Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular Phylogenetics and Evolution 38 (2006) 719–730
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Comparing alignment methods for inferring the history of the new world lizard genus *Mabuya* (Squamata: Scincidae)

Alison S. Whiting [a,*], Jack W. Sites Jr. [a], Katia C.M. Pellegrino [b], Miguel T. Rodrigues [c]

[a] *Department of Integrative Biology and M. L. Bean Museum, Brigham Young University, Provo, UT 84602, USA*
[b] *Pós-Graduação em Ciências Ambientais e Saúde, Universidade Católica de Goiás, Rua 232, No. 128, Setor Universitário, CEP 74210-000 Goiânia, Goiás, Brazil*
[c] *Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, Caixa Postal 11.461, CEP 05422-970 São Paulo, SP, Brazil*

Received 22 May 2005; revised 4 November 2005; accepted 4 November 2005
Available online 20 December 2005

Abstract

The rapid increase in the ability to generate molecular data, and the focus on model-based methods for tree reconstruction have greatly advanced the use of phylogenetics in many fields. The recent flurry of new analytical techniques has focused almost solely on tree reconstruction, whereas alignment issues have received far less attention. In this paper, we use a diverse sampling of gene regions from lizards of the genus *Mabuya* to compare the impact, on phylogeny estimation, of new maximum likelihood alignment algorithms with more widely used methods. Sequences aligned under different optimality criteria are analyzed using partitioned Bayesian analysis with independent models and parameter settings for each gene region, and the most strongly supported phylogenetic hypothesis is then used to test the hypothesis of two colonizations of the New World by African scincid lizards. Our results show that the consistent use of model-based methods in both alignment and tree reconstruction leads to trees with more optimal likelihood scores than the use of independent criteria in alignment and tree reconstruction. We corroborate and extend earlier evidence for two independent colonizations of South America by scincid lizards. Relationships within South American *Mabuya* are found to be in need of taxonomic revision, specifically complexes under the names *M. heathi*, *M. agilis*, and *M. bistriata* (sensu, M.T. Rodrigues, Papeis Avulsos de Zoologia 41 (2000) 313).

Keywords: Alignment; Direct optimization; Molecular phylogeny; Oceanic dispersal; *Mabuya*

1. Introduction

Phylogenetic inference has become a powerful tool for establishing and testing specific evolutionary hypotheses. Increased ability to generate large quantities of DNA sequence data has required refinement of analytical methodologies to handle large datasets, and to incorporate some aspects of molecular evolutionary processes into tree reconstruction. For instance, it is now feasible to analyze relatively large datasets with model-based methods via Bayesian techniques, and recent advances in this approach allow different models to be applied to different data partitions (Huelsenbeck and Ronquist, 2001). While great emphasis has been placed on the tree reconstruction phase of phylogenetic analysis, the development of analytical techniques which incorporate specific models of DNA evolution into the alignment of sequence data has received much less attention. Given that some commonly used markers (e.g., ribosomal RNAs and nuclear intron sequences) have proven difficult to align, multiple alignment can influence a phylogenetic topology more so than the method of tree reconstruction (Morrison and Ellis, 1997) or the specific models of sequence evolution used during tree reconstruction (Ogden and Whiting, 2003). Moreover, if a specific model of nucleotide substitution is used in reconstructing the phylogeny for a given dataset, then to be logically consistent, some have argued that the identical model should be used to produce the multiple

* Corresponding author. Fax: +1 801 422 0900.
E-mail address: as77@email.byu.edu (A.S. Whiting).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.11.011

alignment on which the tree is based (Wheeler et al., 2005). This underscores the fact that multiple sequence alignment, like tree reconstruction, is an inference process, and that the same level of sophistication should be brought to the alignment process as to the tree reconstruction process.

One methodology that attempts logical consistency throughout the entire analytical procedure is direct optimization (DO, originally called optimization alignment, Wheeler, 1996; Wheeler et al., 2005). DO obviates the need to reconstruct a multiple sequence alignment by dynamically optimizing alignment simultaneously with tree reconstruction. Thus one can reconstruct a tree without postulating a specific multiple alignment, and consistently apply a single criterion throughout the entire analytical procedure. While a theoretical justification for this methodology is beyond the scope of this paper (but see Wheeler, 1996), it is important to note that once an optimal topology is found, an implied alignment can be generated from that topology and subjected to more traditional methods of analysis (Giribet, 2005; Wheeler, 2003). DO as implemented in the computer program POY (Wheeler et al., 2003) can perform analyses using complex models of sequence evolution under a likelihood framework, allowing the same model to be applied simultaneously to both alignment and tree reconstruction. Furthermore, POY can be used as a tool to generate implied alignments, which can then be used as a multiple alignment for further analyses. For parsimony, it has been demonstrated that an implied alignment from POY will produce trees that are significantly shorter than those which can be produced under other methods of multiple sequence alignment (Ogden and Whiting, 2003; Wheeler, 2003). However, what has yet to be demonstrated is whether alignments generated under the likelihood criterion using models of evolution will result in trees with more optimal likelihood scores than can be found with other widely used methods of multiple sequence alignment.

A brief perusal of recently published empirical phylogenetic studies will show that the two most common methods for aligning DNA sequence data are: the computer program Clustal (version W or X; Thompson et al., 1994, 1997) using the default parameters, and manual alignment. As the number of taxa and genes increase, a full manual alignment quickly becomes unwieldy and impractical, aside repeatability issues. Currently, the most common method of manual alignment entails first aligning each gene in ClustalX, then manually adjusting variable regions by eye or using secondary structure (when this can be inferred), and finally discarding regions deemed "hypervariable." In this paper, we compare the performance of model-based implied alignment methods with the more commonly used methods of ClustalX and manual alignment.

The lizard family Scincidae (skinks), one of the largest families of squamate reptiles, is thought to have originated in Africa and then diversified and spread through Asia and Australia to its current worldwide distribution (Greer, 1970). Until recently, one of the largest genera of skinks was *Mabuya*, with a wide distribution primarily in South America, Africa, and Asia. In 2002, Mausfeld et al. analyzed molecular data from the mitochondrial 12S and 16S rRNA regions, and upon recovering four monophyletic species groups within *Mabuya*, subsequently split the genus into four corresponding genera based on the geographic distribution of each clade (see Fig. 1A). Asian species were placed in *Eutropis*, African and Malagasy species became *Euprepis* [and subsequently *Trachylepis* (Bauer, 2003)], species from the Cape Verde islands were returned to *Chioninia*, while the South American species retained the name *Mabuya*. Not long after, Carranza and Arnold (2003) published a paper based on cytochrome *b*, 12S and 16S rRNA data with more extensive sampling, and did not find support for the four monophyletic geographic lineages found by Mausfeld et al. (see Fig. 1B). We feel that the splitting of *Mabuya* was premature and will probably not withstand a more extensively sampled study. Therefore, herein we will not follow the nomenclature put forth by Mausfeld et al. (2002), but rather retain the generic name *Mabuya* for all taxa.

In South America, skinks are represented solely by the genus *Mabuya*, which is currently estimated to include 18–20 species ranging across much of the continent as well as many offshore islands. The exact number of species is disputed as the nomenclature and taxonomic status of many species is especially chaotic and in need of revision (Rodrigues, 2000). Herein, to avoid nomenclatural problems, we adopt the names *Mabuya bistriata* and *Mabuya ficta* according to Rodrigues (2000). South American *Mabuya* are unusual and of general evolutionary interest in two respects. First, all mainland species are viviparous even though they are confined to low-elevation tropical and subtropical habitats. This observation stands in sharp contrast to the well-supported "cold climate" hypothesis for the origin of viviparity in many clades of squamates (Méndez de la Cruz et al., 1998; Shine, 1985). The first new placental morphotype (Type IV) in over half a century was recently described to accommodate the extreme specializations seen in the chorioallantoic placenta of New World *Mabuya* (Blackburn and Vitt, 2002). Virtually all nutrients required for embryonic development are provided through a maternal–fetal nutrient transfer system that appears to be convergent with that of Eutherian mammals.

