

2012

## PhyloMarker—A Tool for Mining Phylogenetic Markers Through Genome Comparison: Application of the Mouse Lemur (Genus *Microcebus*) Phylogeny

Runhua Lei

Thaine W. Rowley

Lifeng Zhu

Carolyn A. Bailey

Shannon E. Engberg

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unomaha.edu/biofacpub>

 Part of the [Biology Commons](#)

---

**Authors**

Runhua Lei, Thaine W. Rowley, Lifeng Zhu, Carolyn A. Bailey, Shannon E. Engberg, Mindy L. Wood, Mary C. Christman, George H. Perry, Edward E. Louis Jr., and Guoqing Lu

---

SOFTWARE REVIEW

**OPEN ACCESS**Full open access to this and thousands of other papers at <http://www.la-press.com>.

## PhyloMarker—A Tool for Mining Phylogenetic Markers Through Genome Comparison: Application of the Mouse Lemur (Genus *Microcebus*) Phylogeny

Runhua Lei<sup>1,#</sup>, Thaine W. Rowley<sup>2,#</sup>, Lifeng Zhu<sup>3</sup>, Carolyn A. Bailey<sup>1</sup>, Shannon E. Engberg<sup>1</sup>, Mindy L. Wood<sup>1</sup>, Mary C. Christman<sup>2</sup>, George H. Perry<sup>4</sup>, Edward E. Louis Jr<sup>1</sup> and Guoqing Lu<sup>2</sup><sup>1</sup>Center for Conservation and Research, Omaha's Henry Doorly Zoo and Aquarium, Omaha, NE 68107, United States.<sup>2</sup>Department of Biology, University of Nebraska at Omaha, NE 68135, United States. <sup>3</sup>College of Life Science, Nanjing Normal University, Jiangsu 210097, China. <sup>4</sup>Department of Anthropology, Pennsylvania State University, PA 16802, United States. #Authors contributed equally to this work.Corresponding authors email: [edlo@omahazoo.com](mailto:edlo@omahazoo.com); [glu3@unomaha.edu](mailto:glu3@unomaha.edu)

**Abstract:** Molecular phylogeny is a fundamental tool to understanding the evolution of all life forms. One common issue faced by molecular phylogeny is the lack of sufficient molecular markers. Here, we present PhyloMarker, a phylogenomic tool designed to find nuclear gene markers for the inference of phylogeny through multiple genome comparison. Around 800 candidate markers were identified by PhyloMarker through comparison of partial genomes of *Microcebus* and *Otolemur*. In experimental tests of 20 randomly selected markers, nine markers were successfully amplified by PCR and directly sequenced in all 17 nominal *Microcebus* species. Phylogenetic analyses of the sequence data obtained for 17 taxa and nine markers confirmed the distinct lineage inferred from previous mtDNA data. PhyloMarker has also been used by other projects including the herons (Ardeidae, Aves) phylogeny and the Wood mice (Muridae, Mammalia) phylogeny. All source code and sample data are made available at <http://bioinfo-srv1.awh.unomaha.edu/phylomarker/>.

**Keywords:** phyloinformatics, single copy gene markers, genome comparison, *Microcebus* phylogeny

*Evolutionary Bioinformatics* 2012:8 423–435

doi: [10.4137/EBO.S9953](https://doi.org/10.4137/EBO.S9953)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



## Introduction

Assembling the tree of life is an ultimate goal in biology. Utilizing many nuclear gene markers that are distributed throughout different chromosomes is one of the fundamental ways to resolve incongruence in large-scale phylogenies.<sup>1</sup> The inclusion of additional characters from multiple independent genes could reduce sampling errors and systematic biases while reinforcing phylogenetic signals. However, only a limited number of nuclear markers are available for the analysis of deep phylogeny, especially for non-model organisms.<sup>2</sup> Moreover, one of the future challenges is to develop optimized procedures for the detection and selection of orthologous genes with low levels of saturation.<sup>3</sup>

To address the above issues, Li et al<sup>4</sup> initiated a phylogenomic approach that systematically compares genomes to identify candidate nuclear gene markers and presented a case study in constructing a fish tree of life. This initial effort was focused on the large exon markers that are suitable for the analysis of deep phylogeny. However, other genomic markers such as introns and exon-primed intron-crossing markers are more appropriate for small-scale phylogeny and population genetics.<sup>5,6</sup> We developed PhyloMarker, a phylogenomic tool that can be easily used by biologists to find both intron and exon markers through genome comparison. With the advent of next generation sequencing technologies, sequencing subgenomic regions or transcriptomes is becoming common practice.<sup>7</sup> PhyloMarker can take advantage of such large genomic databases for mining phylogenetic markers. Here, we present the conceptual algorithm of PhyloMarker and introduce its implementation and usage. Additionally, we demonstrate its utility with a case study that evaluates the phylogeny of the nominal mouse lemur species, genus *Microcebus*, along with summarizing the results for several vertebrate groups.

## Molecular phylogeny of mouse lemurs

With advances in molecular technology and improvements of analytical tools, the number of cryptic species described in the last two decades has proliferated in many taxonomic groups (eg, amphibian and reptiles,<sup>8,9</sup> bats,<sup>10</sup> fishes,<sup>11</sup> tenrecs<sup>12</sup>). This expansion is especially true for diversity within lemur genera as the taxonomy has been dramatically revised.<sup>13–15</sup>

Found throughout Madagascar, including regions with substantial anthropogenic changes,<sup>16</sup> the mouse lemur, genus *Microcebus*, is considered the smallest, most abundant and widespread lemur.<sup>17</sup> Initially classified as a single species, *M. murinus* Miller was divided into two regionally defined species, *M. murinus* in the west and *M. rufus* in the east.<sup>18</sup> In the last two decades, the genus *Microcebus* has expanded incredibly from two to 18 species based primarily on mitochondrial DNA (mtDNA) fragment sequences and morphological data.<sup>13,14,19–25</sup>

Tattersall<sup>26</sup> questioned whether the increase in lemur species was taxonomic inflation and suggested that more data was needed to delimit or validate the legitimacy of these new taxa. Based on the intra- and interspecific genetic distance estimation and Population Aggregation Analysis simulation, Markolf et al<sup>27</sup> argued that several recently named lemur taxa should be re-evaluated with additional data from multiple nuclear and sex-specific genetic loci. Heckman et al<sup>28</sup> found that comparisons between mtDNA and nuclear DNA (nucDNA) data were fundamentally congruent, but incomplete lineage sorting and low mutation rates of nucDNA data may limit the phylogenetic resolution. Weisrock et al<sup>29</sup> verified the high lineage diversity in *Microcebus* based on nucDNA and mtDNA sequence data, but raised questions on the validity of *M. mambiratra* and suggested that additional cryptic species were included within the current distributions of *M. murinus*, *M. myoxinus* and *M. simmonsii*.

