, R., J. B. Slowinski, B. I. Crother, and F. T. Burbrink. 2005. Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes. *Mol. Phyl. Evol.*, 37:581–601.
- Losos, J. B. 1990. A phylogenetic analysis of character displacement in Caribbean *Anolis* lizards. *Evolution*, 44:558–569.
- Losos, J. B. and D. B. Miles. 2002. Testing the hypothesis that a clade has adaptively radiated: iguanid lizard clades as a case study. *Am. Nat.* 160:147–157.
- Losos, J. B. 2009. Lizards in an Evolutionary Tree, Ecology and Adaptive Radiation of Anoles. University of California Press. Berkeley CA.
- Maddison, W. P. and D.R. Maddison. 2010. Mesquite: a modular system for evolutionary analysis. Version 2.73 <http://mesquiteproject.org>
- Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56: 701–710
- Martins, M., O. A. V. Marques, and I. Sazima. 2008. How to be arboreal and diurnal and still stay alive: Microhabitat use, time of activity, and defense in neotropical forest snakes. *South American Journal of Herpetology* 3:58–67.
- Masaki, H., A. Takahiro and H. Michio. 2007 Handed snakes: convergent evolution of asymmetry for functional specialization. *Biol. Lett.* 3:169–173
- Monroe, J. M. and F. Bokma. 2009. Do speciation rates drive rates of body size evolution in Mammals? *The American Naturalist*, 174:912–918.

- Pough, F. H and J. D. Groves. 1983. Specialization of the body form and food habits of snakes. Amer. Natur. 115:92–112
- Posada, D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–6.
- Rage, J-C. 1987. Fossil Record, p. 51–76. In Snakes, Ecology and evolutionary biology 422 (ed. R. A. Seigel, J. T. Collins & S. S. Novak), McGraw-Hill Publishing
- Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecology Letters 8:336–352.
- Sànchez-Herrera, O., Smith, H.M., and Chiszar, D. 1981. Another suggested case of ophidian deceptive mimicry. Trans. Kansas Acad. Sci, 84: 121–127
- Savage, J. M, F.L.S, and J. B. Slowinski. 1992. The colouration of the venomous coral snakes (family Elapidae) and their mimics (families Aniliidae and Colubridae). Biological Journal of the Linnean Society. 45: 235–254.
- Scanferla, C. A. 2006. The oldest record of *Clelia* (Serpentes – Colubridae) in South America. C. R. Palevol. 5:721–724
- Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.
- Shine, R., W. R. Branch, P. S. Harlow, J. K. Webb, and T. Shine. 2006. Biology of Burrowing Asps (Atractaspididae) from Southern Africa. Copeia 2006:103–115.
- Simpson, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.
- Szyndlar . Z. 1991. Ancestry of the Grass Snake (*Natrix natrix*): Paleontological Evidence Journal of Herpetology,25: 412–418
- Vincent, S. E., P. D. Dang, A. Herrel, & N. J. Kley. 2006. Morphological integration and adaptation in the snake feeding system: a comparative phylogenetic study. Journal of European. Society for Evolutionary Biology.19:1545–1554.

- Vincent, S. E., M. C. Brändle, A. Herrel, M. E. Alfaro. 2009 Convergence in trophic morphology and feeding performance among piscivorous natricine snakes. *Journal of Evolutionary Biology*, 22:1203–1211.
- Uetz, P. & Etzold, T. 2007 The Reptile Database, [www.reptile-database.org](http://www.reptile-database.org).
- Van der Pijl, L. & Dodson, C. H. 1966. *Orchid Flowers: Their Pollination and Evolution*. University of Miami Press, Coral Gables, Florida.
- Walls, G. L. 1943. The vertebrate eye and its adaptive radiation. *Arch. Ophthal.* 29:1040.
- Williams, E. E. 1972. The origin of faunas. Evolution of lizard congeners in a complex island fauna: A trial analysis. *Evolutionary Biology* 6:47–89.
- Wiens, J. J., M. C. Brändle, and T. W. Reeder. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, 60:2006, pp. 123–141
- Zaher, H., Z. Felipe, G. Grazziotin, J. E. Cadle, R. W. Murphy, J. C. de Moura-Leite, and S. L. Bonatto. 2009 Molecular phylogeny of advanced snakes (Serpentes, Caenophidia) with an emphasis on South American Xenodontines: a revised classification and descriptions of new taxa Volume 49:115–153.

## SUPPLEMENTARY MATERIALS

GenBank accession numbers for specimens used in chronogram building part of this study.—

Species	16S	12S
<i>Alsophis manselli</i>	AF158459	AF158528
<i>Apostolepis assimilis</i>	GQ457781	GQ457724
<i>Apostolepis dimidiata</i>	GQ457782	GQ457725
<i>Arrhyton taeniatum</i>	AF158453	AF158522

Species	16S	12S
<i>Alsophis manselli</i>	AF158459	AF158528
<i>Atractus albuquerquei</i>	GQ457783	GQ457726
<i>Atractus badius</i>	AF158425	AF158485
<i>Atractus flammigerus</i>	AF158402	AF158471
<i>Atractus schach</i>	AF158427	AF158486
<i>Atractus trihedrurus</i>	GQ457784	GQ457727
<i>Atractus zidoki</i>	AF158426	AF158487
<i>Borikenophis portoricensis</i>	AF158448	AF158517
<i>Caaeteboia amarali</i>	GQ457807	GQ457747
<i>Calamodontophis paucidens</i>	GQ457786	GQ457728
<i>Carphophis amoenus</i>	AY577013	AY577022
<i>Clelia bicolor</i>	GQ457787	GQ457729
<i>Clelia clelia</i>	AF158403	AF158472
<i>Contia tenuis</i>	AY577021	AY577030
<i>Crisantophis nevermanni</i>	AF158405	AF158475
<i>Diadophis punctatus</i>	AF544765	AF544793
<i>Dipsas catesbyi</i>	Z46459	Z46496
<i>Dipsas indica</i>	AF158421	AF158488
<i>Dipsas indica</i>	GQ457789	GQ457730
<i>Dipsas neivai</i>	GQ457790	GQ457731
<i>Dipsas variegata</i>	AF158406	AF158476
<i>Drepanoides anomalus</i>	AF158407	AF158477
<i>Drepanoides anomalus</i>	GQ457791	GQ457732
<i>Erythrolamprus aesculapii</i>	AF158462	AF158531
<i>Erythrolamprus aesculapii</i>	GQ457795	GQ457736
<i>Elapomorphus quinquelineatus</i>	GQ457794	GQ457735
<i>Farancia abacura</i>	Z46491	Z46467
<i>Farancia erythrogramma</i>	AY577017	AY577026
<i>Gomesophis brasiliensis</i>	GQ457796	GQ457737
<i>Grayia ornata</i>	AF158434	AF158503
<i>Helicops angulatus</i>	AF158408	AF158478
<i>Helicops angulatus</i>	GQ457797	GQ457738
<i>Helicops gomesi</i>	GQ457798	GQ457739
<i>Helicops infrataeniatus</i>	GQ457799	GQ457740
<i>Helicops infrataeniatus</i>	GQ457800	GQ457741
<i>Heterodon nasicus</i>	GQ457801	AF158494
<i>Heterodon platirhinos</i>	AY577019	AY577028
<i>Heterodon simus</i>	AY577020	AY577029
<i>Hydrodynastes bicinctus</i>	AF158430	AF158479
<i>Hydrodynastes bicinctus</i>	GQ457802	GQ457742
<i>Hydrodynastes gigas</i>	GQ457803	GQ457743
<i>Hydrops triangularis</i>	AF158415	AF158499
<i>Hydrops triangularis</i>	GQ457804	GQ457744
<i>Ialtris dorsalis</i>	AF158456	AF158525
<i>Imantodes cenchoa</i>	AF158429	AF158495
<i>Imantodes cenchoa</i>	GQ457805	GQ457745
<i>Imantodes lentiferus</i>	AF158463	AF158473
<i>Leptodeira annulata</i>	AF158404	AF158473
<i>Liophis breviceps</i>	AF158464	AF158533
<i>Liophis jaegeri</i>	GQ457809	GQ457749
<i>Liophis juliae</i>	AF158445	AF158514
<i>Liophis miliaris</i>	AF158409	AF158480
<i>Liophis reginae</i>	AF158433	AF158501

Species	16S	12S
<i>Alsophis manselli</i>	AF158459	AF158528
<i>Liophis typhlus</i>	AF158410	AF158481
<i>Liophis typhlus</i>	GQ457811	GQ457751
<i>Lygophis elegantissimus</i>	GQ457808	GQ457748
<i>Lygophis meridionalis</i>	GQ457810	GQ457750
<i>Micrurus surinamensis</i>	AF544770	AF544799
<i>Natrix natrix</i>	AF158461	AF158530
<i>Oxyrhopus clathratus</i>	GQ457815	GQ457754
<i>Oxyrhopus formosus</i>	AF158411	AF158482
<i>Oxyrhopus melanogenys</i>	AF158422	AF158489
<i>Oxyrhopus rhombifer</i>	GQ457816	GQ457755
<i>Phalotris lemniscatus</i>	GQ457817	GQ457756
<i>Phalotris nasutus</i>	GQ457818	GQ457757
<i>Philodryas aestiva</i>	GQ457819	GQ457758
<i>Philodryas agassizi</i>	GQ457823	GQ457762
<i>Philodryas baroni</i>	AF158469	AF158534
<i>Philodryas mattogrossensis</i>	GQ457820	GQ457759
<i>Philodryas olfersi</i>	AF158417	AF158484
<i>Philodryas patagoniensis</i>	GQ457821	GQ457760
<i>Philodryas viridissima</i>	AF158419	AF158474
<i>Phimophis guerini</i>	GQ457822	GQ457761
<i>Pseudalsophis elegans</i>	AF158401	AF158470
<i>Pseudoboa coronata</i>	AF158412	AF158483
<i>Pseudoboa coronata</i>	GQ457824	GQ457763
<i>Pseudoboa neu edii</i>	AF158423	AF158490
<i>Pseudoboa nigra</i>	GQ457825	GQ457764
<i>Pseudoeryx plicatilis</i>	AF158418	AF158500
<i>Pseudotomodon trigonatus</i>	GQ457827	GQ457766
<i>Psomophis genimaculatus</i>	GQ457828	GQ457767
<i>Psomophis joberti</i>	GQ457829	GQ457768
<i>Ptychophis flavovirgatus</i>	GQ457830	GQ457769
<i>Rhabdophis subminiatus</i>	AF544776	AF544805
<i>Rhinobothryum lentiginosum</i>	AF158465	AF158535
<i>Sibon nebulata</i>	AF544777	AF544806
<i>Sibynomorphus garmani</i>	GQ457831	GQ457770
<i>Sibynomorphus mikanii</i>	GQ457832	GQ457771
<i>Siphlophis cervinus</i>	AF158466	AF158536
<i>Siphlophis compressus</i>	AF158467	AF158537
<i>Siphlophis pulcher</i>	GQ457834	GQ457773
<i>Tachymenis peruviana</i>	GQ457835	GQ457774
<i>Taeniophallus nicagus</i>	AF158414	AF158502
<i>Taeniophallus brevirostris</i>	GQ457793	GQ457734
<i>Thamnodynastes lanei</i>	GQ457836	GQ457775
<i>Thamnodynastes pallidus</i>	AF158420	AF158492
<i>Thamnodynastes rutilus</i>	GQ457837	GQ457776
<i>Tomodon dorsatus</i>	GQ457838	GQ457777
<i>Tretanorhinus variabilis</i>	AF158460	AF158529
<i>Tropidodryas striaticeps</i>	GQ457838	GQ457778
<i>Xenodon dorbignyi</i>	GQ457812	GQ457752
<i>Xenodon histricus</i>	GQ457813	GQ457753
<i>Xenodon neu edii</i>	GQ457841	GQ457779
<i>Xenodon severus</i>	Z46474	Z46449
<i>Xenodon werneri</i>	AF158468	AF158538