A second point of evolutionary interest is the hypothesized transoceanic colonization of South America by skinks from Africa in the Miocene (Horton, 1973). Off of the NE coast of Brazil (375 km) lies the small volcanic island group of Fernando de Noronha (Almeida, 2000), which is home to one endemic species of skink (*Mabuya atlantica*). Previous workers have suggested that *M. atlantica* is more closely related to African species of *Mabuya* than to the mainland South American *Mabuya*, and that consequently this species arrived via an independent colonization event. Morphological characters, including presacral vertebrae counts, keeled dorsal scales, coloration, and oviparity, have been presented to support the close relationship between the African species and the Fernando de Noronha endemic

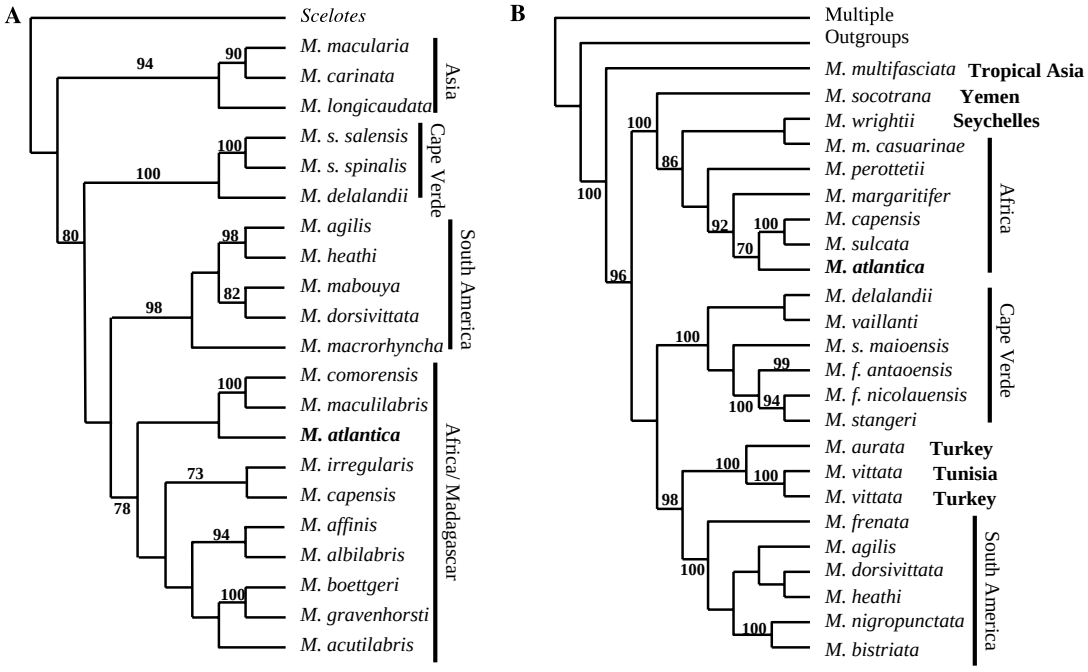

Fig. 1. Topologies from previous molecular studies on *Mabuya* (redrawn and simplified; branch lengths are not to scale and duplicate species have been removed). (A) Mausfeld et al. (2002) maximum likelihood (ML) topology based on 12S and 16S rRNAs. Numbers above branches are bootstrap proportions based on 100 pseudoreplicates, and values below 50% are not shown. The geographic localities of species are listed to the left. (B) Carranza and Arnold (2003) ML topology (greatly simplified with duplicate taxa removed), based on cyt *b*, 12S and 16S rRNAs. Numbers on the tree are Bayesian posterior probabilities, with values below 50 not shown.

(Greer et al., 2000; Horton, 1973; Travassos, 1948). Recent studies have also provided mitochondrial DNA (mtDNA) evidence supporting a second colonization event to account for the origin of *M. atlantica* (Carranza and Arnold, 2003; Mausfeld et al., 2002), and the results of these studies are summarized in Fig. 1. The Mausfeld et al. (2002) study was based on 524 bp of 16S rRNA and 399 bp of 12S rRNA for 21 species of *Mabuya* from Africa, Madagascar, South America, Asia, and the Cape Verde islands, and *Scelotes mirus* as the outgroup. They recovered *M. atlantica* within the Africa/Madagascar clade with 88% maximum parsimony (MP) bootstrap support, and 78% maximum likelihood (ML) bootstrap support. Carranza and Arnold (2003) used 305 bp of cyt *b*, 379 bp of 12S rRNA, and 388 bp of 16S rRNA for 38 individuals representing 24 species of *Mabuya*, and 8 outgroup taxa. Carranza and Arnold also recovered *M. atlantica* nested within a primarily African clade, with a Bayesian posterior probability of 1.0. Here, we extend these studies by sampling more taxa within South American *Mabuya*, including multiple nuclear gene regions, and implementing a frequentist test of the one- versus two-colonization hypothesis.

2. Materials and methods

2.1. Sampling

A total of 84 individuals of *Mabuya* were sequenced including 12 species from South America, 11 species from South and West Africa, 7 species from Madagascar, and 5 species from tropical Asia. On the basis of the prior study (Whiting et al., 2003), outgroup taxa chosen from Scincidae included representatives of the genera *Eumeces*, *Tiliqua*, *Lamprolepis*, and *Typhlosaurus*, and all trees were rooted with *Acontias* (for detailed specimen information see Appendix A). Molecular data collected included ~5000 bp of DNA sequence from the mitochondrial 12S rRNA, 16S rRNA, and cytochrome *b* (cyt *b*) gene regions, the nuclear proto-oncogene C-mos, and nuclear introns from the alpha-enolase (Enol), glyceraldehyde-3-phosphate dehydrogenase (Gapdh), and myosin heavy chain 2 (MYH2) genes. GenBank accession numbers for newly generated sequences are DQ238867–DQ239427.

DNA was extracted from liver or muscle tissue preserved in 95–100% ethanol using the Qiagen DNeasy kit (Valencia, CA, USA). DNA templates and controls were amplified using standard PCR techniques in 50 μL reactions, and products were visualized via 2% agarose gel electrophoresis. Primers and protocols for the amplification of 16S, 12S, Enol, C-mos, cyt *b*, and Gapdh are detailed elsewhere (Whiting et al., 2003, 2004). MYH2 was amplified with the primers MYH2-F (5′-GAA CAC CAG CCT CAT CAA CC-3′) and MYH2-R (5′-TGG TGT CCT GCT CCT TCT TC-3′; Dolman and Phillips, 2004; Lyons et al., 1997, 1999), using Amplitaq Gold (Perkin Elmer, Foster City, CA), and the following cycling profile: 95 °(10:00); 94 °(0:30), 62 °(0:45), 72 °(0:45) for 35 cycles; 72 °(5:00). Target products were purified using the Montage PCR_{96} Filter Plate Kit (Millipore Co., Bedford, MA) and sequenced using the Perkin Elmer Big Dye version 3 cycle sequencing

kit. Sequencing reactions were purified using Sephadex in MultiScreen Durapore PVDF plates (Millipore Co., Bedford, MA). Purified sequencing reactions were analyzed on either an ABI 3100 or an ABI 3730 automated sequencer. To insure the accuracy of sequences, negative controls were included, complementary strands were sequenced, and sequences were manually checked using the original chromatograph data in the program Sequencher 4.2 (GeneCodes Co., Ann Arbor, MI).