Until recently, only a limited number of nuclear loci have been utilized in phylogenetic studies of *Microcebus*, all of which were previously developed from other taxonomic groups.<sup>28,29</sup> Although Horvath et al<sup>30</sup> has developed a phylogenomic “toolkit” for lemurs consisting of seven previously utilized markers and 11 novel loci, this toolkit has not been utilized extensively with lemurs. The amount of genomic sequence available generated from next generation sequencing has increased significantly, which has promoted the use of multilocus phylogenetic approaches to resolve the phylogenies of amphibians, birds, fishes, and primates.<sup>4,6,31–33</sup> Using genomic data available in the Ensembl database, we utilized PhyloMarker to extract candidate nuclear phylogenetic markers shared between the two partial genomes of the Gray mouse lemur, *Microcebus murinus*, and

the Northern greater galago, *Otolemur garnettii*, generating nucDNA sequence from nine exons to evaluate the phylogenetic relationships among 17 nominal *Microcebus* species.

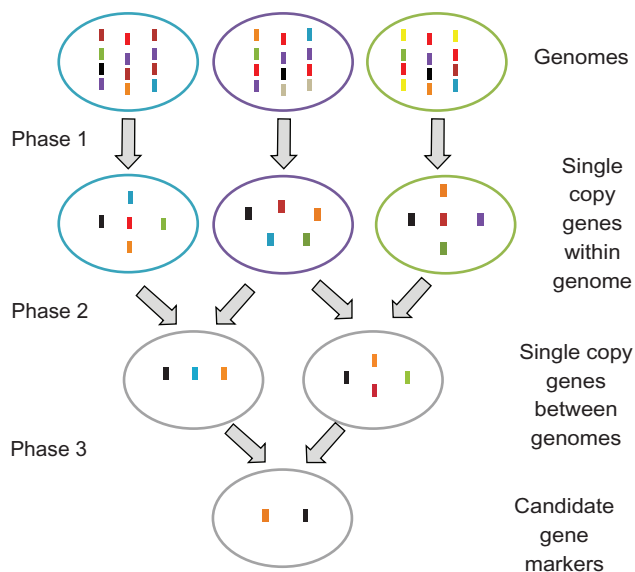
## Methods

### PhyloMarker

PhyloMarker was designed to find single copy nuclear gene regions that are relatively conserved across a variety of species. The process consists of three steps: (1) exon or intron sequence extraction; (2) intra-genome comparison; and (3) inter-genome comparison (Fig. 1).

In Step 1, extraction of exon or intron sequences from GenBank input files is performed using the BioPerl package. The exon positions are available for each gene entry, which can be used to extract specific sequences. The intron positions need to be calculated from the location of consecutive exons. Extracted sequences that are longer than the user-defined sequence length and have less than 20% ambiguous nucleotides are written to a FASTA file.

In Step 2, each FASTA file acts as both the query database and the subject database with the sequences being compared using the BLAST algorithm with minor modification to compute the alignment coverage.<sup>34</sup> Sequences with coverage and identity values below the user-defined thresholds are extracted and written to a file that consists of single copy exons or introns for each given genome.



**Figure 1.** Conceptual design of PhyloMarker.

In Step 3, the single copy nuclear gene sequences predicted in the first or reference genome are compared to the single copy nuclear sequences from each of the subsequent genomes. Those sequences with coverage and identity values above the user-defined thresholds are extracted from the compared genomes. These pairwise single copy “orthologs” are further evaluated, and each candidate is selected only if the exon or intron from the reference genome has an ortholog from each of the other genomes. Finally, the information from the single copy nuclear gene candidates are assembled and formatted into an Excel file.

### Case study in genus *Microcebus*

#### Sample collection

Seventeen tissue samples were collected from Madagascar as detailed in Andriantompohavana et al<sup>22</sup> and Louis et al<sup>13,14</sup> which represent the 17 nominal *Microcebus* species (*M. arnholdi*, *M. sambiranensis*, *M. mairatra*, *M. margotmarshae*, *M. myoxinus*, *M. berthae*, *M. rufus*, *M. lehilahytsara*, *M. mittermeieri*, *M. tavaratra*, *M. jollyae*, *M. bongolavensis*, *M. ravelobensis*, *M. danfossi*, *M. murinus*, *M. griseorufus*). A sample was not available for the recently described *M. gerpi*, and thus, was not included in this study.<sup>25</sup> Total genomic DNA was extracted from the tissue samples using a whole genome amplification kit (WGA; GE Healthcare, USA). Samples extracted using the WGA method, which did not yield a PCR product were re-extracted with the standard PCI protocol (phenol/chloroform isoamyl alcohol).<sup>35</sup>

#### Mining for phylogenetic markers and primer design

The partial genomic sequences of *M. murinus* and *O. garnettii* were retrieved from the Ensembl database (<http://www.ensembl.org/index.html>). Exon and intron sequences with length greater than 800 bp were extracted from the genome database utilizing PhyloMarker (<http://bioinfo-srv1.awh.unomaha.edu/phylomarker/>). PCR and sequencing primers were designed based on the aligned sequences of *M. murinus* and *O. garnettii*. MacVector™ 7.2.2 (Accelrys, San Diego, CA) was utilized to design the primers. Initial default parameters for primer design were as follows: 200–800 bp product size, 17–30 bp primer length, 15%–60% GC content, and the 3-end nucleotide was a G or C clamp.



Gradient- Polymerase Chain Reaction (PCR) design was employed to optimize amplification for each gene in the panel of 17 *Microcebus* species. The thermocycler profile conditions were as follows: 95 °C for 1 min; 34 cycles of 94 °C for 30 sec, 50 °C–60 °C (Table 1) for 45 sec, 72 °C for 45 sec; 72 °C for 10 min. PCR amplifications were carried out in 25 µl reaction volumes containing 2–5 ng of total genomic DNA, 12.5 µM of each primer, 200 µM dNTPs, 10 mM Tris-HCl, 1.5 mM MgCl<sub>2</sub>, 100 mM KCl (pH 8.0) and 0.5 units of BIOLASE™ Taq DNA Polymerase (Bio-line USA Inc., Randolph, MA). The samples were electrophoresed on a 1.2% agarose gel to verify the PCR product and subsequently purified with Exonuclease I and Shrimp Alkaline Phosphatase (SAPEXO).<sup>36</sup> The purified products were cycle-sequenced using a Big Dye terminator sequencing kit (Life Technologies™, Grand Island, NY). The sequences were analyzed by capillary electrophoresis with an Applied Biosystems Prism 3130 genetic analyser. The PCR and sequencing primer suite in Table 1 were used to generate only exon sequence data in this study.