Species	16S	12S
<i>Alsophis manselli</i>	AF158459	AF158528
<i>Xenoxybelis argenteus</i>	AF158413	AF158493
<i>Daboia russelii</i>	DQ305413	AY352712
<i>Causus rhombeatus</i>	DQ305409	AJ275752
<i>Echis carinatus</i>	EU852313	EU852319
<i>Atheris nitschei</i>	AY223650	AY223663
<i>Atheris squamigera</i>	AF544762	EU624279
<i>Ophryacus undulatus</i>	AF057209	AF057256
<i>Crotalus viridis</i>	AF259253	AF259145
<i>Crotalus durissus</i>	AF259248	AF259140
<i>Porthidium lansbergii</i>	AY223655	AY223668
<i>Crotalus basiliscus</i>	AF259244	AF259136
<i>Crotalus atrox</i>	AF259258	AF259150
<i>Gloydius blomhoffi</i>	EF012803	AY352719
<i>Crotalus triseriatus</i>	AF259234	AF259124
<i>Porthidium dunni</i>	AY223654	AY223667
<i>Bothriechis schlegelii</i>	AF057213	AF057260
<i>Bothriechis aurifer</i>	DQ305425	DQ305448
<i>Bothriechis rowleyi</i>	DQ305427	DQ305450
<i>Bothriechis lateralis</i>	AF057211	AF057258
<i>Bothriopsis bilineata</i>	AF057214	AF057261
<i>Bothrops asper</i>	AF057218	AF057265
<i>Bothrops atrox</i>	AY223659	AY223672
<i>Cerrophidion godmani</i>	EU684303	DQ305442
<i>Bothriechis nigroviridis</i>	AF057212	AF057259
<i>Bothrops jararaca</i>	EU867254	EU867266
<i>Agkistrodon contortrix</i>	AF057277	AF057276
<i>Atropoides nummifer</i>	AF057207	DQ305445
<i>Micrurus diastema</i>	Z46484	Z46454
<i>Acrochordus javanicus</i>	AF512745	AF512745

### Specimen used in morphological part of study.—

*Agkistrodon contortrix* CAS 203556, CAS 203547, CAS 203548, CAS 203551, CAS 203553, CAS 203550, CAS 203559, CAS 203558, CAS 203557, CAS 203560, CAS 203563, CAS 203562, CAS 203564, CAS 90151, CAS 13930, CAS 62639, CAS 62638, CAS 62641, CAS 62640, CAS 62637, CAS 203565, CAS 71637, CAS 203567; *Alsophis antillensis* KU 267045, KU 267042, KU 267044, KU 267023, KU 267025, KU 267028, KU 267036, KU 267038, KU 267031; *Alsophis cantherigaster* KU 267073, KU 267074, KU 267077, KU 267094, KU 267098, KU 68909, KU 94549, KU 300663; *Alsophis portoricensis* KU 267164, KU 267161, KU 267157, KU 267155, KU 267151, KU 267168, KU 267182, KU 267181; *Arrhyton funerum* KU 268374, KU 268370, KU 268375, KU 268367, KU 268372, KU 268359, KU 268368, KU 268360; *Atheris nitschei* CAS 201768, CAS 176931, CAS 201656, CAS 201707, CAS 201706, CAS 201709, CAS 201708, CAS 201655, CAS 201653, CAS 201651, CAS 201654, CAS 201652; *Atheris squamigera* CAS 122703, CAS 122748, CAS 122746, CAS 122745, CAS 122744, CAS 85296, CAS 111851, CAS 153468, CAS 148629, CAS 147907, CAS 150984, KU 173095; *Atractus flammigerus* KU 125994, KU 125998, KU 125990, KU 175400, KU 125987, KU 214843, KU 214905; *Atractus badius* FMNH 43722, FMNH 40046; *Atropoides nummifer* CAS 4414, CAS 7501, CAS 163731, CAS 163772, CAS 163771; *Azemiops*

*feae* KU 312228, 312229; *Bothriechis aurifer* CAS 67049; *Bothriechis lateralis* CAS 79030, CAS 79031; *Bothriechis nigroviridis* CAS 178120; *Bothriechis rowleyi* CAS 163753; *Bothriechis schlegeli* KU 179509, KU 30954, KU 30953, KU 34001, KU 31996, KU 63921; *Bothrops asper* CAS 127442, CAS 114058, CAS 71773, CAS 73655, CAS 150330, CAS 114091, CAS 119605; *Bothrops atrox* CAS 139482, CAS 13699, CAS 94646; *Bothrops jararaca* CAS 94570, CAS 116330; *Causus maculatus* CAS 146291, CAS 103735, CAS 103736, CAS 103737, CAS 125500, CAS 123677, CAS 146040, CAS 146041, CAS 146032, CAS 146033, CAS 146035, CAS 146037, CAS 146042, CAS 146329, CAS 146328, CAS 146308, CAS 146326, CAS 146385, CAS 146379; *Cerrophidion godmani* CAS 163947, CAS 67028, CAS 67032, CAS 67031, CAS 67026, CAS 67027, CAS 67029, CAS 67033, CAS 67030, CAS 163774, CAS 163944; *Cerrophidion tzotzilorum* CAS 163520, CAS 163521, CAS 163518, CAS 163519, CAS 163770, CAS 163903; *Crotalus atrox* CAS 229233, CAS 100129, CAS 156188, CAS 192774, CAS 192772, CAS 192783, CAS 192782, CAS 192771, CAS 141800, CAS 89760, CAS 9831, CAS 141898, CAS 192792, CAS 192784, CAS 9832, CAS 35297, CAS 65085; *Crotalus basiliscus* CAS 192764, CAS 192765, CAS 192766, CAS 95764, CAS 159399, CAS 159398, CAS 24095; *Crotalus durissus* CAS 165295, CAS 154141, CAS 135256, CAS 135253; *Crotalus lepidus* CAS 48027, CAS 48026, CAS 48028, CAS 48029, CAS 48031, CAS 48030, CAS 48021, CAS 48023, CAS 48024, CAS 48022, CAS 48025, CAS 100132; *Crotalus triseriatus* CAS 103588, CAS 87175, CAS 98553, CAS 135680, CAS 5276, CAS 5846, CAS 135352, CAS 103613; *Crotalus viridis* CAS 223404, CAS 182534, CAS 223730, CAS 182539, CAS 223729, CAS 223384, CAS 204070, CAS 223558; *Daboia russelii* CAS 215454, CAS 215460, CAS 210536, CAS 206671, CAS 210838, CAS 210836, CAS 21053, CAS 210538; *Dipsas catesbyi* CAS 93331, CAS 93330, KU 300815, KU 148301, KU 148300, KU 142937, KU 175403, KU 175405, KU 158780, KU 175406, FMNH 168380, FMNH 35715, FMNH 40241, FMNH 40055, FMNH 165554; *Dipsas indica* CAS 93332, KU 112255, KU 152514, KU 105399, KU 148305, FMNH 35723, FMNH 165847; *Dipsas variegata* C 94653, CAS 14548, FMNH 217221, FMNH 217218, FMNH 215837, FMNH 215834, FMNH 217217; *Echis carinatus* CAS 174027, CAS 174028, CAS 130876, CAS 130841, CAS 130850, CAS 134075, CAS 131416, CAS 129745, CAS 129749, CAS 131532, CAS 130874; *Erythrolamprus bizona* KU 31890, KU 35732, KU 31889; *Farancia abacura* KU 204335, KU 92704; *Gloydius blomhoffi* CAS 14586, CAS 14585, CAS 14608, CAS 146603, CAS 14621, CAS 14605, CAS 14613, CAS 14610; *Helicops angulatus* KU 128255, KU 109836, KU 126026, KU 300939, KU 300941, KU 300942; *Heterodon platirhinos* KU 207296, KU 207106; *Imantodes cenchoa* C 57313, C 59095, C 51667, C 59094, KU 75700, KU 110134, KU 75696, KU 110132, KU 110133, KU 75697, KU 110135; *Imantodes lentiferus* KU 148331, KU 142941, KU 158783, KU 175415, KU 121898, KU 121895, KU 148325, KU 105407, KU 148322, KU 148326; *Leptodeira annulata* C 66605, C R191, C 66613, C 66612, CAS 13308, CAS 8722, CAS 16320, CAS 103393, CAS 103454, CAS 93203, CAS 139479, CAS 231779, CAS 94640, CAS 231456, CAS 142475, CAS 156686, CAS 169590, CAS 141007, CAS 21056, CAS 21055, KU 107711, KU 107703, KU 107706, KU 107704, KU 107710, KU 107694, KU 107708, KU 107699, KU 107698, KU 80596, KU 107697, KU 80597, KU 75714; *Leptodeira splendida* CAS 96885; *Liophis epinephelus* KU 121322, KU 135183, KU 142805, KU 121323, KU 142809, KU 110725, KU 110721, KU 110722; *Liophis reginae* KU 214886, KU 214885, KU 158535, KU 148353; *Micrurus diastema* CAS 73636, CAS 114086, CAS 4410, CAS 141872, CAS 154137; *Ophryacus undulatus* CAS 169618, CAS 154143; *Oxyrhopus*