2.2. Alignment—ClustalX

All ribosomal and intron gene regions were input to ClustalX individually and aligned using the default parameters for gap opening (10), gap extension (0.20), delay divergent sequences (30%), and DNA transition weight (0.50; Thompson et al., 1997). 16S and 12S rRNAs were also input to ClustalX in conserved and variable regions corresponding to the POY analyses to assess any effect this may have on alignment.

2.3. Alignment—manual

The protein coding genes C-mos and cyt *b* were aligned according to conservation of the amino acid reading frame in Sequencher 4.2 (GeneCodes Co., Ann Arbor, MI). Alignments were unambiguous with no indels present in cyt *b*, and a single 3 bp indel in the C-mos sequences of *Mabuya rudis*, *Mabuya multifasciata*, and *Mabuya macularia*. The resulting alignments of C-mos and cyt *b* were used for individual analyses of these genes, as well as in total data analyses performed under all alignment criteria.

Manual alignment of the ribosomal and intron gene regions were performed by first aligning each gene in ClustalX, and then manually adjusting the resulting alignments in MacClade (Maddison and Maddison, 2000). It is a common practice in manual alignment of sequences to discard hypervariable regions as unalignable. In order to make a fair comparison between alignment methods, data cannot be excluded from some analyses and not others. Previous studies have found that up to 50% of the total characters in a dataset can be noise (random data similar to third positions or gapped sites), without having an effect on results (Wenzel and Sidall, 1999). Therefore, all analyses were performed with all the data included.

2.4. Alignment—POY

For all POY analyses, each gene was input to POY as a separate file, with 16S and 12S being divided into smaller conserved and variable regions (similar to stem and loop regions). The protein coding genes were input as "prealigned," so the alignments are not revised for these genes but the data are used in the DO of the total dataset. The combined data were analyzed in POY under both ML and MP optimality criteria. Under the ML criterion, POY uses a 10-parameter model, including possible transitions between all 4 bases and gaps, with base frequencies, invariable sites, and rate heterogeneity estimated from the data for each analysis. POYML analyses were performed under multiple submodels including: s1—all transformations including gaps are equally probable, s1g—all nucleotide transformations are equally probable with a separate parameter for gaps, s2g—separate rates for transitions, transversions, and gaps, s3g—a three parameter model plus gaps, and s6g—separate rates for all symmetric nucleotide transformations with all gaps treated equally. We performed searches under each submodel consisting of five replicates with subtree pruning regrafting (SPR) and tree bisection-reconnection (TBR) branch swapping, and tree drifting and fusing. The $-\ln L$ scores of the trees resulting from each model were compared using Akaike Information Criterion (AIC) to select the best-fit model of evolution for the data.

Once the appropriate model was determined, a search of 100 replicates was performed on an IBM 1320 Linux cluster supercomputer with tree drifting and fusing, SPR and TBR branch swapping, and an implied alignment output. The specific commands used are as follows: "—repintermediate—catchslaveoutput—parallel—likelihood—norandomizeoutgroup—sprmaxtrees 1—tbrmaxtrees 3—maxtrees 10—holdmaxtrees 100—slop 5—checkslop 10—replicates 100—multirandom—treefuse—fuselimit 10—fusemingroup 5—fusemaxtrees 100—numdriftchanges 30—driftspr—numdriftspr 10—drifttbr—numdrifttbr 10—slop 10—checkslop 10—noestimateparamsfirst—submodel s6g—invariantsitesadjust—gammaclasses 4—trullytotallikelihood—impliedalignment—seed − 1." The best tree resulting from the 100 replicate search was used as the starting tree for an additional 10 swapping runs to increase the probability that the optimal tree was found. The implied alignment corresponding to the best $-\ln L$ score was then used as the POYML alignment for all further analyses. For comparison, MP alignments were also performed in POY with the cost of gaps, transversions, and transitions being equal (1:1:1), 200 replicates, and a swapping strategy similar to that of the ML analysis described above.

2.5. Tree reconstruction

The aligned datasets were partitioned by gene region, and the AIC, as implemented in Modeltest (Posada and Crandall, 1998), was used to determine the appropriate model of sequence evolution for each gene in the POYMP, POYML, ClustalX, and manual alignments. Modeltest was also used to analyze the concatenated total datasets from all four alignment methods to find the appropriate single model for analysis of each alignment.

Each gene region from all four alignment methods was analyzed individually, as well as combined into partitions consisting of mtDNA, nuclear DNA (nucDNA), ribosomal DNA (rDNA), and intron regions (introns). These individual and partitioned analyses were used to compare alignment methods.

All seven gene regions were concatenated to form combined datasets for each alignment method. Bayesian analyses were performed in MrBayes 3.0 (Huelsenbeck and Ronquist, 2001) with 2,500,000 generations, four chains (three heated and one cold), and trees sampled every 1000 generations. Three individual Bayesian analyses were run for each combined dataset to ensure consistency, and "burn-in" was determined by plotting $-\ln L$ scores. Trees prior to stationarity were discarded, and a 50% majority rule consensus was taken from all remaining data points to obtain the final topology. After "burn-in" trees were discarded, all remaining tree scores from each of the three independent runs were combined and the mode computed, resulting in a single $-\ln L$ score for comparison across different analyses.

Each combined dataset was subjected to a partitioned Bayesian analysis using the models selected for each gene by Modeltest, with parameters allowed to vary between partitions. We believe these analyses to be the most accurate and appropriate as the specific models of evolution used were justified for each gene region from each individual alignment, thereby accounting for individual variation. In order to remove any bias in resulting $-\ln L$ scores from the use of different models in the tree reconstruction of different alignments, additional Bayesian analyses were performed on all four alignments using a single model (GTR + G + I) of evolution for all gene regions. Lastly, all alignments were subjected to an MP analysis in PAUP* (Swofford, 2002), with 10,000 random sequence additions, equal weighting of all characters, TBR branch swapping, and gaps treated as a fifth state. The three different types of analyses will be referred to, from this point on, as mixed model Bayesian, single model Bayesian, and MP, respectively. For further hypothesis testing, the optimal tree was selected on the basis of likelihood score.

2.6. Hypothesis testing

To further test the hypothesis of two colonization events of South America from Africa by *Mabuya*, we performed parametric bootstrapping (Huelsenbeck et al., 1996; Swofford et al., 1996) on the mtDNA partition, the nucDNA partition, and the combined dataset. For each of the above datasets, PAUP* (Swofford, 2002) was used to search for the best tree given the constraint of a single Neotropical colonization by *Mabuya* (null hypothesis). This constraint tree, along with the model and parameters estimated for the specific partition, was used to simulate 1000 datasets in Seq-Gen 1.2.7 (Rambaut and Grassly, 1997). The MP length differences between the constraint tree and the optimal tree for each simulated dataset were used to generate the expected distribution of length differences under the null hypothesis. For the mtDNA, nucDNA, and the combined datasets, the observed length difference from each original dataset was compared to the null distribution of tree length differences from the simulated datasets, and the null hypothesis was rejected if the observed difference was greater than 95% of the expected differences (i.e., at a probability level of $p = 0.05$).