The sequence fragments were aligned to generate consensus sequences using Sequencher 4.10 (Gene Codes Corporation, Ann Arbor, MI). All sequences (accession numbers JX017385–JX017537) have been deposited in GenBank. Single haplotypes for COII and COIII to ND4 (PAST) mtDNA fragments representing each of the 17 recognized *Microcebus* species were analyzed and compared to the novel nuclear gene data set.<sup>13,14,22</sup> Comparative sequence data was mined from three taxa as follows: *O. garnettii* and *M. murinus* from the partial genomes available from Ensembl and the draft genome for the Aye-aye, *Daubentonia madagascariensis*.<sup>37</sup>

### Phylogenetic analysis

Sequences of nine novel nuclear markers from the 17 nominal mouse lemurs, along with two outgroups, the Northern greater galago and Aye-aye, were tested in phylogenetic analyses to evaluate their effectiveness. Sequences were aligned utilizing MAFFT according to the default settings.<sup>38</sup> Initial sequence comparisons and measures of variability were performed using MEGA version 4.0.<sup>39</sup> We analyzed three concatenated data sets as follows: nine nucDNA sequence fragments, COII and PAST mtDNA sequence fragments, and nucDNA and mtDNA sequence data sets.

Phylogenetic trees were estimated from the data sets using Maximum likelihood (ML) and Bayesian inference methods, utilizing the computer packages PAUP\* 4.0b10 and MrBayes v3.1.2.<sup>40–42</sup>

An optimal nucleotide substitution model for each data set was chosen using the Akaike Information Criterion as implemented in Modeltest v3.7.<sup>43</sup> The ML analyses were performed for each data set under the best model with PAUP 4.0b10 software.<sup>40</sup> A heuristic search was performed using random sequence addition (n = 10) and TBR branch swapping. We performed 1000 bootstrap replicates with TBR branch swapping to test the support for nodes in the topology.

Bayesian inference analyses were conducted using MrBayes v3.1.2.<sup>41,42</sup> A Markov Chain Monte Carlo (MCMC) with four simultaneous chains and 5,000,000 generations was performed under the GTR+I+G model selected by MrModeltest v2.2 for the DNA sequence data set.<sup>44</sup> For every 100 generations, the tree with the best likelihood score was saved, resulting in 50,000 trees. Topologies prior to—ln likelihood of equilibrium were discarded as burnin and clade posterior probabilities (PP) were computed from the remaining trees. These trees were condensed into a majority rule consensus tree using PAUP\* 4.0b10.<sup>40</sup> Branch supports were presented as posterior probabilities on the consensus tree.

The coalescent-based Bayesian species tree inference method implemented in the software \*BEAST (an extension of BEAST v1.6.1).<sup>45,46</sup> The software \*BEAST also implements a Bayesian MCMC analysis, and is able to co-estimate species trees and gene trees simultaneously.<sup>45</sup> The input file was formatted with the BEAUti utility included in the software package, using the same partition scheme of the concatenated analysis.

Although \*BEAST does not require the inclusion of outgroups for rooting purposes, Northern greater galago and Aye-aye sequences were incorporated in the analysis. The \*BEAST analysis was conducted utilizing a strict molecular clock model (no loci violated a strict clock assumption, data not shown), a random starting tree, and a speciation Yule process as the tree prior. In the absence of an independent and reliable calibration point (ie, dated fossil record), relative evolutionary rates (ie, branch lengths) were estimated in substitutions per site by setting the mean clock rate equal to 1.0 following Drummond and Rambaut.<sup>46</sup>

**Table 1.** New nuclear markers developed by PhyloMarker for mouse lemur phylogeny, with primer sequences and PCR annealing temperature (Tm).

Gene	Primers	Sequences	Tm	
IOHDZUNO18	IOHDZUNO18aF	5' ACTCTCGCCCATCACCTATC 3'	54	
	IOHDZUNO18aR	5' ACTGTCCTGTTGTCCACGC 3'		
	IOHDZUNO18bF	5' TGGCTCTCCTGTCTCAATG 3'	60	
	IOHDZUNO18bR	5' GGGGATGGGCACTGTTTC 3'		
	IOHDZUNO18cF	5' GACACCCTCCTGAACAGACG 3'	60	
	IOHDZUNO18cR	5' ATTTGACCCGCCCTTC 3'		
IOHDZUNO20	IOHDZUNO20aF	5' CCTGAGAATCCGAACGCTG 3'	60	
	IOHDZUNO20aR	5' CCCGTCCCCTGTTTTTTG 3'		
	IOHDZUNO20bF	5' ATGATGGTAGGCTGAGGAATG 3'	62	
	IOHDZUNO20bR	5' GGTGGTAACAGTATTGGGTGC 3'		
	IOHDZUNO20cF	5' ATGATGGTAGGCTGAGGAATG 3'	62	
	IOHDZUNO20cR	5' ACTATTTGAGGAACTTGGAGACTG 3'		
	IOHDZUNO20dF	5' ACACTGCCATTTCTGCCTG 3'	60	
	IOHDZUNO20dR	5' AAGTCGCCAACATTGAACG 3'		
	IOHDZUNO23	IOHDZUNO23aF	5' CAACAACGATTCCTCTACCACC 3'	60
		IOHDZUNO23bR	5' TGAGTGACGGTCCCCTGT 3'	
IOHDZUNO23aR*		5' CACCAGCCTCATCTACGGG 3'		
IOHDZUNO23bF*		5' TAAAGGAAGAGAAAATGGAAATAGA 3'		
IOHDZUNO28	IOHDZUNO28aF	5' AACAAGCAGAAGAAATAATCCG 3'	60	
	IOHDZUNO28bR	5' AGGAAAGAAGAGGTTGGAGTTG 3'		
	IOHDZUNO28aR*	5' GTCAGAATCATCCAGCCGA 3'	60	
	IOHDZUNO28bF*	5' ATTGACTGTGAGAAAGGGTGG 3'		
	IOHDZUNO28cF	5' TCCAGCCTTATGTCCTTCG 3'	60	
	IOHDZUNO28cR	5' CCTTCAGTTTATCCTTTCTTTAG 3'		
	IOHDZUNO28dF	5' AGATGGGACTTTGCTACCG 3'	60	
	IOHDZUNO28dR	5' CGAAGAGATGACCTGTTTTTG 3'		
IOHDZUNO30	IOHDZUNO30aF	5' CAGAAGGAAGAAGCAAAGAACTACTA 3'	58	
	IOHDZUNO30bR	5' AGAAACCCAGGAGGACGG 3'		
	IOHDZUNO30aR*	5' CCAACTGATGAAAACCTCCCC 3'		
	IOHDZUNO30bF*	5' TCTTCAGGAGGTGCCCAA 3'		
IOHDZUNO33	IOHDZUNO33aF	5' GAATGGTCTTCGGGCAGAG 3'	58	
	IOHDZUNO33aR	5' ATGCGGCGGTGACAAAG 3'		
	IOHDZUNO33bF	5' TTGCTGGTGACCTGGGAC 3'	60	
	IOHDZUNO33bR	5' CCATCTCAATGCCTTTAGGG 3'		
	IOHDZUNO33cF	5' CCTGGGTGGCAGATAAACG 3'	56	
	IOHDZUNO33cR	5' GGAGGACTTCTTGGCTTGTTTC 3'		
	IOHDZUNO33dF	5' AGGACCTGAAGCAAAGCAC 3'	60	
	IOHDZUNO33dR	5' CGTAGAACCTTGACCTCCATAAC 3'		
	IOHDZUNO33AR1*	5' GAATGGTGGTTGTGCTGGTC 3'	60	
	IOHDZUNO33DF1*	5' GACTATGAGTTCACAGAGGGCAC 3'		
IOHDZUNO53	IOHDZUNO33DR1*	5' GTGCCCTCTGTGAACTCATAGTC 3'	60	
	IOHDZUNO53aF	5' CAGAACACGCTTGAAACTATG 3'		
	IOHDZUNO53aR	5' CCACTGGACTTGAGGCTACTGT 3'		
	IOHDZUNO53bF	5' CAGAACAAAAGAACC GAATGAT 3'		
	IOHDZUNO53bR	5' AACTGGCTACACTGGATTTCC 3'		
MERRF11	MERRF11aF	5' ACTCTCAGTGAATGGGGTTTG 3'	56	
	MERRF11aR	5' GCGGAGGAGGATTTGGA 3'		
	MERRF11bF	5' GCTCAGATACAGACTTCCTTTTAGA 3'	60	
	MERRF11bR	5' CCTGGATTTGGGTGCTTG 3'		
	MERRF11bF1*	5' GCACCTATAGCGATGAAGACAG 3'	60	
	MERRF11bR1*	5' CTTTGCTGCTGACGTACTTGG 3'		