*melanogenys* KU 148386, KU 147200, KU 112286, KU 204903; *Philodryas viridissimus* KU 301227, KU 69835, KU 289206; *Porthidium dunni* CAS 169559, CAS 73647, CAS 119919; *Porthidium lansbergii* CAS 116158, CAS 116160, CAS 116209, CAS 116210, CAS 116159, CAS 21062, CAS 116226, CAS 98362; *Pseudoboa coronata* KU 127277; *Pseudoboa neuwiedi* KU 107807, KU 112430, KU 117051, KU 182722; *Sibon nebulata* C c7875, C s7870, C s7847, C s7973, CAS 138041, CAS 138042, CAS 79002, CAS 79003, CAS 95750, CAS 94647, KU 174303, KU 157616, KU 301271, KU 157614, KU 152612, KU 112478; *Sibynomorphus catesbyi* C 53504, C R2034, C R2034; *Sibynomorphus mikanii* CAS 94294; *Siphlophis cervinus* KU 175433, KU 121926, KU 112483, KU 214901, KU 204933; *Siphlophis compressus* KU 112507, KU 167613; *Tretanorhinus variabilis* KU 268666, KU 268962, KU 268959, KU 268963, KU 268978, KU 268972, KU 268969, CAS 14455, CAS 14453, CAS 14449, CAS 14452, CAS 14458, CAS 14448, CAS 14447, CAS 14457, CAS 14459, CAS 14450, CAS 14484, CAS 14498, CAS 14496, CAS 14490, CAS 14489, CAS 14487, CAS 14493, CAS 14492, CAS 14497, CAS 14485, CAS 14488; *Xenodon rabdocephalus* KU 112516, KU 112519, KU 55731, KU 174430; *Xenodon severus* KU 105428; *Xenoxybelis boulengeri*, KU 214891, KU 204928, KU 207788;

Calibrations for the chronogram used in this study.–

*Clelia* MRCA. We used the oldest *Clelia* (Xenodontinae) fossil available. For a minimum age of the calibration we used the lower bound of the specimen horizon, 0.781 MYA (Middle Pleistocene). For the maximum age we used upper bound of the oldest known Elapidae fossil from 16 MYA. Lognormal calibration values: Offset 0.781; mean 1.921; SD 0.489.

*Heterodon* MRCA. We used *Paleoheterodon* (Xenodontinae) fossil from the Early Miocene. For a minimum age of the calibration we used 23.03 MY. For the maximum age we used the split of the Boidea/Pythonid split of 65 MYA. Lognormal Calibration Values: Offset 23.03; mean 3.044; SD 0.422

*Natrix* MRCA. We used a stem fossil of *Natrix natrix* (Natricinae) from the Lower Pleistocene. For a minimum we used the lower age of the fossil's horizon, the Pleistocene epoch (1.806). For

the maximum age we used the oldest known Elapidae fossil from 16 MYA. Lognormal calibration values: Offset 1.806; mean 1.96; SD 0.421

PCA scores for morphological measurements.–

Species	PC1	PC2
<i>Micrurus diastema</i>	0.293995375	0.071277944
<i>Heterodon platirhinos</i>	-0.494554123	-0.022759445
<i>Leptodeira annulata</i>	0.177073253	-0.005806644
<i>Sibynomorphus mikianii</i>	0.405637921	-0.005333646
<i>Dipsas variegata</i>	0.409458603	-0.039178973
<i>Sibon nebulatus</i>	0.16523857	0.014591012
<i>Dipsas catesbyi</i>	0.389251313	-0.008161277
<i>Sibynomorphus garmani</i>	0.397811622	0.064064931
<i>Dipsas indica</i>	0.257839817	-0.06827563
<i>Tretanorhinus variabilis</i>	0.121190102	0.064745432
<i>Imantodes cenchoa1</i>	0.379613716	0.222962862
<i>Imantodes cenchoa</i>	0.367902156	0.244752828
<i>Atractus badius</i>	0.547682983	-0.117396211
<i>Atractus flammigerus</i>	0.381094116	-0.03132055
<i>Philodryas viridissimus</i>	0.15536868	0.047665185
<i>Xenoxybelis argenteus</i>	0.213878886	0.226693711
<i>Alsophis antillensis</i>	0.025984903	0.09987967
<i>Alsophis cantherigerus</i>	-0.284065549	0.119175534
<i>Alsophis portoricensis</i>	-0.01737693	0.042575057
<i>Arrhyton taeniatum</i>	0.488312827	-0.157936078
<i>Liophis miliaris</i>	0.133800853	-0.015286324
<i>Liophis reginae</i>	0.065788776	-0.033139954
<i>Erythrolamprus aesculapii</i>	0.160521711	0.013662105
<i>Xenodon werneri</i>	-0.171404498	-0.028508559
<i>Siphlophis cervinus</i>	0.343568884	0.087333788
<i>Siphlophis compressus</i>	0.31309075	0.08278185
<i>Pseudoboa coronata</i>	-0.0244793	0.12115941
<i>Pseudoboa neuwiedii</i>	-0.020659828	0.112197632
<i>Oxyrhopus clathratus</i>	0.1039299	0.115314054
<i>Helicops angulatus</i>	0.21645119	-0.136808565
<i>Farancia abacura</i>	-0.350508261	0.173702068
<i>Azemiops feae</i>	-0.040927624	0.107171202
<i>Daboia russellii</i>	-0.535401312	0.061182377
<i>Atheris squamigera</i>	-0.132458095	-0.052227259
<i>Atheris nitschei</i>	0.004812984	-0.130136987
<i>Causus rhombeatus</i>	0.074638367	-0.051154291
<i>Echis carinatus</i>	0.050373134	-0.083504065
<i>Crotalus lepidus</i>	-0.057658798	-0.041683484
<i>Crotalus triseriatus</i>	0.059040428	-0.126457972

Species	PC1	PC2
<i>Crotalus viridis</i>	-0.572769496	0.072469135
<i>Crotalus atrox</i>	-0.236975092	-0.084986584
<i>Crotalus basiliscus</i>	-0.404595701	0.025608616
<i>Crotalus durissus</i>	-0.268034489	-0.074659702
<i>Agkistrodon bilineatus</i>	-0.265423972	-0.039928708
<i>Porthidium lansbergii</i>	0.152146163	-0.175029763
<i>Porthidium dunni</i>	-0.043160275	-0.113041306
<i>Cerrophidion godmani</i>	0.040042425	-0.144579894
<i>Atropoides nummifer</i>	-0.526517407	-0.133462288
<i>Bothriechis nigroviridis</i>	-0.142748207	-0.060350306
<i>Bothriechis lateralis</i>	-0.17836379	-0.056918478
<i>Bothriechis rowleyi</i>	-0.314830699	-0.018479467
<i>Bothriechis aurifer</i>	-0.618545401	-0.011239131
<i>Bothrops jararaca</i>	-0.173304425	0.014298043
<i>Bothrops asper</i>	-0.440005865	0.07072582
<i>Bothrops atrox</i>	-0.289664638	0.057545424
<i>Ophryacus undulatus</i>	-0.296635859	-0.026118519
<i>Bothriechis schlegelii</i>	0.209842383	-0.202339342
<i>Gloydius blomhoffi</i>	-0.20431314	-0.037326286

PCA rotation and importance of components.--

Rotation:	PC1	PC2
HW	<b>-0.4934415</b>	-0.25587418
HWAtEyes	<b>-0.4598592</b>	-0.21564495
HeadHeight	<b>-0.4166410</b>	-0.08539657
HeadLength	-0.3348264	0.13892910
JawLength	<b>-0.4478050</b>	-0.02352219
SVL	-0.2425227	<b>0.92783538</b>
Standard deviation	0.297	0.101
Proportion of Variance	0.860	0.099
Cumulative Proportion	<b>0.860</b>	<b>0.959</b>

## **CHAPTER 3**

### MODERATE LEVELS OF MISSING DATA LEAD TO UNDERESTIMATES OF PHYLOGENETIC BRANCH LENGTHS

Hugo Alamillo<sup>1</sup> and Jonathan M. Eastman<sup>1</sup>

<sup>1</sup> School of Biological Sciences  
Washington State University  
PO Box 644236  
Pullman, WA 99164-4236

Running title: Missing data, branch lengths and gamma

Key words: missing data, branch lengths, simulation, gamma statistic, macroevolution

## ABSTRACT

The increased availability of molecular data sets has afforded neontologists the opportunity to test long-standing hypotheses about macroevolutionary patterns. Of particular interest has been testing proposed adaptive radiations by estimating net diversification rates or calculating measures of node distributions based on the branch lengths of a phylogenetic tree. In this study, we test the hypothesis that using data matrices with missing data and a fast rate of evolution in one clade of the phylogeny, affects estimates of branch lengths. We re-estimated trees from perturbed data sets that were missing up to 75% of the molecular data. Missing data were distributed randomly or phylogenetically and had evolved along clades that were up to 25 times faster than the original trees. Our study had several methodological problems built into the simulation design that prevented us from making solid recommendations about the use of data sets with missing data. Nevertheless, we discuss trends in our results. Unexpectedly, the influence of having a clade radiating fast did not affect the phylogenetic estimates compared (branch lengths, gamma statistic, molecular evolution). The level of missing data had the greatest effect on branch-length estimates.

## INTRODUCTION

Time-calibrated phylogenies have become a major tool for the investigation of evolutionary patterns (Ayala, 1997; Sanderson et al., 1997; Thorne et al. 1998). Whether the investigator assumes a hard molecular clock or relaxes this assumption, the desired product is a framework of species relationships connected by branches that represent time since divergence (Drummond et al., 2006). For almost two decades (Nee et al., 1992; Nee, 2001), these time-calibrated branches have been used to test long-standing macroevolutionary hypotheses such as the insular radiation of the Hawaiian silverswords (Baldwin and Sanderson, 1998) or the angiosperm global takeover (Magallón and Sanderson, 2001). Yet these studies have not been without major assumptions about the phylogenetic tools used for estimating the topologies and branch lengths employed to test these ideas.

One major user-assumption supposes that incomplete molecular matrices (i.e. missing data) affect branch-length estimations throughout the tree equally. This assumption might have special impact when testing the macroevolutionary hypothesis that a subclade radiates rapidly at the beginning of its history and then slows in lineage diversification after some time (i.e. early burst model; Simpson, 1944; Schluter, 2000; Harmon et al., 2010). Such a scenario where a subclade experiences different evolutionary pressures than the rest of the tree can lead to a branch-length distribution pattern of short branches at the base of the fast-radiating subclade, and long branches leading to the tip taxa (Gavrilets and Losos, 2009). This pattern of long branches could exacerbate the effects of parameter estimation under a likelihood framework given that if the distances used in the formulae that calculate transition probabilities of molecular data under different substitution models are great (>40% distances), they could involve large sampling

errors (Yang, 2006). Other problems that can be exacerbated by this assumption include issues with model underparameterization leading to wrong patterns of cladogenesis (Revell et al., 2005). Therefore, the assumption that missing data affects all parts of the tree equally under all tempos of evolution might not be accurate.