3. Results

3.1. Patterns of variation

Uncorrected pairwise sequence divergence for each gene across all taxa, within *Mabuya*, and within South American *Mabuya* are shown in Table 1. There is a great variation in divergence levels across the gene regions used in this study, with different genes bringing resolution to different parts of the tree. It should be noted that the MYH2 gene has a large insertion (~500 bp) in all South American *Mabuya*, accounting for the large disparity in divergence between *Mabuya* and South American *Mabuya*. Mean base composition for each gene region is also shown in Table 1.

3.2. Individual and partitioned analyses

The most appropriate model for each gene from all four alignment methods was found using the AIC in Modeltest, and is shown in Table 2. The number of total aligned bases, as well as the resulting $-\ln L$ score for the Bayesian analysis of each gene region is shown in Table 3.

3.3. Combined analyses

The single gene trees and all combinations of data recover many identical clades with moderate-to-strong support (0.9–1.0 PP). We do find moderate-to-strong conflict in the ClustalX alignment of the Gapdh gene and the placement of four taxa in the enolase gene tree, but in both the

Table 1
Maximum uncorrected pairwise sequence divergence is listed across various taxonomic levels for each gene region used in this study

	C-mos	16S	Enol	12S	Gapdh	cyt*b*	Myh2
All taxa (%)	11.1	14.8	18.4	22.1	23.4	24	33
Mabuya (%)	6.4	13.4	17.8	19.8	15.2	23.2	26.8
South American *Mabuya* (%)	3.2	8.5	14.8	10.9	8.4	17.1	7.6
	0.301	0.320	0.239	0.356	0.232	0.283	0.283
Mean base composition (A, C, G, T)	0.193	0.247	0.229	0.259	0.180	0.304	0.223
	0.224	0.210	0.240	0.184	0.235	0.136	0.267
	0.282	0.222	0.292	0.202	0.353	0.276	0.228

Mean base composition for each region with all taxa included is also shown. All computations were based on the POYML implied alignment.

Table 2
Modeltest results (AIC) for each gene from all alignment methods

Alignment	mtDNA			nucDNA			
	16S	12S	cyt *b*	C-mos	Enol	Gapdh	MYH2
POYML	TrN + *I* + *G*	TrN + *I* + *G*	N/A	N/A	HKY + *G*	HKY + *G*	HKY + *G*
POYMP	TrN + *I* + *G*	TrN + *I* + *G*	N/A	N/A	HKY + *G*	K81uf + *G*	K81uf + *G*
ClustalX	TrN + *I* + *G*	TrN + *I* + *G*	N/A	N/A	HKY + *G*	K81uf + *G*	K81uf + *G*
Manual	TrN + *I* + *G*	TrN + *I* + *G*	TVM + *I* + *G*	HYK + *G*	HKY + *G*	K81uf + *G*	K81uf + *G*

Table 3
Results of Bayesian analysis of individual genes as well as mitochondrial DNA (mtDNA), nuclear DNA (nucDNA), ribosomal RNA (rRNA), and intron region (intron) partitions from each of the four alignment methods

	ClustalX			Manual			POYMP			POYML		
	−ln *L*	#bp	IS	−ln *L*	#bp	IS	−ln *L*	#bp	IS	−ln *L*	#bp	IS
C-mos	—	—	—	2979.133	597	132	—	—	—	—	—	—
Cyt *b*	—	—	—	17,897.661	979	465	—	—	—	—	—	—
12S	12,413.843	906	388	11,435.269	889	385	**11,137.689**	1011	383	11,177.986	992	383
16S	6630.408	581	186	6314.007	577	185	**5487.987**	664	175	5495.949	660	167
Enol	2112.189	424	97	**1666.414**	230	86	1759.642	259	90	1755.187	258	93
Gapdh	3477.107	406	173	3142.808	370	157	2972.637	446	141	**2765.314**	460	129
MYH2	11,274.481	1110	461	8734.610	1029	372	7039.896	1147	330	**6807.027**	1127	334
mtDNA	37,625.136	2466	1039	36,176.177	2445	1035	**34,805.47**	2654	1023	34,820.145	2631	1015
NucDNA	21,061.981	2537	863	17,117.99	2226	747	14,960.889	2449	693	**14,525.237**	2442	688
Intron	18,543.189	1940	731	14,181.829	1629	615	11,998.172	1852	561	**11,689.493**	1845	556
rRNA	19,272.445	1487	574	17,857.443	1466	570	**16,679.14**	1675	558	16,701.395	1652	550

The number of aligned base pairs (#bp) is listed, along with the number of parsimony informative sites (IS) for each region with all taxa included. The −ln *L* score is the mode of all Bayesian scores from the analysis after the burn-in trees are discarded. For each gene and partition, the best likelihood score is listed in bold.

intron and the nuclear data partitions we see that both of these issues are resolved, and trees from all four alignment methods show an increase in both resolution and nodal support from the individual gene trees. Therefore, we consider this conflict to be negligible and combined all data for analysis.

Likelihood ratio tests identified the s6g model (modified GTR plus gaps with all parameters estimated from the data) as the best justified model under POY, and this model was used to generate the implied alignment for all subsequent POYML analyses. The implied alignment from the POYML topology (−ln *L* = 54,049.92) consisted of 5073 characters, 1703 of which are parsimony informative. The implied alignment from the POYMP topology (length 11,483) consisted of 5103 aligned bases, and 1717 parsimony informative sites. The ClustalX alignment resulted in 5003 characters with 1902 parsimony informative sites, while the manual alignment consisted of 4671 total characters and 1782 parsimony informative sites. The models and parameters chosen for each partition in all four alignments are shown in Table 2.

All Bayesian analyses reached stationarity before 100,000 generations, and after discarding the first 100 trees (burn-in) from each of the three separate runs per alignment, the −ln *L* scores for all the remaining trees were combined and the mode computed to obtain the final score for that alignment. Results from the mixed model Bayesian analyses are shown in Table 4. The POYML mode −ln *L* score was 50,496.073, with the POYMP alignment mode at 51,611.928, the ClustalX alignment is −61,489.630, and the manual alignment mode is 55,329.945. Similarly, when the GTR + *I* + *G* model is used for all data partitions, the mode −ln *L* score after burn-in for the POYML alignment is 51,014.655, the POYMP alignment mode is 51,554.225, while the mode for the ClustalX alignment is 61,548.268, and the manual alignment mode is 55,858.397.

Table 4
Results of Bayesian analysis of combined datasets for all alignment methods

	ClustalX	Manual	POYMP	POYML
Mixed model Bayesian	61,489.630	55,329.945	51,611.928	**50,496.073**
Single model Bayesian	61,548.268	55,858.397	51,554.225	**51,014.655**
MP tree length	15,154	20,341	**11,483**	11,702

Each aligned dataset was analyzed in a partitioned Bayesian analysis with models for each gene as listed in Table 2, and parameters allowed to vary across partitions (mixed model Bayesian). A separate Bayesian analysis was also performed for each alignment using a single model (GTR + *I* + *G*) across all partitions (single model Bayesian). And finally, each alignment was subjected to a maximum parsimony analysis (MP). The −ln *L* scores for the mixed model Bayesian and single model Bayesian results are the mode of all Bayesian scores from the analysis after the burn-in trees are discarded. The MP score is the resulting tree length. For each type of analysis, the best tree score is listed in bold; see text for details on alignment and tree reconstruction.