(Continued)

**Table 1.** (Continued)

Gene	Primers	Sequences	Tm
MHDZNPC3	MERRFI1bR2*	5' CTGTCTTCATCGCTATAGGTGC 3'	60
	MHDZNPC3aF	5' CTCTCCATCTGGCATCCTAAC 3'	
	MHDZNPC3bR	5' ATCTCCACTTTCAAATCCAGC 3'	
	MHDZNPC3aR*	5' CTGACTGCTCTCTCCTTTGAAG 3'	
	MHDZNPC3bF*	5' TGACCATCAAGGCATCCC 3'	

**Note:** \*Internal primer for sequencing.

The final analysis was run for 500 million generations, with 50 million trees discarded as burnin with every five thousandth tree kept thereafter. Convergence of the MCMC was assessed by examining trace plots and histograms in Tracer v1.5 after obtaining an effective sample size greater than 200 for all model parameters.<sup>47</sup> A maximum clade credibility tree was generated using the program TreeAnnotator v1.6.1 provided in the BEAST package, with a burnin of 5000 (10%) and visualized in FigTree v1.3.1.<sup>46,48</sup>

## Results

### PhyloMarker development

Two versions of PhyloMarker were developed with the same core procedures. The Web PhyloMarker was built using the LAMP (Linux, Apache, MySQL and Perl/PHP) architecture. The standalone PhyloMarker was programmed in PERL. The latest version of PhyloMarker has many improvements including a global function controlling all the procedures, updated coverage/identity and sequence extraction algorithms, the coverage and identity flags separated for single copy exon and ortholog comparisons, and the ability to use more than two genome data sequence files simultaneously. PhyloMarker can be used for either exon or intron marker searches. The bioinformatics pipeline is presented in Figure 1.

PhyloMarker uses NCBI BLAST core programs for sequence comparison. The program was tested with BLAST 2.2.24, but other versions prior to BLAST+ should work as well. For standalone PhyloMarker, the user needs to download the BLAST package from NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release/2.2.24/>) and install it on a local computer. Several additional PERL packages, including BioPerl, Data, Error, and FindBin, that are used in PhyloMarker are available via CPAN (Comprehensive Perl Archive Network, <http://www.cpan.org/>).

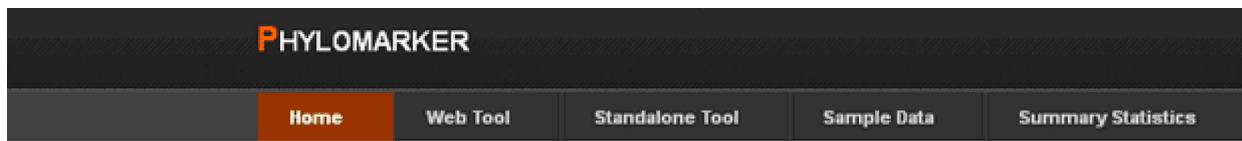
### The use of PhyloMarker

To run Web PhyloMarker, the following four steps are necessary: selecting marker type (exon or intron), choosing data type, setting up parameters, and uploading files (Fig. 2A). For selecting data type, the program accepts either GenBank or FASTA format. The GenBank files are usually much larger and thus Web PhyloMarker requires gunzipped (gzipped) files as input. The gzipped files of genomes can be downloaded from Ensembl (<http://www.ensembl.org/info/data/ftp/index.html>). Genome sequences downloaded from NCBI (<ftp://ftp.ncbi.nih.gov/genomes/>) or other public repositories need to be gzipped for Web PhyloMarker to use. Alternatively, Web PhyloMarker accepts sequences in FASTA format, a universal data format for nucleotide and protein sequences. This feature is particularly useful for high throughput sequence data. The Web PhyloMarker program is suitable for a data set less than 10 MB.

For parameter settings, the user needs to provide the maximum and minimum size of the markers and the minimum sequence identity and coverage values used to identify orthologous genes (inter-genome comparison). Setting identity and coverage values to isolate single copy genes (intra-genome comparison) is currently not available in Web PhyloMarker.