The problem of using data sets that have missing data (i.e. incomplete gene portions) in phylogenetic analyses has been investigated to document levels of incorrect lineage placement (Wiens, 2003a and 2003b) and branch length inaccuracies (Lemmon et al., 2009). These simulation studies have sought to address the confounding effects that incomplete character matrices might have in finding a true topology. Several important results involving topological effects, decreased phylogenetic accuracy, and parameter estimates have been documented. Incorrect topologies were initially thought to be the result of large proportions of the total data matrices being scored as missing data. However, large proportions of ambiguous data do not affect lineage placement as much as a threshold of the amount of informative characters in the data, where even a very small amount can rescue phylogenetic estimates from improper lineage placement. Secondly, decreased phylogenetic accuracy is not a necessity if certain characters have missing data for some taxa. Therefore using incomplete character sets is either beneficial or harmless (Wiens, 2003a and 2003b).

To date the study that has directly tested effects of missing data on Maximum Likelihood branch-length estimation has been the Lemmon et al. (2009) study. This study considered two 500 base pair loci across 4-taxon ultrametric trees under six evolutionary rates, from invariable to saturated. One of the loci had various levels of missing data (0–100%, in increments of 50 base pairs). Their comparison of branch-length effects relied on the deviation from ultrametricity of estimated four-taxon trees. The major finding, with respect to branch lengths, was that even

low amounts of missing data led their estimated trees away from ultrametricity. Furthermore, the missing data had to be in a data set that had a fast-evolving gene and a slow-evolving gene (their Fig. 5). Although this last study is an important advancement in our understanding of how missing data affects topology reconstruction and branch-length estimation, the study estimates trees using a Jukes-Cantor model which assumes equal rates of evolution making it difficult to decipher whether the bias comes from the missing data or rate heterogeneity not being modeled. Furthermore, because the scope of this study was a comparative approach to examine the effects on ML and Bayesian estimation, no macroevolutionary scenarios were considered.

Given that studies have not provided consistent suggestions about how to handle data sets with missing data (de Queiroz and Gatesy, 2007; Kearney, 2002) and that new interest has developed in understanding diversification patterns using branch-length information, we took a two-fold approach to modeling the effects of missing data on phylogenetic estimates under a clade with fast molecular evolution. The scenario we focus on is one where one clade has evolved faster than others in terms of molecular evolution (i.e. trees that have a subclade with a long branch leading to it). We first compared branch lengths of original and simulated trees. Secondly, we quantified the effects of missing data by comparing gamma statistic values (Pybus and Harvey, 2000), a statistic that examines changes in diversification rate through time, between simulated and true trees.

By comparing branch lengths and gamma values on large scale phylogenies, we attempted a broad scale survey of the effects of missing data. This real-scenario approach that uses larger phylogenies deviates from smaller four-taxon tree simulation studies to understand the complications faced by currently undertaken diversification analyses.

## MATERIALS AND METHODS

### *Simulated Trees*

The effects of missing data on phylogenetic branch-length estimation were investigated using 5,033 simulated 60-taxon phylogenies using the *Geiger* and *Ape* packages in R v.2.10 (Harmon et. al 2008; Paradis, 2006; Fig. 1). There was an odd number of trees because random cluster jobs had errors and never finished. All starting phylogenies were generated under the same speciation and extinction rates,  $s$  and  $e$  respectively, ( $s = 0.2$  and  $e = 0.02$ ) and were ultrametric. Each tree had a subclade selected at random and transformed with a branch length multiplier to simulate rapid evolution. One original tree branch length subtending a selected clade was multiplied by a constant (either 10, 15, 20, or 25).

### *Simulated Molecular Data Sets*

Four loci were simulated across phylogenies (each 250 base pairs). The four loci were then concatenated for a total of 1000 base pairs per molecular data set. Each locus was evolved under a HKY model of evolution, but different substitution rate matrices (**Q**) were used to simulate varying degrees of molecular evolution. The four expected substitutions per site per million years treatments used were 0.01 (*slowest*), 0.1 (*slow*), 1 (*fast*), and 10 (*fastest*). To simulate rate variation among sites, a gamma-distributed rate heterogeneity function was used with shape parameter  $\alpha = 0.5$ .

*Pruning Strategies.*— Two main pruning strategies were done to each data set that was simulated, *randomly* and by *subclade*. The *random* pruning strategy allowed missing data to occur anywhere in the data matrix at the different proportions mentioned below. This was reflective of a stochastic process of data deletion, and has been likened to how morphological characters get sampled in fossil taxa (Wiens, 2003b). The *subclade* pruning strategy picked one clade that met several characteristics (e.g. matched a threshold size of subclade to be pruned) from the phylogenies generated and was therefore a phylogenetically driven pruning strategy.

Once it was decided if the data set was to be pruned *randomly* or by *subclade*, the pruning strategies were enforced and quantified as a percentage of the total amount of base pairs in each data set. Six proportions were deleted from each data set: 6, 12, 18, 25, 50, and 75%.

#### *Phylogenetic and Branch-Length Estimation*

Phylogenetic analyses were run using the MPI version of *Garli* using defaults (Zwickl, 2006). We set the *stopgen* parameter, which controls the maximum number of generations to 50,000 and the time limit to 5,000,000, while logging every 1000 generations. To optimize the runs and have them finish faster, we used the *treerejectionthreshold* set to 50. This setting controls the minimum difference in branch-length score between the best known tree and alternate trees to be considered. For more details see the *example.conf* file (Supplementary Materials).

Because the aim of this study was to compare effects of missing data on branch lengths and other parameters associated with the permuted data sets, one comparison we needed to make was between branch lengths of true and estimated trees. Having begun with ultrametric trees upon which simulations were based an appropriate comparison would be from trees transformed

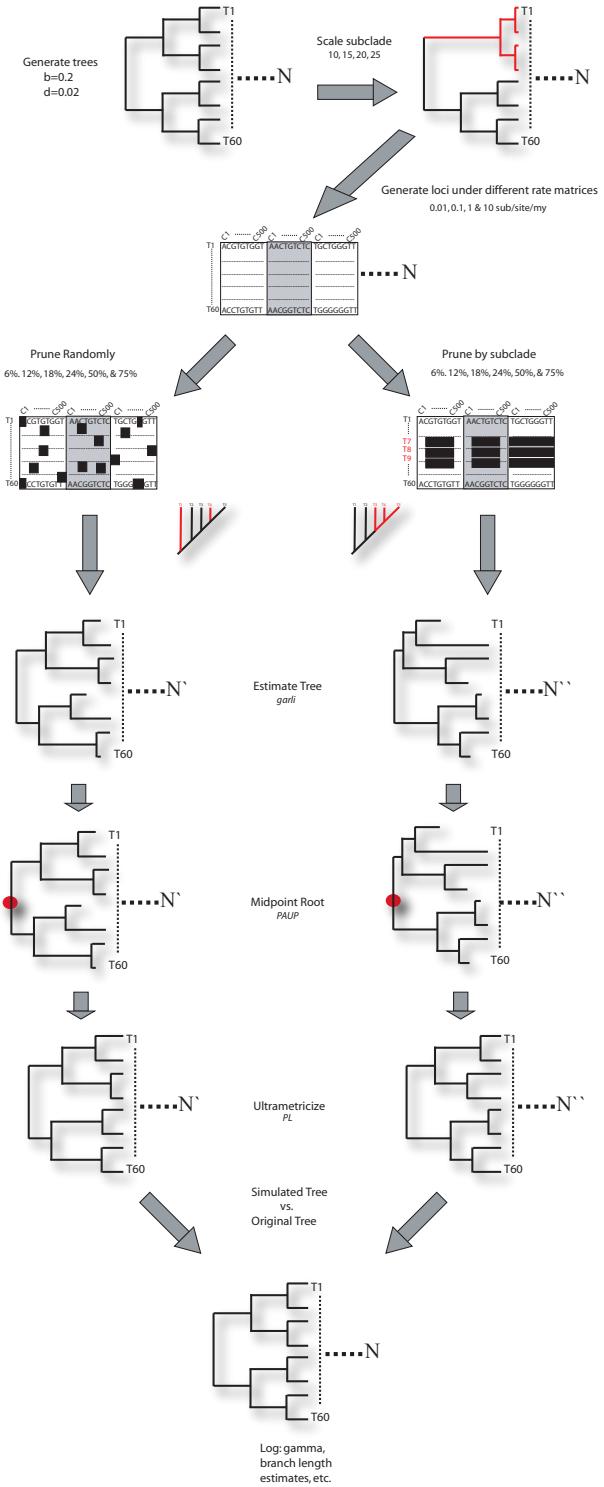


Figure 1. Flow chart of steps and methods used in this study.

to be ultrametric. To compare ultrametric trees we needed to root the simulated trees in a standardized manner, therefore all the trees were midpoint re-rooted using the Unix version of *PAUP\** v.4.0 (Swofford, 2003). With midpoint rooting the tree is rooted at the midpoint of the connection between the two taxa with the longest path between them.

Once the trees were midpoint rooted, they were ultrametricized using the Penalized Likelihood (PL) method used in the *R* package *Ape* (Sanderson, 2004; Paradis, 2006). This is a semiparametric approach which penalizes fast rate changes from branch to branch. A smoothing parameter is used to control how much penalization happens. If the smoothing parameter is large, then the model will be clock-like; if it is small then large amounts of rate variation will be allowed. To ensure an accurate estimation of the smoothing parameter, an initial calibration step was done in addition to the cross-validation procedure already done by *PL*. This step involved giving an initial range of values ( $10^{-3}$ – $10^6$ ), then calculating the maximum likelihood value over a tree. *PL* was done on the simulated trees to compare their branch lengths.

The following metrics were used to assess differences among trees:

1. *Total branch lengths of a tree*.—One way of quantifying the effects of having missing data in our molecular data sets was to examine the total branch lengths of original and estimated phylogenetic trees. For any given ultrametricized tree we summed the branch lengths within it. Although this does not point at effects on a specific branch length it serves as a proxy of the effects on branch-length estimation.
2. *Statistical test*.—To identify significant differences between pruning treatments, molecular evolution measures, and proportions pruned we used Kruskal-Wallis Rank Sum (KW) tests and Multiple Comparison (MC) test between treatments. The KW test is a non-parametric

method that compares if medians in one population are significantly different from another population (Corder and Foreman, 2009). It is analogous to a one-way ANOVA, but it does not assume a normal distribution. A significant KW test indicates that at least one group is different from the rest, but the MC test is necessary to determine which groups are different.