The POYML combined tree (Fig. 2) recovers a monophyletic Asian *Mabuya* (1.0 PP) which is sister group (1.0 PP) to a monophyletic Afro-malagasy *Mabuya* (including Fernando de Noronha endemic *M. atlantica*; 1.0 PP) + a monophyletic South American *Mabuya* (1.0 PP). Malagasy *Mabuya* are monophyletic (1.0 PP) and sister to one clade

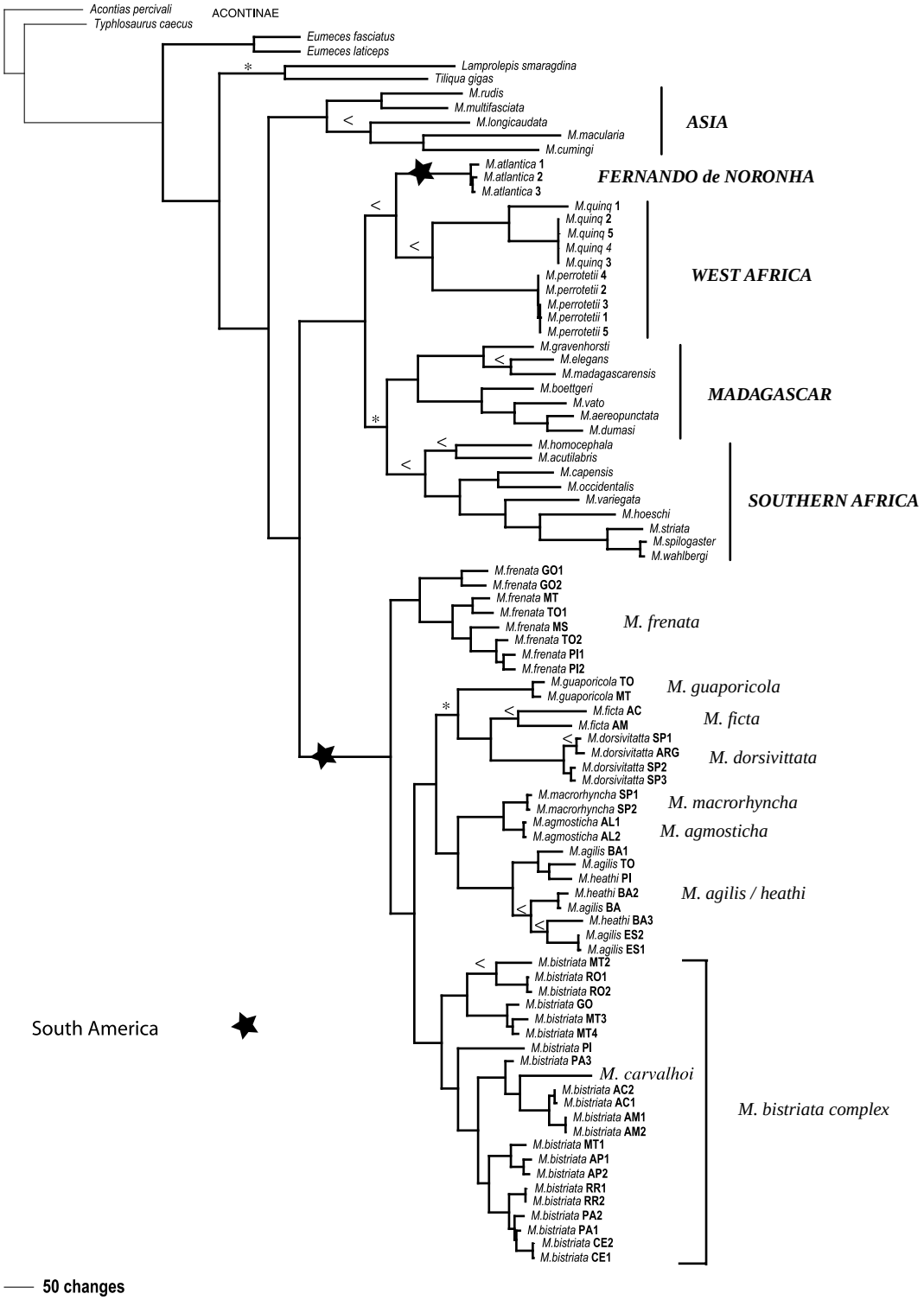

Fig. 2. Partitioned Bayesian analysis of the POY maximum likelihood (POYML) combined data implied alignment, 50% majority rule consensus. Nodes with posterior probabilities below 0.90 are marked <, while 0.90–0.94 are marked *, and all unmarked nodes have values of 0.95–1.0.

of South African taxa, while *M. atlantica* is sister to the remaining African species. Within South American *Mabuya*, multiple species and groups are well supported (PP $\geqslant$ 0.95) including *M. frenata*, *M. guaporicola*, *M. dorsivittata*, *M. macrorhyncha*, *M. agmosticha*, *M. macrorhyncha* + *M. agmosticha*, *M. ficta* + *M. dorsivittata*, *M. bistriata* + *M. carvalhoi*, and *M. agilis* + *M. heathi*. The *M. agilis*/*heathi* clade is sister to *M. macrorhyncha* + *M. agmosticha* (1.0 PP), and this group is sister to a clade composed of *M. dorsivittata*, *M. ficta*, and *M. guaporicola*. *M. frenata* is sister to the clade of seven species listed above + the *M. bistriata*/*carvalhoi* assemblage.

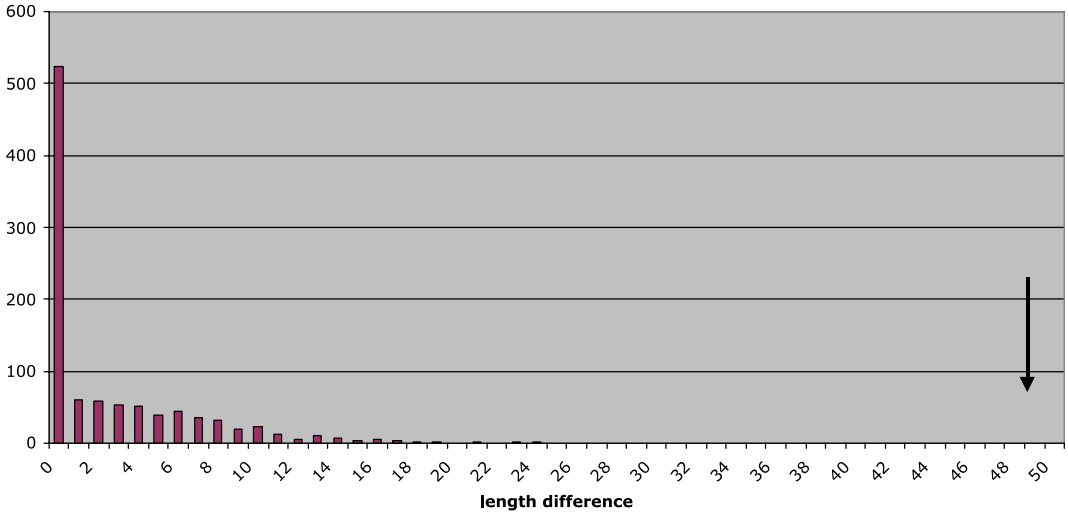

Fig. 3. Parametric bootstrap distribution of tree length differences for all data combined between constrained and optimal topologies constructed from 1000 simulated datasets. Empirical length difference is denoted with the arrow.

The POYMP topology differs from the POYML tree by placing *M. frenata* as sister to the *M. bistriata*/*carvalhoi* group (0.90 PP), and *M. atlantica* is sister to *Mabuya quinquintaeniata* (0.89 PP). *M. ficta* is not monophyletic, and *M. ficta* + *M. dorsivittata* is sister to all remaining *Mabuya* except *M. guaporicola* (0.78 PP) which is basal within the genus (0.81 PP), although the last two results are poorly supported. Outside of this the POYMP and POYML topologies are identical.