The last steps are to decide the number of genomes to compare and to upload the files. Once the number of genomes is entered, click “Ok”, and the user can upload sequence files in either gzipped GenBank format or FASTA format. Once the “Run PhyloMarker” button is clicked, Web PhyloMarker will be executed with the results page displayed (Fig. 2B). Several links will be shown at the bottom of this page where the user can download single copy exons (or introns) in each genome, pairwise Blast results, and candidate markers. The resulting Excel file includes detailed information of the markers as follows: Gene ID,





## Welcome

PhyloMarker is developed to find single copy nuclear gene markers for the inference of phylogeny. This page acts as a portal to access [web tools](#) and resources (including [source code](#) and [candidate markers](#)).

Both [web](#) and [standalone](#) PhyloMarker share the same core procedures. Web PhyloMarker is easy to use with a size limitation of the input data (<10 MB). If your data files are large and need other options that are not offered by the web program, you are advised to run the standalone program.

A detailed [README](#) file is available in the Standalone Tool page. [Sample files](#) are provided for you to test both the web and standalone programs. PhyloMarker has been used to intron or exon markers in several vertebrate groups (See the [Results](#) page). If you are interested in using PhyloMarker or want to collaborate with us, please [email us](#).



1. Custom Parameter Settings  
Using Custom Identity of: 75  
Using custom coverage of: 50

2. Finding Single Copy Sequences.  
...Adding exons from: TakifuguExample/  
...Extracting exons from: TakifuguExample/  
...Formatting BlastDB for: TakifuguExample/TakifuguExample\_exons  
...Running Blast on: TakifuguExample/TakifuguExample\_exons  
...Finding Single Copy exons for: TakifuguExample/TakifuguExample\_exons  
We parsed 58 records!  
Wrote 19 putative Single Copy exons to file.  
...Adding exons from: OryziasSample/  
...Extracting exons from: OryziasSample/  
...Formatting BlastDB for: OryziasSample/OryziasSample\_exons  
...Running Blast on: OryziasSample/OryziasSample\_exons  
...Finding Single Copy exons for: OryziasSample/OryziasSample\_exons  
We parsed 28 records!  
Wrote 15 putative Single Copy exons to file.

3. Finding Single Copy Orthologs!  
...Formatting BlastDB for: TakifuguExample/TakifuguExample\_single\_copy\_exons  
...Running Blast on: OryziasSample/OryziasSample\_single\_copy\_exons  
...Finding Single Copy exons for: TakifuguExample/TakifuguExample\_single\_copy\_exons.fas and OryziasSample/OryziasSample\_single\_copy\_exons.fas  
We parsed 15 exons!  
Wrote 2 putative Orthologous exon Sequences to file.  
Results.xls

Single Copy exons: [TakifuguExample](#) [OryziasSample](#)  
Blast Results: [TakifuguExample](#) [OryziasSample](#)  
Candidate Markers: download the file [Here](#)

otolemur					microcebus					otolemur vs microcebus				
Gene ID	exon ID	Size	Sequence	GC conter	Gene ID	exon	Size	Sequence	GC conter	E-value	Identity(%)	Coverage	Bit Score	Alignm
ENSOGAG	ENSOGAE1	659	TTTCAGTT	0.652504	ENSMICG	ENSMICE0	656	GTGCGAG	0.676829	0	94.8	99.7	0	65
ENSOGAG	ENSOGAE1	602	GTCGCTG	0.579734	ENSMICG	ENSMICE0	602	TCAGTAG	0.611296	0	95.02	100	0	60
ENSOGAG	ENSOGAE1	538	CTTAAAA	0.460967	ENSMICG	ENSMICE0	538	TTAGTTTA	0.45539	0	94.94	99.26	0	53
ENSOGAG	ENSOGAE1	526	CTTGTCTC	0.65019	ENSMICG	ENSMICE0	534	AAAATCC	0.726592	8.00E-162	88.39			53
ENSOGAG	ENSOGAE1	1026	GTAAGTGC	0.562378	ENSMICG	ENSMICE0	1026	CTACCGTA	0.572125	0	90.65	99.7	0	101
ENSOGAG	ENSOGAE1	722	CGAGGAC	0.645429	ENSMICG	ENSMICE0	722	GCCTCGGA	0.674515	0	95.01		0	72
ENSOGAG	ENSOGAE1	1209	GCTCCTTT	0.565757	ENSMICG	ENSMICE0	606	CTGTTCTG	0.534653	0	93.89		0	60
ENSOGAG	ENSOGAE1	586	CTTGCTCT	0.573379	ENSMICG	ENSMICE0	741	AGGTAAC	0.573549	2e-24	84.13	99.7	4	124
ENSOGAG	ENSOGAE1	999	ATGAAAG	0.627628	ENSMICG	ENSMICE0	999	ATGAAAG	0.653654	0	93.59	100	0	99
ENSOGAG	ENSOGAE1	1503	TGTTTCATA	0.656687	ENSMICG	ENSMICE0	1034	ACGGCCC	0.663443	0	90.66	68.38	0	70
ENSOGAG	ENSOGAE1	572	CAGTTTGC	0.391608	ENSMICG	ENSMICE0	572	CAAAAAA	0.38986	0	94.76	100	0	57
ENSOGAG	ENSOGAE1	996	TGGTTGCA	0.601406	ENSMICG	ENSMICE0	977	ATGGTCA	0.610031	6.00E-130	82.37	84.65	6e-	83
ENSOGAG	ENSOGAE1	1005	TCAAAAC	0.645771	ENSMICG	ENSMICE0	921	CTGTTCCC	0.677524	0	97.14	98.7	0	90

Figure 2. PhyloMarker Web Tool input page (A), result page (B), and resulting Excel table (C).



Intron or Exon ID, GC content, sequences, identity and coverage (Fig. 2C).

To run the standalone PhyloMarker, only a few steps are required. A detailed readme file with instructions is provided to demonstrate how to utilize PhyloMarker. The PhyloMarker program and the readme file can be downloaded from the “Standalone Tool” page. We have also created small tutorial files located on the “Testing” page.

### Genus *Microcebus* as a test case for PhyloMarker

Utilizing the partial genomes of *Microcebus murinus* (1.93 X) and *Otolemur garnettii* (1.50 X) available from the Ensembl database, we utilized standalone PhyloMarker to extract single copy nucDNA candidate markers in common between the two prosimians. To test the practical value of potential phylogenetic markers identified by PhyloMarker (Table 2), 20 genes were randomly picked out of 253 candidate markers identified by PhyloMarker v2.0 and assessed for the 17 nominal *Microcebus* species. Conserved flanking regions of the candidate markers identified from the *M. murinus* genome were utilized to design PCR primer pairs for each *Microcebus* species (Table 2). Due to the large overall size of each fragment, the PCR primer pairs were devised to produce overlapping segments. We successfully amplified and assembled a consensus contig for nine candidates. The length of the consensus sequences ranged from 527 to 1,588 bp. Additional characteristics of the data set were presented in Table 3.