These tests were done using the *stats* and *pgirmess* v.1.4.7 (Giraudoux, 2010) R packages.

3. *Gamma statistic*.— The gamma statistic ( $\gamma$ ) is a summary of the diversification information contained within the waiting times of the phylogeny (Pybus and Harvey, 2000). Therefore if a phylogeny has a  $\gamma$ -value of 0 the nodes are more evenly distributed, while if the  $\gamma$ -value is negative or positive, the distribution of nodes are clustered at the base of the tree or at the tips of the tree, respectively.

All phylogenetic analyses were run using the University of Idaho *iBest* 192 CPU “*Firefly*” cluster. All simulation and analyses code was written in R v. 2.10. Sun Grid Engine scripts were written for job submittals. PERL v. 5.8.8 was used to write parsing scripts.

## RESULTS

We report results for phylogenetic estimates of branch lengths and gamma statistic values. Each of these was contrasted with prune strategy, proportions pruned, and molecular evolutionary rates. All locus rate branch-length and gamma values were joined to form the observed simulated

distribution that was compared to the original tree distribution, except when examining effects of molecular evolutionary rates.

### *Estimated Branch Lengths*

*Prune Strategy*.—The two pruning strategies, *random* and by *subclade*, recovered different branch-length estimates across all scaling treatments and underestimated the branch lengths (Fig. 2, Table 1). The smallest p-value was calculated for the comparison between the *random* and *subclade* strategies of the bl20 multiplier (p-val. = 1.76e-06). No difference is observed between the various branch-length multiplier treatments.

*Proportion Pruned*.—The *random* and *subclade* strategies recovered similar results of branch length estimation, where most treatments hovered around 100 for total branch lengths (Fig. 3, Table 1). Total sums of branch lengths for simulated trees were not significantly different across randomly pruned strategies (p-vals.>0.05 for all clade multipliers). Although a trend of increasing branch-length estimation as missing data increased was recovered, most of the subclade pruned data sets were not significantly different. However, two proportions of missing data, 50 and 75% missing data were significantly different from other pruning proportions in scaled factors for bl10 and bl15. The significantly different 75% missing data treatment was also recovered as branch lengths were multiplied for the bl20 and bl25 clades (p-vals. << 0.05).

*Molecular evolution*.—Significantly different groups for both the *random* and *subclade* strategies were nearly the same across loci and across clade multipliers (Fig. 4, Table 1). The only differing significant locus comparison was a *fastest/slow* result absent in clade multipliers greater than 10 for the *random* pruning, and in the bl25 for *subclade* pruning, (see dashed comparison line Fig. 4).

### *Estimated Gamma*

*Prune Strategy.*—The Kruskal-Wallis tests were significantly different for the *random* and *subclade* comparisons (Fig. 5, Table 2). Gamma values were also overestimated from the original trees. However, no major differences were noted among clade multiplier treatments.

*Proportion Pruned.*—*Random* pruning strategies across clade multiplier treatments did not differ for the different proportions of missing data enforced, but overestimated the value of gamma (Fig. 6, Table 2). *Subclade* pruning strategies showed a trend of decreasing gamma value as missing data increased. The molecular matrices missing 75% of the data were always significantly different from other missing taxa proportions for the *subclade* pruning.

*Molecular Evolution.*—All comparisons between all groups were significantly different for estimation of gamma, except the comparison between *fast* and *slow* loci (bar with asterisk, Fig. 7, Table 2). Among groups of both random and subclade pruning, a trend of the *fast* and *slow* loci estimating the highest gamma values is documented, while the *slowest* locus always estimated the lowest gamma value.

### Prune Strategy vs. Branch Lengths Estimated

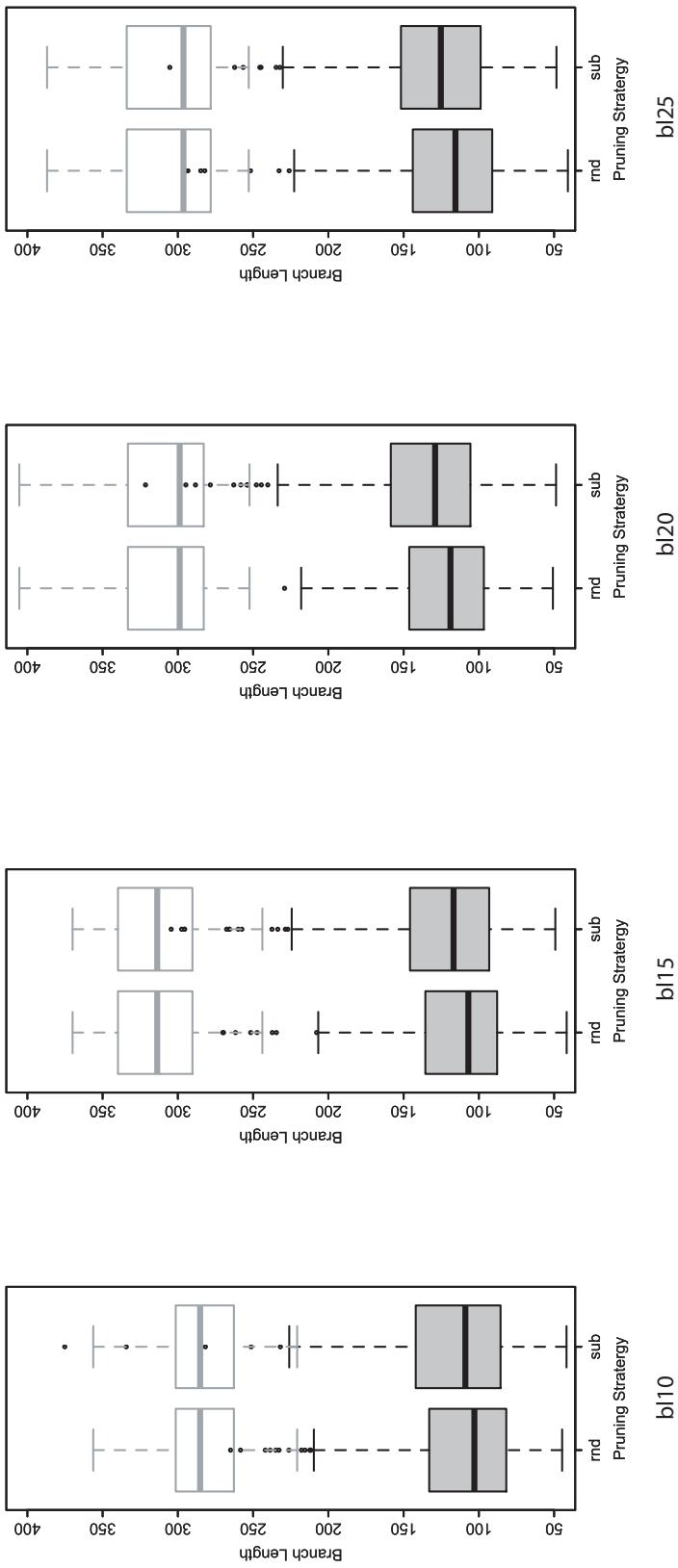


Figure 2. Pruning Strategy vs. Branch Lengths. Grey-filled box plots indicate within-group significant results according to the Kruskal-Wallis rank sum statistic and the multiple comparisons tests. Faded box plots indicate true branch lengths per pruning treatment. Total branch lengths estimated per simulated tree are outlined on the y-axis and prune strategy on the x-axis. rnd = random pruning and sub = subclade pruning. The “bl” labels refer to the multiplier used for the set of trees that simulated an adaptive radiation. Refer to Table 1 for p-values.

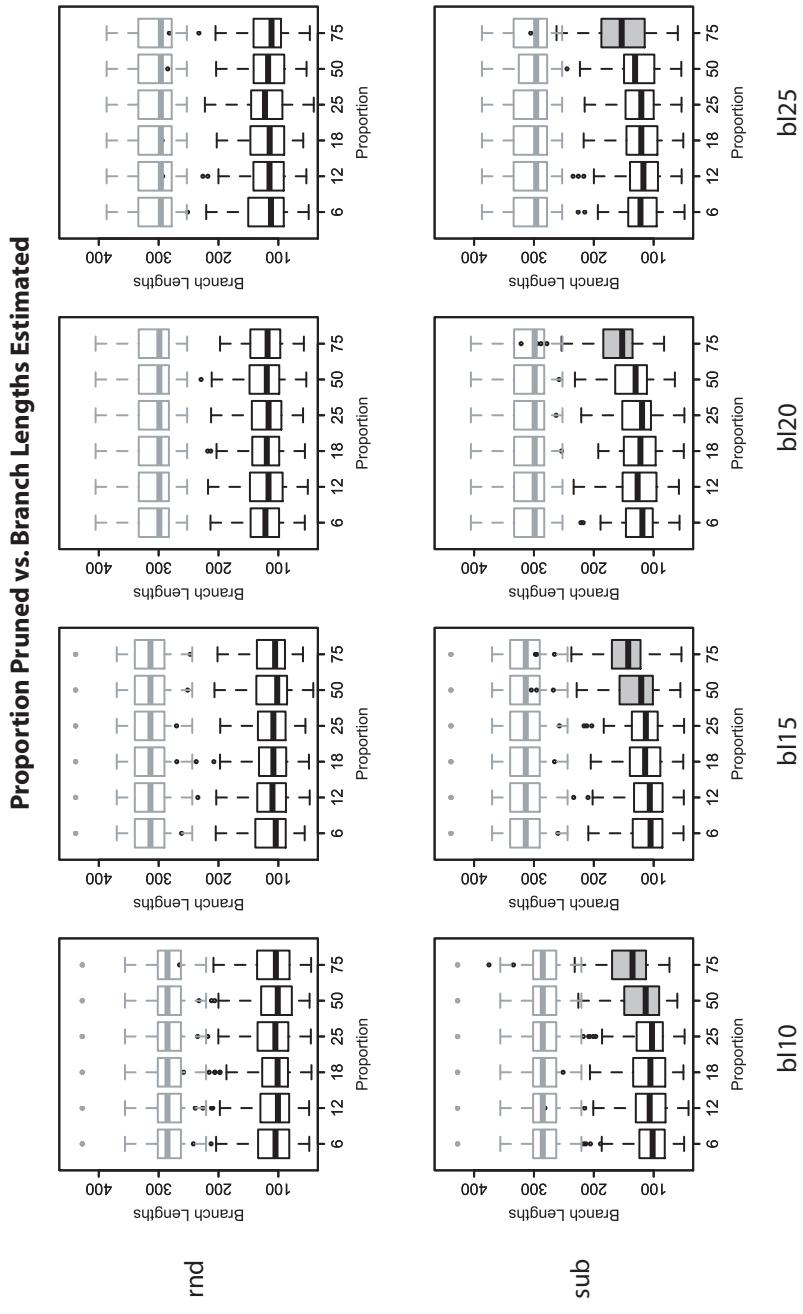


Figure 3. Proportion Pruned vs. Branch Lengths. Grey-filled box plots indicate within-group significant results according to the Kruskal-Wallis rank sum statistic and the multiple comparisons tests. Faded box plots indicate true branch lengths per pruning proportion. Total branch lengths estimated per simulated tree are outlined on the y-axis and proportion pruned on the x-axis, where 6, 12, 18, 25, 50, and 75 refer to the percentage of the data that was deleted from the molecular data matrices. rnd = random pruning and sub = subclade pruning. The “bl” labels refer to the multiplier used for the set of trees that simulated an adaptive radiation. Refer to Table 1 for p-values.