The ClustalX and the manual alignment trees are identical with the exception of the relationships within the *M. bistriata* clade, placement of *Mabuya acutilabris* and *Mabuya homocephala*, and the sister group of *M. quinquintaeniata* + *Mabuya perrotetii*; although none of the above relationships are well supported in either tree. Both the ClustalX and manual alignment trees strongly support the monophyly of tropical Asian *Mabuya* (PP = 1.0), South American *Mabuya* (PP > 0.95), and Afro-Malagasy *Mabuya* (plus *M. atlantica*) (PP > 0.95), and both recover the same relationships among these groups [Asia (South America + Afro-Malagasy)] with PP = 1.0. However, both also recover *M. carvalhoi* as unresolved within the *M. bistriata* clade, find a nonmonophyletic *M. ficta*, and do not resolve relationships among the three main clades of Afro-Malagasy *Mabuya*, or the four main clades of South American *Mabuya*.

Because the POYML topology is the most likely, we select this as our optimal topology for further hypothesis testing and as our working hypothesis for relationships within *Mabuya*.

3.4. Hypothesis testing

The parametric bootstrap tests significantly reject the null hypothesis of a single colonization of South America for the mtDNA partition, nucDNA partition, and the total combined dataset (with $P < 0.001$ for all three datasets). For the combined data, the difference between the optimal tree and the single colonization tree in the original dataset was 50, whereas the largest difference between trees in the simulated datasets was 24 (Fig. 3). Results for the nucDNA and mtDNA partitions were similar, with a minimum of nine steps between the empirical tree length difference and the simulated tree length difference.

4. Discussion

4.1. Alignment methods

Results from both the individual gene trees and the combined analyses suggest that an alignment generated under DO in POY results in more optimal ML tree topologies relative to those recovered from a ClustalX or manual alignment. The combined data analyses show that likelihood scores improve successively from the ClustalX alignment to the manual, to the POYMP, and finally to the POYML alignment—resulting in the optimal tree. This is typically the same pattern that we see with the individual gene analyses, although with the ribosomal genes the POYMP scores are slightly better than for POYML although the scores from the two methods are quite close. The one exception to this basic pattern is the enolase gene, in which the manual alignment results in the most likely tree score.

These results for likelihood parallel those found for parsimony, in that POY results in more optimal topologies (i.e., better likelihood score) than ClustalX (Ogden and Whiting, 2003; Wheeler, 2003). Intuitively, this is in large part because ClustalX generates an alignment from a single guide tree found under a simple clustering algorithm, while POY uses heuristic methods to evaluate large number of guide trees using more sophisticated criteria before selecting an optimal alignment (Wheeler, 2003). The most troubling discovery relating to Clustal is that a manual adjustment of a ClustalX alignment in every case produces

a more optimal topology than the original ClustalX alignment. In the case of the enolase gene in which the manual alignment produced the most optimal tree, there was a single divergent sequence that ClustalX had a difficult time aligning and adjustment of parameters may have greatly improved this alignment. POYML and POYMP had a similar problem with the Enolase region to a lesser degree. The disparity in resulting likelihood scores may be due to the lack of parameter investigation in ClustalX, but whatever the cause, we suggest that simple ClustalX alignments should not be used without careful scrutiny.

It is important to note that the POYML implied alignment produces the most optimal likelihood topologies and likewise the POYMP alignment produces the most parsimonious topologies (see Table 4). This result was theoretically expected, but to our knowledge is the first time to be demonstrated for likelihood analyses. This difference emphasizes the importance of consistent use of an optimality criterion throughout alignment and tree reconstruction (Wheeler et al., 2005). Similarly, the improvement in likelihood score seen with the addition of more complex models and partitioned analyses is not as great as the difference between likelihood scores resulting from different alignments. As such, these results join those found previously (Morrison and Ellis, 1997; Ogden and Whiting, 2003; Ogden et al., 2005; Wheeler, 2003) in emphasizing the importance of alignment in the tree reconstruction process, and the disproportionate lack of emphasis currently placed on alignment.

We recognize that these results should be tested further with other data sets as there are many factors that should be considered. One of the most important issues not dealt with in this study is the treatment of gaps. All alignment programs must deal with gaps, while most commonly used tree reconstruction methods simply treat gaps as missing data, although this is not a necessity of the method but just a lack of implementation (Felsenstein, 2003). All of the Bayesian analyses done here treat gaps as missing data, while the MP analyses of the combined datasets were performed with gaps treated as a fifth state. We recognize that the effect of gap treatment is still a large question mark, but that is beyond the objectives of this study.

This study was not meant to be a comprehensive look at all alignment programs, and was primarily intended to bring added attention to the subject of alignment and the influence it plays in phylogenetic tree reconstruction. We did not include secondary structure alignment, nor did we evaluate different parameter settings in ClustalX or POYMP, because our objective was not to do a full comparison of the methods, but rather to contrast the methods most commonly used today in multi-gene phylogenetic studies. Some will argue that by not cutting out the "hypervariable" or "unalignable" regions we biased our analyses toward POY. There is no set standard for "unalignable" data, but we think most will agree that if a dataset does not contain variable regions, then any basic alignment method including something as simple as a similarity criterion (such as is used in the program Sequencher) would give a reasonable alignment. The purpose of rigorous alignment programs is to deal with the regions that are difficult to align, therefore, we feel justified in using these regions to compare alignment methods. Another objection may be that tree score is not an appropriate measure of alignment quality. We agree that the alignment that produces the most parsimonious or the most likely tree may not be the true alignment or the way the sequences evolved historically. Nevertheless, we consider that just as with tree reconstruction we cannot guarantee the "true" tree, we must choose an optimality criterion and find both the tree and the alignment, that best satisfy that criterion.

It should be recognized that POY is not intended to be a multiple alignment program in the traditional sense, and that the implied alignment represents the optimization of sequence data for a given (optimal) topology. Wheeler (2003) recommends that the topology from POY be favored rather than using the implied alignment for further tree reconstruction, since this will result in the globally most optimal solution and be logically consistent. Nonetheless, POY provides an estimate for a likelihood alignment, and this likelihood alignment produces likelihood topologies that are more optimal than those generated via standard multiple alignment methods. In combination with similar results found under other optimality criterion, these results support the conclusion that consistent use of an optimality criterion, models, and parameters in both alignment and tree reconstruction leads to the most optimal trees for the given criteria, and appears to be more influential on final tree score than the choice of specific models or partitioning of models in tree reconstruction.

4.2. South American Mabuya

While the topologies resulting from the combined analysis of the four different alignment methods differ in some details, there are many relationships and groups present in all four trees, including: the monophyly of tropical Asian, Afro-malagasy (including Fernando de Noronha's *M. atlantica*), and South American *Mabuya*, with the Asian clade sister to the other two groups (as reported by Mausfeld et al., 2002). It should be noted that our sampling was not designed to test the monophyly of the Mausfeld et al. newly proposed genera, and more complete sampling of the border regions between Africa and Asia (similar to Carranza and Arnold, 2003) will be necessary to resolve this taxonomic issue. Within the Afro-Malagasy clade, Malagasy species are recovered as monophyletic by all alignments, in contrast to the Mausfeld et al. (2002) trees that suggested paraphyly, although *Mabuya comorensis* was not included in our study. Also recovered in all alignments are monophyletic *M. atlantica*, *M. quinquitaeniata*, *M. perrotetii*, and a clade containing all remaining African species except *M. homocephala*, and *M. acutilabris*. Relationships among the above-mentioned clades are unresolved or poorly supported by all alignment methods and therefore remain tentative.