Using concatenated sequences of all nine nucDNA markers (8,007 bp), a phylogeny of the 17 nominal

*Microcebus* species in addition to two outgroups was inferred (Fig. 3A). The taxonomic framework of the nucDNA phylogeny was mostly congruent to the phylogenetic tree based on COII and PAST fragment combined mtDNA data with most of the differences confined to shifts among groups of species (Fig. 3 and 4). A multispecies coalescent approach was taken to infer the most likely species tree on the basis of gene tree topologies estimated for each molecular marker as obtained with \*BEAST (Fig. 5). Compared to the phylogenetic tree based solely on mtDNA sequence data, most sister taxa relationships in Figure 5 are congruent to Figure 3B except for the position of *M. tavaratra* and *M. simmonsii*. The phylogenetic relationships among 17 nominal *Microcebus* species inferred from species tree and concatenated analyses followed a regional geographic distribution except for *M. tavaratra* and *M. jollyae*. For the nuclear and mtDNA combined data set, both phylogenetic tree and species tree are congruent (Figs. 4 and 5B).

### Discussion

PhyloMarker is a unique tool that can be used to find and develop single copy nuclear gene markers for the inference of large scale phylogenies. The computer package has been used by several research programs to identify phylogenetic markers in model based mammals and birds to be utilized in specific non-model species (Table 2). For instance, the complete genomes of chicken and zebra finch were compared and identified 730 exon markers and 37 intron candidate markers for heron phylogeny

**Table 2.** Genomic statistics and the numbers of single copy genes and candidate markers identified by PhyloMarker in three vertebrate groups.

Species	Genes	Exons	Single copy exons	Introns	Single copy introns	Exon markers	Intron markers
Chicken	17934	9768	4380	39871	9077		
Zebra finch	18581	5865	2288	31053	10567		
Chicken vs. zebra finch						730	37
Rat	29516	16110	4896	49314	4176		
Mouse	36822	41018	6681	76337	2988		
Rat vs. mouse						1038	595
Galago	28085	8557	2998	26779	5982		
Mouse lemur	24994	4979	1788	17928	4684		
Galago vs. mouse lemur						576	220

**Note:** Mouse lemur represents the *Microcebus murinus* genome and the Galago represents the *Otolemur garnettii* genome.

**Table 3.** Summary information of nine novel exon genes identified by PhyloMarker amplified in the 17 nominal *Microcebus* species.\*

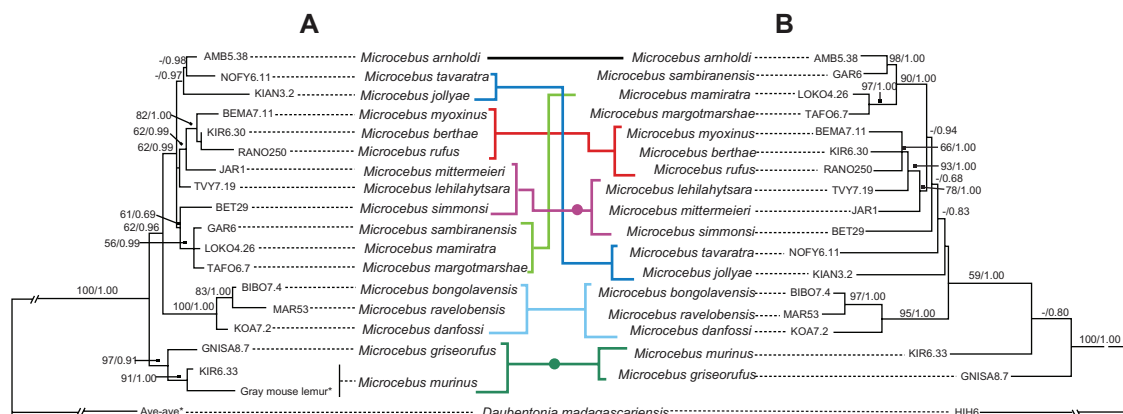
Marker name	Chromosome	Start position	Length (bp)	No. of var.	No. of PI	GC%	Genetic distance (%)
IOHDZUNO18	1	16577336	842	66/8	17/4	56.3	1.1 ± 0.1/0.2 ± 0.1
IOHDZUNO20	2	149247023	1002	49/14	15/7	47.1	0.7 ± 0.1/0.2 ± 0.1
IOHDZUNO23	8	124265861	697	58/7	9/3	43	0.9 ± 0.1/0.1 ± 0.1
IOHDZUNO28	10	91177180	972	97/21	42/13	43.6	1.4 ± 0.1/0.4 ± 0.1
IOHDZUNO30	13	36909363	655	71/9	17/4	44.1	1.4 ± 0.2/0.2 ± 0.1
IOHDZUNO33	14	70633572	1547	98/17	37/8	50.9	0.9 ± 0.1/0.2 ± 0.1
IOHDZUNO53	X	147743462	741	73/23	28/6	46.1	1.5 ± 0.2/0.5 ± 0.1
MERRFI1	1	8073387	883	90/24	20/3	47.5	1.0 ± 0.1/0.2 ± 0.0
MHDZNPC3	1	182443077	671	131/17	48/7	53.5	2.6 ± 0.2/0.3 ± 0.1

**Abbreviations:** \*bp, base pairs; var., variable sites (Ingroup + Outgroup/Ingroup); PI, parsimony informative sites (Ingroup + Outgroup/Ingroup); Genetic distance, average uncorrected distance (Ingroup + Outgroup/Ingroup).

(<http://bioinfo-srv1.awh.unomaha.edu/phylomarker/results.php>). Furthermore, a comparison of rat and mouse genomes resulted in 1,038 exon and 595 intron candidate markers for Wood mice phylogeny. Another salient feature of PhyloMarker is its utility to identify duplicated genes through intra-genome comparison, an important task in the study of molecular evolution and species phylogeny. Moreover, PhyloMarker can take advantage of high throughput genomic or transcriptomic sequence data to select candidate phylogenetic markers, as long as the assembled sequences are in the FASTA format. In regard to future upgrades, we will incorporate algorithms to improve sensitivity and accuracy for sequence comparison and add support for newer versions of BLAST+. The chromosomal positions of candidate markers are currently