Pruning Strategy vs. Branch Lengths Estimated				
	KW			
<b>bl10</b>	0.002			
<b>bl15</b>	0.0001			
<b>bl20</b>	1.76E-06			
<b>bl25</b>	0.001			
Proportion Pruned vs. Branch Lengths Estimated				
	KW	MC		
	<u>rnd</u>	<u>sub</u>	<u>rnd</u>	<u>sub</u>
<b>bl10</b>	0.984	<b>2.76E-11</b>	—	<b>6-50, 6-75</b>
<b>bl15</b>	0.996	<b>7.87E-11</b>	—	<b>6-50, 6-75</b>
<b>bl20</b>	0.974	<b>2.98E-11</b>	—	<b>6-75</b>
<b>bl25</b>	0.991	<b>2.15E-05</b>	—	<b>6-75</b>
Molecular Evolution vs. Branch Lengths Estimated				
	KW	MC		
	<u>rnd</u>	<u>sub</u>	<u>rnd</u>	<u>sub</u>
<b>bl10</b>	<b>&lt; 2.2e-16</b>	<b>&lt; 2.2e-16</b>	NOT fast-fastest, NOT fast-slow	NOT fast-fastest, NOT fast-slow
<b>bl15</b>	<b>&lt; 2.2e-16</b>	<b>&lt; 2.2e-16</b>	NOT fast-fastest, NOT fast-slow, NOT fastest-slow	NOT fast-fastest, NOT fast-slow
<b>bl20</b>	<b>&lt; 2.2e-16</b>	<b>&lt; 2.2e-16</b>	NOT fast-fastest, NOT fast-slow, NOT fastest-slow	NOT fast-fastest, NOT fast-slow
<b>bl25</b>	<b>&lt; 2.2e-16</b>	<b>4.59E-14</b>	NOT fast-fastest, NOT fast-slow, NOT fastest-slow	NOT fast-fastest, NOT fast-slow, NOTfastest-slow

Table 1. Comparisons of branch lengths estimated. All significant results are in bold font. Names of non-significant (p-val. > 0.05) comparisons are listed for the molecular evolution. KW = Kruskal-Wallis; MC = multiple comparison.

### Molecular Evolution vs. Branch Lengths Estimated

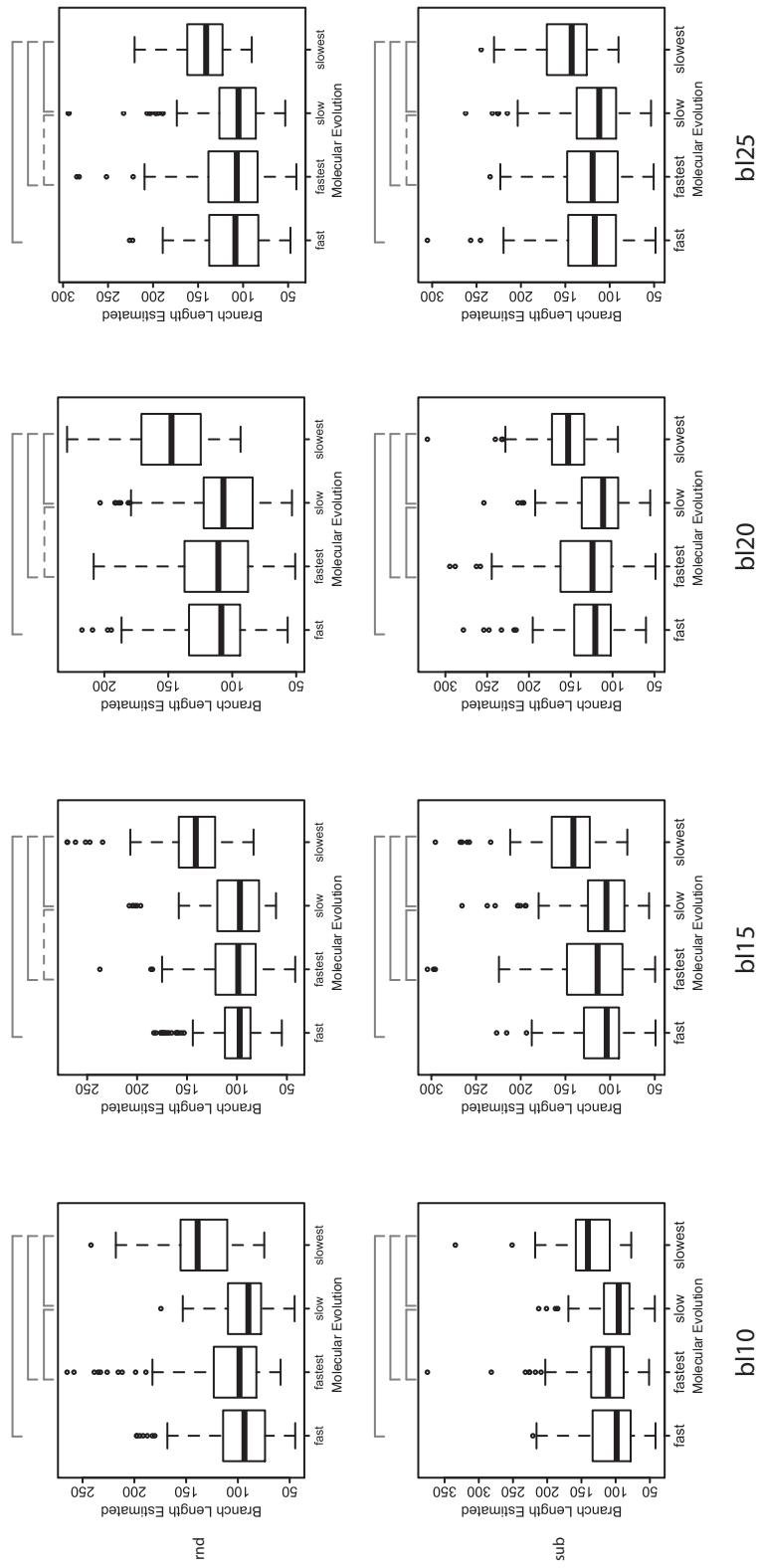


Figure 4. Molecular Evolution vs. Branch Lengths. Grey lines above box plots indicate within-group significant matching results according to the multiple comparisons test, after the Kruskal-Wallis rank sum statistic. Total branch lengths estimated per simulated tree are outlined on the y-axis and molecular evolution on the x-axis, where fast, fastest, slow, and slowest refer to the locus used to infer a tree. rnd = random pruning and sub = subclade pruning. The “bl” labels refer to the multiplier used for the set of trees that simulated an adaptive radiation. Dashed lines signify significant intergroup comparisons. Refer to Table 1 for p-values.

## DISCUSSION

Given the important role phylogenies play in our understanding of species diversification patterns (Nee, 2001; Drummond et al., 2006; Ricklefs, 2007), assessing the effects of missing data on phylogenetic trees with high levels of molecular evolution in one clade is of key importance. In this study, we addressed the problem of molecular data matrices with missing data by examining parameter estimates of branch lengths and one macroevolutionary test, the gamma statistic (Pybus and Harvey, 2000). Three general trends were highlighted from the results of the simulations done here. The first was that after enforcing even low levels of missing data in the data matrices, the estimated branch lengths were shorter (i.e. shorter sum-of-branch-lengths). Moreover, when the missing data is located in one subclade, very high levels of ambiguous characters (75%) will elongate branches. The second was an effect of decreased estimated gamma values as the missing data increased. Lastly, the presence of an increasingly fast molecular rate in one clade did not have an effect on the branch-length estimates or gamma results.

*Factors that affect branch-length estimation*

In general all treatments had a decreased estimation of branch lengths (Fig. 2, normal vs. faded comparisons). This was both true for pruning strategy and missing data proportion treatments. Underestimation of branch lengths seemed slightly better for *subclade* pruning given it had closer to true branch lengths than the *random* pruning strategies. This could be