Relationships within South American *Mabuya* that are present and well supported in all alignment trees include monophyly of *M. guaporicola*, *M. dorsivittata*, *M. macrorhyncha*, *M. agmosticha*, *M. macrorhyncha* + *M. agmosticha*, *M. frenata*, *M. bistriata* + *M. carvalhoi*, *M. agilis* + *M. heathi*, and *M. agilis* + *M. heathi* + (*M. macrorhyncha* + *M. agmosticha*). The sister relationship of *M. macrorhyncha* and *M. agmosticha* is found in every individual gene tree as well as in all combined trees, confirming earlier evidence provided for this grouping (Rodrigues, 2000). Also well supported in all combined trees is the sister relationship between the above stated two species and a clade composed of *M. agilis*, and *M. heathi*. Both *M. agilis* and *M. heathi* have been previously recognized as problematic groups (Rodrigues, 2000). Our results confirm the nonmonophyly of these species and indicate that at least two entities may be involved in the *agilis*/*heathi* complex, the first is mainly associated with Cerrados (dry savanna); the second with the Caatingas and habitats near the Eastern Brazilian coast. More extensive geographic sampling of the *agilis*/*heathi* complex will need to be undertaken before species definitions can be confidently assigned.

Another clade within *Mabuya* that will require further attention in the future is the assemblage we have labeled as the *M. bistriata* complex. Our results indicate that this clade is composed of a large species complex consisting of a minimum of six entities, to which either the name *M. bistriata* (sensu Rodrigues, 2000) or *M. nigropunctata* (Ávila-Pires, 1995) has been applied. One clade assembles specimens from the Central Brazilian Cerrados and those from their contact areas with Amazonia in states of Mato Grosso and Rondonia. Its sister clade has a specimen from Uruçuí-Una in the Cerrados of the state of Piauí (*M. bistriata*-PI; Fig. 2) as the sister of all other *bistriata* complex. Within the latter, two large and heterogeneous groups are present; the first includes specimens from isolated patches of Atlantic Forest in the Caatingas, specimens from Eastern Amazonia and one from Aripuanã (state of Mato Grosso), confirming the sympatric occurrence of two species in that locality. Nested within the second clade is *M. carvalhoi*, a recently described and distinct morphological species. The specific relationship of *M. carvalhoi* within the *M. bistriata complex* does vary between different alignments, but the species is always placed within this group (PP $\geqslant$ 0.95). In Fig. 2, *M. carvalhoi* is sister to a group of Western Amazonia *M. bistriata* and both are sisters to specimens from Santarém, state of Pará. These results reveal the presence of multiple monophyletic groups within the *M. bistriata complex*, and the need for further detailed morphological and molecular work and extensive sampling in order to properly diagnose these taxonomic entities. The morphological distinctiveness of *M. carvalhoi* also reveals that additional sampling may help to confirm its placement.

Other relationships within South American *Mabuya* are more tentative, including the monophyly of *M. ficta*, and the placement of *M. guaporicola*. The different alignment methods support different placements of *M. guaporicola*, although its placement is not well supported in any tree. The POYMP alignment places *M. guaporicola* as sister to all remaining South American *Mabuya*, while the other three alignment methods recover *M. guaporicola* in a clade with *M. ficta* and *M. dorsivittata*. This clade shows consistently long branch lengths, which may be the source of error associated with not only the placement of *M. guaporicola* but also the monophyly of *M. ficta*. Further geographic sampling within both *M. ficta* and *M. guaporicola* may help to shorten these long branches and stabilize placement of these species.

Mabuya frenata is well supported as monophyletic in all alignments, but its placement does vary. Our most likely topology (POYML; Fig. 2) does place *M. frenata* as sister to all the remaining *Mabuya* species, so we accept this relationship as the best working hypothesis for the group.

4.3. Multiple colonizations of South America

The parametric bootstrap is thought to be the most rigorous and appropriate test for an a priori hypothesis of relationships (Huelsenbeck et al., 1996; Swofford et al., 1996). Our significant rejection of the single colonization hypothesis by both the mtDNA and nucDNA partitions as well as the combined dataset further corroborate the recently reported molecular results (Carranza and Arnold, 2003; Mausfeld et al., 2002), and the long suspected morphological evidence (Greer et al., 2000; Horton, 1973; Travassos, 1948). Fernando de Noronha is currently inhabited by only three lizard species, *M. atlantica*, *Hemidactylus mabouia*, and *Tupinambis merianae*, of which the latter two are known to have been introduced to the islands in the second half of the 20th century (Carranza and Arnold, in press). While our results strongly reject the hypothesis that *M. atlantica* could have been introduced from Brazil, the archipelago of Fernando de Noronha has historically been visited by sailors and traders, and the possibility of human introduction from Africa cannot be addressed by these analyses. However, the explorer Americus Vespucius visited Fernando de Noronha in August of 1503, and the following is an excerpt from a letter to Piero Soderini detailing his fourth voyage to America: "This island we found uninhabited. It had plenty of fresh-water, and an abundance of trees filled with countless numbers of land and marine birds, which were so simple, that they suffered themselves to be taken with the hand. We took so many of them that we loaded a boat with them. We saw no other animals, except some very large rats and lizards with two tails, and some snakes (Lester, 1886). The "large rats" mentioned by Vespucius were the now extinct *Noronhomys vespucii*, while the "snakes" were most likely *Amphisbaena ridleyi*, and the "lizards with two tails" were probably specimens of *M. atlantica* with aberrantly regenerated tails. This visit of Vespucius was just shortly after Columbus' discovery of the Americas, and provides strong evidence that *M. atlantica* was present on Fernando de Noronha before the arrival of humans, thereby adding further support to the hypothesis of two independent colonization events via rafting to the New World.

The transmarine rafting journey of both the ancestor of South American *Mabuya* and the ancestor of *M. atlantica* involved a journey of greater than 3000 km. The colonization of South America and Fernando de Noronha from Africa is but one of several documented cases of long distance marine dispersal of lizards, with distances of up to 6000 km being traveled (Arnold, 1976, 2000; Carranza et al., 2000; de Queiroz, 2005; Nicholson et al., 2005; Raxworthy et al., 2002). For several decades, vicariance has been seen as the most probable explanation for historical biogeography. The multiple colonizations of South America found here are part of a recent resurgence in support for oceanic dispersal (see de Queiroz, 2005 for a review of the topic).

Acknowledgments

We thank the following people for valuable help in the field: F. Curcio, J.M. Martins, G. Skuk, V. Xavier, V. Verdade, D. Pavan, E.M.X. Freire, M.J.J. Silva, R. Amaro-Ghilardi, N.J. da Silva Jr., C.M. Carvalho, H. Zaher, D. Borges, and M. Barbosa. We thank the following people and institutions for tissue samples: M. Kearney and A. Reseter of the Field Museum of Natural History, R. Nussbaum and G. Schneider of the University of Michigan Museum of Zoology, L.J. Vitt and J.P. Caldwell of the Sam Noble Oklahoma Museum of Natural History, F. Sheldon and D. Dittman of the Louisiana State University Museum of Natural Science, A. Bauer of Villanova University, M. Morando and L. Avila of the Centro Nacional Patagónico, Argentina. We thank the conservation, environmental, and permit issuing authorities of the Republic of South Africa (Northern Cape Province, Western Cape Province, Eastern Cape Province, Mpumalanga, Northern Province, and KwaZulu/Natal), and the Republic of Namibia for permitting to conduct research and collect specimens. We would like to thank Gonzalo Giribet, Salvador Carranza, Ward Wheeler, and two anonymous reviewers for suggestions and comments that greatly improved this manuscript. This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a National Science Foundation graduate research fellowship to A.S.W., and an NSF doctoral dissertation improvement award (DEB 02-00362) to J.W.S. and A.S.W.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2005.11.011.