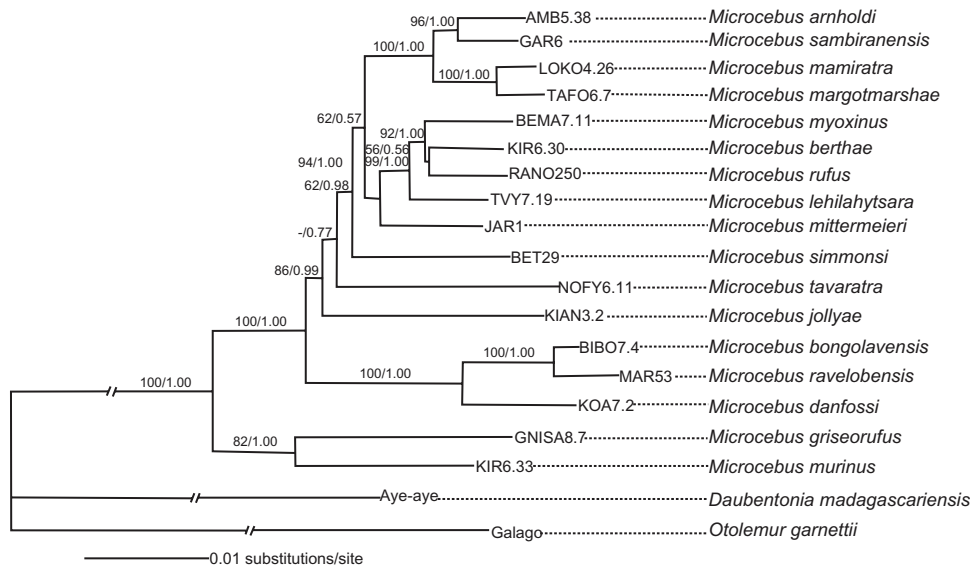
not included in the resulting Excel file. A utility is needed to parse the chromosome information and display the distribution of the markers as a chromosome map. The source code is available and users are welcome to test PhyloMarker.

In regards to the genus *Microcebus* case study, lemur taxonomy has been dramatically revised due to extensive field work to previously unexplored regions and to advances in molecular technology, primarily centered on mtDNA sequencing and analyses. Consequently, the number of lemur species described in the past three decades has increased from 36 to 101 species.<sup>49,50</sup> This proliferation has been questioned as to whether or not the recently described lemurs were unnoticed cryptic species or taxonomic inflation.<sup>26,27</sup> There are many potential sources of discrepancy



**Figure 3.** Phylogenetic relationships among the 17 recognized *Microcebus* species inferred from maximum likelihood and Bayesian approaches, utilizing the combined sequence data of nine novel nucDNA genes identified by PhyloMarker (A) and mtDNA COII and PAST fragments (B) for 18 mouse lemur individuals with two outgroup taxa.

**Notes:** Numbers above the branches represent ML bootstrap support/Bayesian posterior probability support. The sequences drawn from the Ensembl genome database and Perry et al<sup>37</sup> are marked with an asterisk. Between trees (A and B) color lines link shifts among groups of species only within the genus *Microcebus* whereas color lines and solid circles indicate incongruence between the nucDNA and mtDNA data sets between *Microcebus* species.

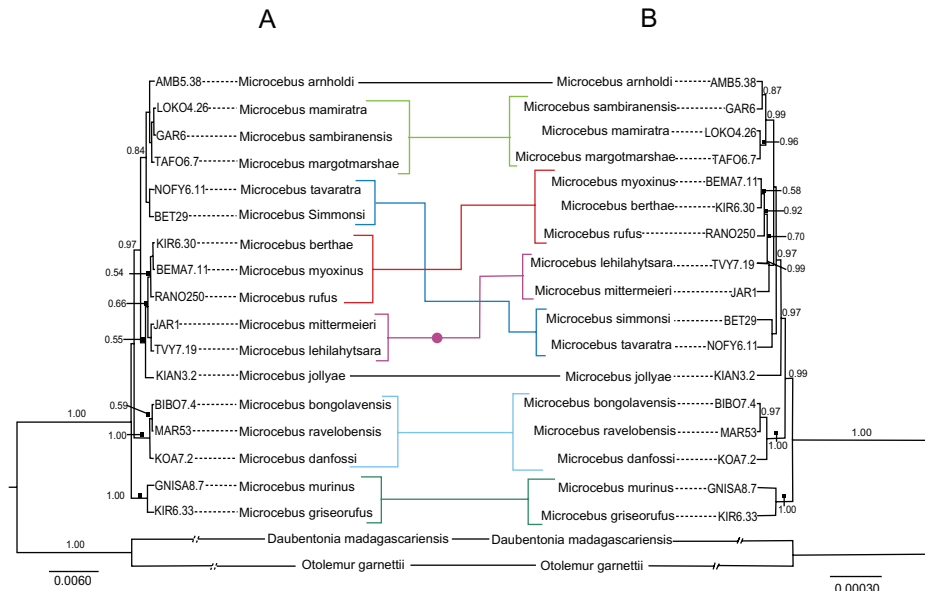


**Figure 4.** Phylogenetic relationships among *Microcebus* species inferred from the ML and Bayesian approaches for the concatenated nucDNA gene identified by PhyloMarker and mtDNA combined sequence data from 17 mouse lemur individuals with two outgroup taxa. **Note:** Numbers above the branches represent ML bootstrap support/Bayesian posterior probability support.

between gene trees and species trees that contribute to this argument, including unresolved genetic issues pertaining to horizontal transfer, lineage sorting, and gene duplication or loss.<sup>28,45,51,52</sup>

In this study, we verified the phylogeny, as previously defined by mtDNA analyses, of the current

17 *Microcebus* species using nine novel nuclear loci identified with PhyloMarker. Incongruence between mtDNA and nucDNA data sets was primarily related to alternative linkages between identical groups of species. However, the original criteria for defining these species including character state differences



**Figure 5.** Maximum clade credibility phylogeny of the genus *Microcebus* inferred by the \*BEAST species tree analyses of nine concatenated nucDNA genes identified by PhyloMarker from 18 mouse lemur individuals with two outgroup taxa (A) and nuclear gene identified by PhyloMarker and mtDNA combined sequence data from 17 mouse lemur individuals with 2 outgroup taxa (B).

**Notes:** Clade posterior probabilities are shown above branches. The node gray bars represent 95% node highest posterior density. Between trees (A and B) color lines link shifts among groups of species only within the genus *Microcebus* whereas color lines and solid circles indicate incongruence between the concatenated nucDNA and the combined nucDNA/mtDNA data sets between *Microcebus* species. The two outgroup sequences were drawn from the Ensembl genome database and Perry et al.<sup>37</sup>



and geographic barriers were not altered, but simply augmented. Debates continue to simmer regarding the validity of lemur taxonomic revisions with important management and conservation decisions dependent on a miniscule percentage of each taxon's genome. With advances in next generation sequencing, additional prosimian genomes will become available to extract comparable single copy exons and introns that function across all lemur taxa. PhyloMarker provides a practical utility to extract numerous single copy genes from large repositories of sequence data across multiple distant or closely related taxa, thus enabling future scientific decisions to be based on sound information.