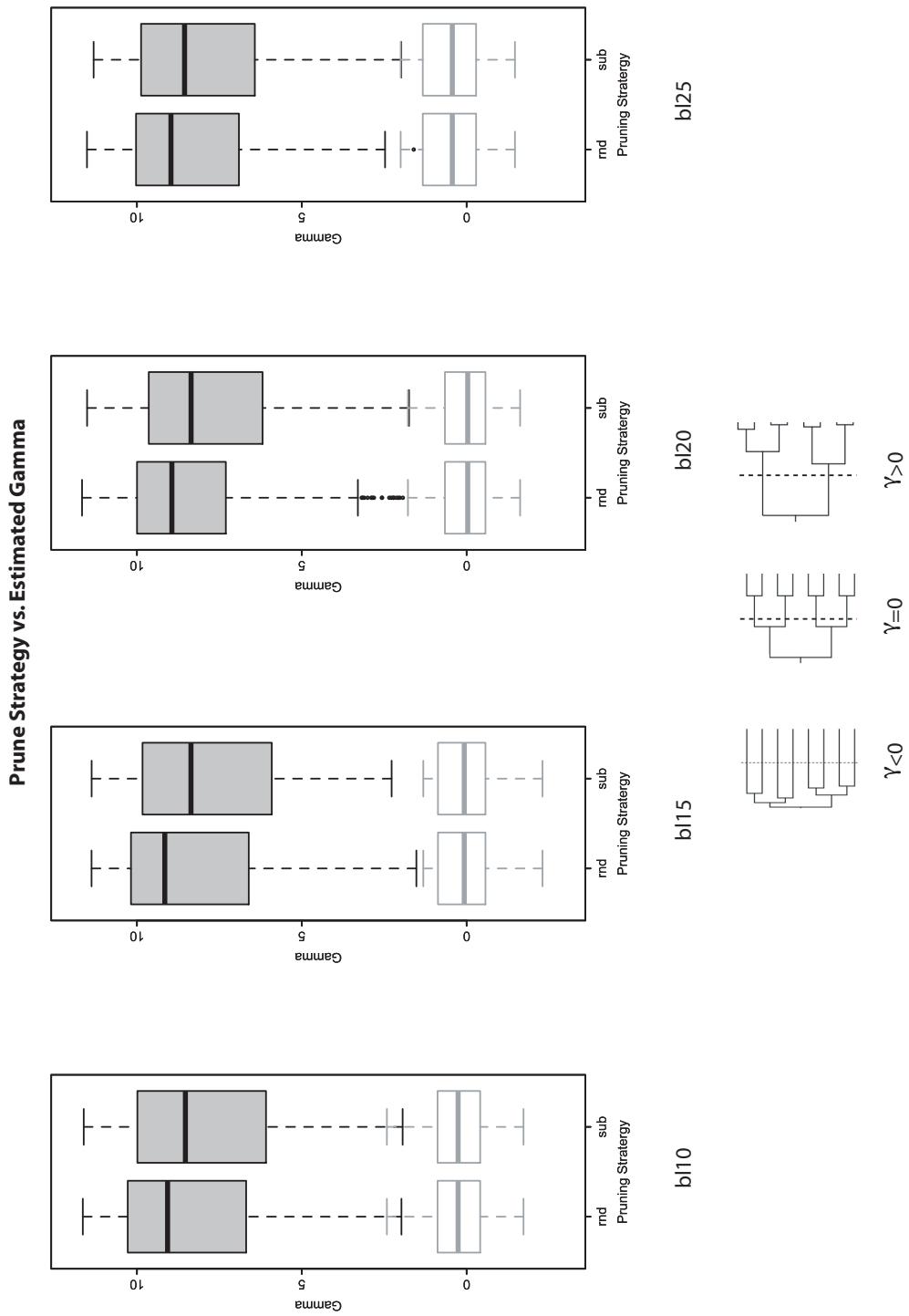
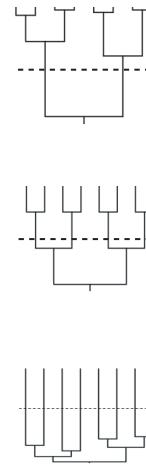
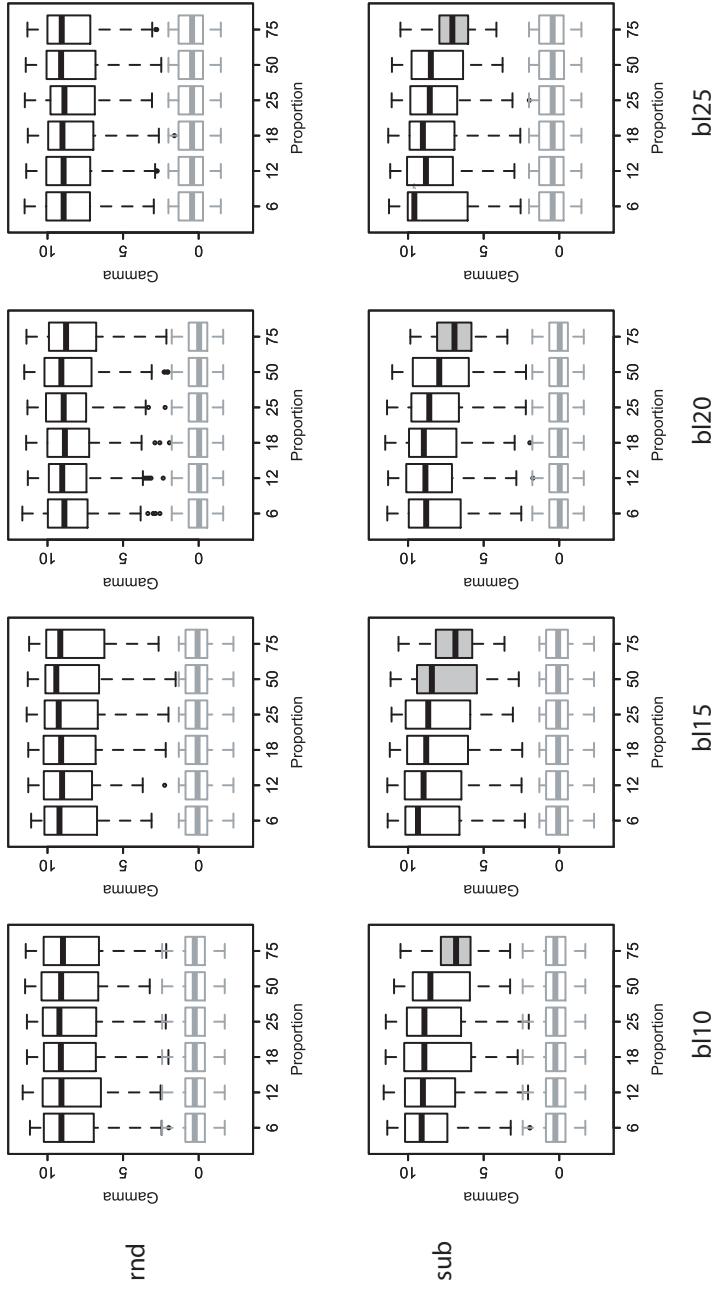


Figure 5. Pruning Strategy vs. Estimated Gamma. Grey-filled box plots were significant according to the Kruskal-Wallis rank sum statistic. Faded box plots indicate true gamma value per pruning strategy. Estimated Gamma was calculated using the Pybus and Harvey (2001) method and plotted on the y-axis. See lower figure for how gamma values match up with node distributions. Prune strategy was plotted on the x-axis. rnd = random pruning and sub = subblade pruning. The “b|” labels refer to the multiplier used for the set of trees that simulated an adaptive radiation. Refer to Table 2 for p-values.

### Proportion Pruned vs. Estimated Gamma



$\gamma < 0$        $\gamma = 0$        $\gamma > 0$

Figure 6. Proportion Pruned vs. Estimated Gamma. Estimated Gamma was calculated using the Pybus and Harvey (2001) method and plotted on the y-axis. See and multiple comparison tests. Estimated Gamma was plotted on the x-axis, where 6, 12, 18, 25, 50, and 75 refer to the percentage of the data that was deleted from the molecular data matrices. rnd = random pruning and sub = subclade pruning. The “bl” labels refer to the multiplier used for the set of trees that simulated an adaptive radiation. Faded box plots indicate true gamma value per pruning proportion. Refer to Table 2 for p-values.

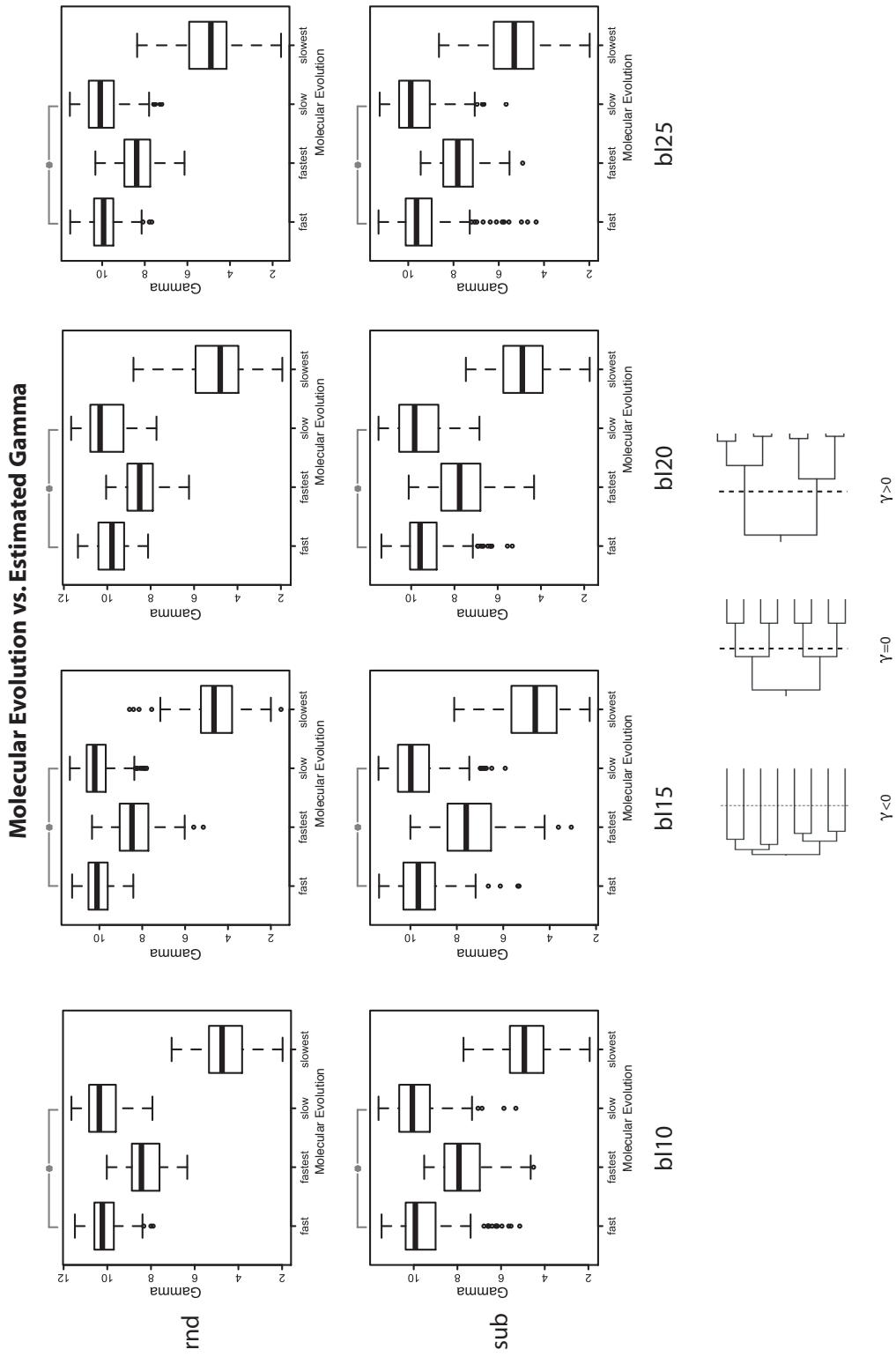


Figure 7. Molecular Evolution vs. Estimated Gamma. Grey lines above box plots with asterisk indicate the only non significant within-group results according to the multiple comparisons test, after the Kruskal-Wallis rank sum statistic. Estimated Gamma was calculated using the Pybus and Harvey (2001) method and plotted on the y-axis. See lower figure for how gamma values match up with node distributions. Molecular evolution is plotted on the x-axis, where fast, fastest, slow, and slowest refer to the locus used to infer a tree. rnd = random pruning and sub = subclade pruning. The “bl” labels refer to the multiplier used for the set of trees that simulated an adaptive radiation. Refer to Table 2 for p-values.