References

Almeida, F.F.M., 2000. The Fernando de Noronha archipelago. In: Schobbenhaus, C., Campos, D.A., Queiroz, E.T., Winge, M., Berbert-Born, M. (Eds.), Sítios Geológicos e Paleontológicos do Brasil. <http://www.unb.br/ig/sigep/witio066/sitio066.htm>.

Arnold, E.N., 1976. Fossil reptiles from Aldabra Atoll, Indian Ocean. Bull. Brit. Mus. (Nat. Hist.) 29, 83–116.

Arnold, E.N., 2000. Using fossils and phylogenies to understand evolution of reptile communities on islands. Bonner Zoologische Beiträge 46, 309–323.

Ávila-Pires, T.C.S., 1995. Lizards of Brazilian Amazonia (Reptilia: Squamata). Zool. Verhand. Leiden 299, 1–706.

Bauer, A.M., 2003. On the identity of *Lacerta punctata* Linnaeus, 1758, the type species of the genus *Euprepis* Wagler, 1830, and the generic assignment of Afro-Malagasy skinks. Afr. J. Herpetol. 52, 1–7.

Blackburn, D.G., Vitt, L.J., 2002. Specializations of the chorioallantoic placent in the Brazilian scincid lizard, *Mabuya heathi*: a new placental morphotype. J. Morph. 254, 121–131.

Carranza, S., Arnold, E.N., 2003. Investigating the origin of transoceanic distributions: mtDNA shows *Mabuya* lizards (Reptilia, Scincidae) crossed the Atlantic twice. Syst. Biodivers. 1, 275–282.

Carranza, S., Arnold, E.N., in press. Systematics, biogeography, and evolution of *Hemidactylus* geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. Mol. Phylogenet. Evol. in press, doi:10.1016/j.mpev.2005.07.012.

Carranza, S., Arnold, E.N., Mateo, J.A., López-Jurado, L.-F., 2000. Long-distance colonization and radiation in gekkonid lizards, *Tarentola* (Reptilia: Gekkonidae), revealed by mitochondrial DNA sequences. Proc. R. Soc. B 267, 637–649.

de Queiroz, A., 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20, 68–73.

Dolman, G., Phillips, B., 2004. Single copy nuclear DNA markers characterized for comparative phylogeography in Australian Wet Tropics rainforest skinks. Mol. Ecol. Notes 4, 185–187.

Felsenstein, J., 2003. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

Giribet, G., 2005. Generating implied alignments under direct optimization using POY. Cladistics 21, 396–402.

Greer, A.E., 1970. A subfamilial classification of scincid lizards. Bull. Mus. Comp. Zool. 139, 151–183.

Greer, A.E., Arnold, C., Arnold, E.N., 2000. The systematic significance of the number of presacral vertebrae in the scincid lizard genus *Mabuya*. Amphibia-Reptilia 21, 121–126.

Horton, D., 1973. Evolution in the genus *Mabuya* (Lacertilia, Scincidae). Ph.D. dissertation, University of New England.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Hillis, D.M., Jones, R., 1996. Parametric bootstrapping in molecular phylogenetics: applications and performance. In: Ferraris, J.D., Palumbi, S.R. (Eds.), Molecular Zoology: Advances, Strategies, and Protocols. Wiley-Liss, New York, NY, pp. 19–42.

Lester, C.E., 1886. The Life and Voyages of Discovery of Americo Vespucius. H. Mansfield, New Haven.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E., O'Brien, S.J., 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat. Genet. 15, 47–56.

Lyons, L.A., Kehler, J.S., O'Brien, S.J., 1999. Development of comparative anchor tagged sequences (CATS) for canine genome mapping. J. Hered. 90, 15–26.

Maddison, D.R., Maddison, W.P., 2000. MacClade, 4.0. Sinauer Associates, Sunderland, MA.

Mausfeld, P., Schmitz, A., Bohme, W., Misof, B., Vrcibradic, D., Rocha, C.F.D., 2002. Phylogenetic affinities of *Mabuya atlantica* Schmidt, 1945, endemic to the Atlantic Ocean archipelago of Fernando de Noronha (Brazil): necessity of partitioning the genus *Mabuya* Fitzinger, 1826 (Scincidae: Lygosominae). Zool. Anz. 241, 281–293.

Méndez de la Cruz, F.R., Villagran Santa Cruz, M., Andrews, R.M., 1998. Evolution of viviparity in the lizard genus *Sceloporus*. Herpetologica 54, 521–532.

Morrison, D.A., Ellis, J.T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14, 428–441.

Nicholson, K.E., Glor, R.E., Kolbe, J.J., Larson, A., Hedges, S.B., Losos, J.B., 2005. Mainland colonization by island lizards. J. Biogeogr. 32, 1–10.

Ogden, T.H., Whiting, M.F., 2003. The problem with the "Paleoptera problem: sense and sensitivity. Cladistics 19, 379–448.

Ogden, T.H., Whiting, M.F., Wheeler, W.C., 2005. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: a reply to Kjer. Cladistics 21, 295–302.

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

Rambaut, A., Grassly, N.C., 1997. Seq-Gen: and application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 235–238.

Raxworthy, C.J., Forstner, M.R.J., Nussbaum, R.A., 2002. *Chameleon* radiation by oceanic dispersal. Nature 415, 784–787.

Rodrigues, M.T., 2000. A new species of *Mabuya* (Squamata: Scincidae) from the semiarid Caatingas of northeastern Brazil. Papeis Avulsos de Zoologia 41, 313–328. Museu de Zoologia da Universidade de Sao Paulo.

Shine, R., 1985. The evolution of viviparity in reptiles: an ecological analysis. In: Gans, C., Billett, F. (Eds.), Biology of the Reptilia. John Wiley and Sons, New York, pp. 605–694.

Swofford, D.L., 2002. PAUP* Phylogenetic Analysis using Parsimony (*and other methods), 4.0. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer Associates, Sunderland, MA, pp. 407–514.

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

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Travassos, H., 1948. Nota sobre a "*Mabuya*" da Ilha Fernando de Noronha (Squamata: Scincidae). Rev. Brasil. Biol. 8, 201–208.

Wenzel, J.W., Sidall, M.E., 1999. Noise. Cladistics 15, 51–64.

Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–10.

Wheeler, W.C., 2003. Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. Cladistics 19, 261–268.

Wheeler, W.C., Gladstein, D.S., De Laet, J., 2003. POY: Phylogeny Reconstruction Via Direct Optimization of DNA and Other Data, 3.0.11. American Museum of Natural History, New York, NY.

Wheeler, W.C., Aagesen, L., Arango, C.P., Faivovich, J., Grant, T., D'Haese, C., Janies, D., Smith, W.L., Varon, A., Giribet, G., 2005. Dynamic Homology and Phylogenetic Systematics: A Unified Approach using POY. American Museum of Natural History, New York.

Whiting, A.S., Bauer, A.M., Sites, J.W.J., 2003. Phylogenetic relationships and limb loss in southern African scincine lizards (Squamata: Scincidae). Mol. Phylogenet. Evol. 29, 582–598.

Whiting, A.S., Sites, J.W.J., Bauer, A.M., 2004. Molecular phylogenetics of Malagasy skinks (Squamata: Scincidae). Afr. J. Herpetol. 53, 135–146.