Prune Strategy vs. Estimated Gamma				
	KW			
bl10	<b>0.001407</b>			
bl15	<b>0.0001151</b>			
bl20	<b>6.86E-05</b>			
bl25	<b>0.01866</b>			
Proportion Pruned vs. Estimated Gamma				
	KW		MC	
	rnd	sub	rnd	sub
bl10	0.998		<b>1.08E-08</b> —	
bl15	0.988		<b>1.88E-06</b> —	
bl20	0.999		<b>6.72E-07</b> —	
bl25	0.971		<b>5.10E-05</b> —	
Molecular Evolution vs. Estimated Gamma				
	KW		MC	
	rnd	sub	rnd	sub
bl10	< 2.2e-16	< 2.2e-16	Not fast-slow	Not fast-slow
bl15	< 2.2e-16	< 2.2e-16	Not fast-slow	Not fast-slow
bl20	< 2.2e-16	< 2.2e-16	Not fast-slow	Not fast-slow
bl25	< 2.2e-16	< 2.2e-16	Not fast-slow	Not fast-slow

Table 2. Comparisons of gamma values estimated. All significant results are in bold font. For brevity, names of non-significant (p-val.> 0.05) comparisons are listed for the molecular evolution. KW = Kruskal-Wallis; MC = multiple comparison.

encouraging since structured missing data is common in real data sets. However the difference was small and this effect seems to be driven by the influence of branch elongation at high missing data categories of 50 and 75% proportions (Fig. 3). Therefore two counterintuitive phenomena are occurring: a general decrease of estimated branch lengths and a slight increase of branch lengths as ambiguous data reaches high levels (>50%).

The two phenomena might be explained by several factors. The generalized shortening of branch lengths can be explained because in our comparison of true and simulated trees, we scaled the branches in the simulated trees by the maximum length of the corresponding true tree. This was done so we could compare equally deep trees. Therefore the only way to explain the shortening of branch lengths is that root-ward branches are longer and tip-ward branches are shorter. The reason why branches are shorter at the tips might be because as missing data accumulates in a data set, the likelihood calculation will have less and less data to estimate the single-site likelihood. When this happens, fewer and fewer expected substitution per sites will be inferred, leading to shorter branches in general.

The case we recover where a subclade's branches are elongated at increasing levels of missing data might be a function of using *Penalized Likelihood*. One issue with *PL* is that there is a penalty for extremely long branches and short branches, so an over-stretching of short branches because of missing data could be leading to the longer estimates. Other problems might be the use of *PL* on phylogenies that have more than 35 taxa (Sanderson, 2004) or *PL*'s discrepancy inferring divergence times when a lineage with an average branch length is adjacent to one where a very long branch is present (Schenk and Hufford, 2010) like we enforced here.

### *Effects on gamma*

Two phenomena need to be considered when explaining the result for estimated gamma values. The first was a general trend to overestimation of the gamma values, which is indicative of nodes being pushed towards the tips of the phylogeny as would be seen in a recent radiation. This result might be expected to some extent given that the trees used here were transformed to have one clade that represented a fast genetic radiation. However, when we examine ultrametric trees that were generated under the *slowest* and *fastest* genetic evolution, both treatments have a pattern of long root-ward and short tip-ward branches. This pattern is counterintuitive to the expectation that the *slowest* locus simulated trees should have more evenly distributed nodes throughout the phylogeny. Therefore, we conclude that there was a systemic bias in the ultrametric transformation by the *PL* method.

The second was a general decrease of gamma values as high levels of missing data are included in the molecular data matrices. This interesting trend was seen as missing data increased in the *subclade* pruning (Fig. 6). Such a trend indicates that the sampled trees have more evenly distributed waiting times as missing data increases. Because branch lengths were increasing as missing data increased (Fig. 3), the signal of the fast genetic radiation decreased in *subclade* pruning.

A possible explanation of the patterns we recover here might be a violation of the assumptions of the gamma distribution. However, we simulated the original phylogenetic trees under a constant birth death model ( $s = 0.2$  and  $e = 0.02$ ) and therefore this assumption is not violated. This result is especially problematic because structuring of missing data is commonly arranged by phylogenetic proximity.

#### *Mid-point rooting problem*

One issue in our study was that we rooted our post-*Garli* trees at the mid-point. This was a problem because mid-point rooting places the root of a tree half way between the tips with the longest branch-length distance. Given that we transformed one clade to have a longer branch (bl multiplier) it is likely that the midpoint rooting always happened on this branch, therefore biasing smoothing with the *PL*.

### *Future work*

Other new approaches would be interesting to implement that can aid in understanding the effects of missing data on large scale phylogenies. One approach to address better the branch length estimation problem would be to calibrate all nodes in a tree and use *BEAST* (Drummond et al., 2006). Within the spectrum of this study, it would also be interesting to explore other effects such as heritability of rates across clades and thereby simulate phylogenetically-correlated rates to test the effects of missing data on such scenarios. And lastly, running more simulations by implementing parallelized versions of the R code (i.e., *snow Simplified* package) and thereby take full advantage of the HPC infrastructure already in use.

### *Conclusions*

Conclusions from this study should be made with a caveat because of the methodological complications we faced. This study identified a detrimental effect to identifying a fast radiation when branch lengths are estimated with high levels (>50%) of missing data and then tested for fast radiation using the gamma statistic. Furthermore, using *PL* to transform branch lengths to time could lead to exacerbation of the problem. Subclade arrangement of missing data seem to worsen the effects, which could be extra problematic if incorporating data from many studies as

would be done in a super-matrix analysis (de Queiroz and Gatesy, 2007). If these characteristics are met by a real data set subjected to phylogenetic analyses and divergence-time estimation, gross underestimating of node dating could take place in time calibrated phylogenies.

## ACKNOWLEDGMENTS

This study would benefit greatly from comments made by several individuals, Drs. Michael Alfaro, Eric Roalson, Barbara Banbury, and Luke Harmon. We are especially indebted to Colby Blair and Matt King for their help writing and running scripts in the WSU and UI computer clusters.

## LITERATURE CITED

- Ayala, F. J. 1997. Vagaries of the molecular clock. *Proc. Nat. Acad. Sci. USA.* 94: 7776–7783.
- Baldwin, B. G. and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci.* 95: 9402–9406.
- Corder, G. W. and D. I. Foreman. 2009. "Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach". New York: Wiley.
- de Queiroz A. and J. Gatesy. 2007. The supermatrix approach to systematics. *Trends Ecol. Evol.* 22:34–41.
- Drummond, A. J., S.Y.W. Ho, M. J. Phillips & A. Rambaut. 2006. PLoS Biology 4 e88 699–710.
- Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates. Sunderland, Massachusetts.
- Gavrilets, S. and J. B. Losos. 2009. Adaptive radiation: Contrasting theory with data. *Science.* 323:732–737
- Giraudoux, P. 2010. pgirmess v1.4.7: data analysis in ecology.
- Harmon, L. J., J. Weir, C. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: Investigating evolutionary radiations. *Bioinformatics* 24:129–131
- Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H., Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. Schulte II, O. Seehausen, B. L. Sidlauskas., O. Torres-Carvajal, J. T. Weir, and A. Ø. Mooers.. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution.* 64:2385–2396
- Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. *Syst. Biol.* 51, 369–381

- Lemmon, A. R., J. M. Brown, K. Stanger-Hall, and E. Moriarty Lemmon. 2009. The effect of ambiguous data on phylogenetic estimates obtained by Maximum Likelihood and Bayesian inference. *Syst. Biol.* 58: 130–145.
- Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. *Evolution*. 55:1762–1780.
- Nee, S., P. H. Harvey, and A. Ø. Mooers. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA*. 89:8322–8326.
- Nee, S. 2001. Inferring speciation rates from phylogenies. *Evolution*, 55:661–668.
- Paradis, E. 2006. *Analyses of Phylogenetics and Evolution with R*. Springer, New York..
- Pybus, O. G. and P. H. Harvey. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 267, 2267–2272.
- Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst. Biol.* 54:973–983.
- Ricklefs, R. 2007. Estimating diversification rates from phylogenetic information. *TRENDS Ecol. Evol.* 22:601–210.
- Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14: 1218–1231.
- Sanderson, M. J. 2004. r8s, version 1.70 User’s Manual.
- Schenk, J. J. and L. Hufford. 2010. Effects of substitution models on divergence time estimates: simulations and an empirical study of model uncertainty using Cornales. *Syst. Botany*. 35:578–592.

- Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.
- Simpson, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.
- Swofford, D. L. 2003. PAUP\*. Phylogenetic Analysis Using Parsimony and Other Methods. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15: 1647–1657.
- Wiens, J. J. 2003a. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst. Biol.*, 52:528–538.
- Wiens, J. J. 2003b. Incomplete taxa, incomplete characters, and phylogenetic accuracy: is there a missing data problem? *J. Vert. Paleontology*. 23:297–310.
- Yang, Z. 2006. Computational Molecular Evolution. University College London, UK
- Zwickl, D. J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

## SUPPLEMENTARY MATERIALS

Example *garli* conf file.–

```
[general]
datafname = hrtTree.1.1.fast.nex
constraintfile = none
streefname = random
ofprefix = hrtTree.1.1.fast
randseed = -1
availablememory = 1024
logevery = 1000
saveevery = 1000
refinestart = 1
outpuťeachbettertopology = 0
enforcetermconditions = 1
genthreshfortopoterm = 5000
scorethreshforterm = 0.05
significanttopochange = 0.02
outputphyliptree = 0
outputmostlyuselessfiles = 0
writecheckpoints = 0
restart = 0
ratematrix = 6rate
statefrequencies = equal
ratehetmodel = gamma
numratecats = 4
invariantsites = none
[master]
nindivs = 4
holdover = 1
selectionintensity = .5
holdoverpenalty = 0
stopgen = 50000
stoptime = 5000000
startoptprec = .5
minoptprec = .01
numberofprecreductions = 20
treerejectionthreshold = 50.0
topoweight = 1.0
modweight = .05
brlenweight = 0.2
randnniweight = 0.1
randsprweight = 0.3
limsprweight = 0.6
intervallength = 100
intervalstostore = 5
limsprrange = 6
meanbrlenmuts = 5
gammashapebrlen = 1000
gammashapemodel = 1000
uniqueswapbias = 0.1
distanceswapbias = 1.0
bootstrapreps = 0
inferinternalstateprobs = 0
```