## Conclusion

PhyloMarker, a phylogenomic tool, is introduced to find single copy nuclear gene markers through genome comparison. It involves intra-genome comparison for detecting single copy exons or introns and inter-genome comparison for orthologous markers. The software is flexible and user friendly since the user can set different threshold values for marker identification and has both web and standalone versions of the same core program. Source code and sample data are available at the project website: <http://bioinfo-srv1.awh.unomaha.edu/phylomarker/>. Users are encouraged to test PhyloMarker, along with suggesting new features that can be included in future upgrades. The power of PhyloMarker in mining new markers for the inference of reliable phylogeny was demonstrated in the case study of mouse lemurs, genus *Microcebus*. Furthermore, additional candidate markers for fish (puffer fish versus rice fish), birds (chicken versus zebra finch), and rodents (mouse versus rat) were developed using PhyloMarker are also available at the project website.

## Acknowledgements

The authors are particularly grateful to C. Li for his initiation of the project and acknowledge F. J. Potmesil and G. Zhang for their help with programming. We would also like to acknowledge G. Orti and other users who beta tested PhyloMarker and provided us feedback. We would like to thank the staff, field assistants and drivers of the Madagascar Biodiversity Partnership for the dedication in collecting the necessary samples in Madagascar, and the staff at the Grewcock Center for Conservation and Research at

Omaha's Henry Doorly Zoo and Aquarium for their expertise with sequence generation. We would like to thank the Ministère des Eaux et Forêts of Madagascar and USFWS for assistance in issuing the necessary collection, exportation and importation permits.

## Funding

The development of PhyloMarker was supported by the National Science Foundation (DEB-0732838) and by the University Committee on Research and Creative Activity (UCRCA) at the University of Nebraska at Omaha. The phylogeny of mouse lemur was supported by a grant from Ahmanson Foundation and the generosity of Bill and Berniece Grewcock and the Theodore F. and Claire M. Hubbard Family Foundation.

## Author Contributions

Conceived and designed the experiments: GL, EEL, and RL. Developed the software: TWR. Generated and analyzed the data: RL, LZ, CAB, SEE, MLW, MCC and GHP. Wrote the manuscript: EEL, GL, RL, CAB, SEE, MLW, and TWL. All authors reviewed and approved of the final manuscript.

## Competing Interests

GP has a grant or grant pending from NSF. Other authors disclose no potential conflicting interests.

## Disclosures and Ethics

As a requirement of publication authors have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

## References

1. Rokas A, Williams BL, King N, Carrol SB. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*. 2003;425:798–804.



2. Chen WJ, Miya M, Saitoh K, Mayden RL. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. *Gene*. 2008;423(2):125–34.
3. Philippe H, Brinkmann H, Lavrov DV, et al. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol*. 2011; 9(3):e1000602.
4. Li C, Ortí G, Zhang G, Lu G. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol Biol*. 2007;7:44.
5. Igea J, Juste J, Castresana J. Novel intron markers to study the phylogeny of closely related mammalian species. *BMC Evol Biol*. 2010;10:369.
6. Li C, Riethoven J-J, Ma L. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evol Biol*. 2010;10:90.
7. Mamanova L, Coffey AJ, Scott CE, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 2010;7(2):111–18.
8. Glaw F, Vences M. *A Field Guide to the Amphibians and Reptiles of Madagascar*. 3rd ed. Cologne: Vences and Glaw Publishers; 2007.
9. Vieites DR, Wollenberg KC, Andreone F, Köhler J, Glaw F, Vences M. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proc Natl Acad Sci U S A*. 2009;106:8267–72.
10. Ramasindrazana B, Goodman SM, Schoeman MC, Appleton B. Identification of cryptic species of *Miniopterus* bats (Chiroptera: Miniopteridae) from Madagascar and the Comoros using bioacoustics overlaid on molecular genetic and morphological characters. *Biol J Linn Soc*. 2011;104: 284–302.
11. Stiassny MLJ, Chakrabarty P, Loiselle PV. Relationships of the Madagascan cichlid genus *Paretroplus* Bleeker 1868, with a description of a new species from the Betsiboka River drainage of northwestern Madagascar. *Ichthyol Explor Fres*. 2001;12(1):29–40.
12. Goodman SM, Raxworthy CJ, Maminirina CP, Olson LE. A new species of shrew tenrec (*Microgalejobihely*) from northern Madagascar. *J Zool (Lond)*. 2006;270:384–98.
13. Louis EE, Coles MS, Andriantompohavana R, et al. Revisions of Mouse Lemurs (*Microcebus*) of Eastern Madagascar. *Int J Primatol*. 2006;27: 347–89.
14. Louis EE, Engberg SE, McGuire S, et al. Revision of the mouse lemurs, *Microcebus* (Primates, Lemuriformes), of northern and northwestern Madagascar with descriptions of two new species at Montagne d'Ambre National Park and Antafondro Classified Forest. *Primate Conserv*. 2008;23:19–38.
15. Lei R, Engberg SE, Andriantompohavana R, et al. Nocturnal lemur diversity at Masoala National Park. *Spec Publ Mus, Texas Tech Univ*. 2008;SP53:1–41.
16. Radespiel U, Reimann W, Rahelinirina, M, Zimmermann E. Feeding ecology of sympatric mouse lemur species (*Microcebus murinus* and *M. ravelobensis*) in northwestern Madagascar. *Am J Primatol*. 2006;59(4): 139–51.
17. Mittermeier RA, Ganzhorn JU, Konstant WR, et al. Lemur Diversity in Madagascar. *Int J Primatol*. 2008;29:1607–56.
18. Martin RD. Adaptive radiation and behavior of Malagasy lemurs. *Philos Trans R Soc Lond B Biol Sci*. 1972;264:295–352.
19. Zimmermann E, Cepok S, Rakotoarison N, Zietemann V, Radespiel U. Sympatric mouse lemurs in north-western Madagascar: A new rufous mouse lemur species (*Microcebus ravelobensis*). *Folia Primatol*. 1998;69: 106–14.
20. Rasoloarison RM, Goodman SM, Ganzhorn JU. Taxonomic revision of mouse lemurs (*Microcebus*) in the western portions of Madagascar. *Int J Primatol*. 2000;21:963–1019.
21. Kappeler PM, Rasoloarison RM, Razafimanantsoa L, Walter L, Roos C. Morphology, behaviour and molecular evolution of giant mouse lemurs (*Mirza* spp.) Gray, 1870, with description of a new species. *Primate Rep*. 2005;71:3–26.
22. Andriantompohavana R, Zaonarivelo JR, Engberg SE, et al. The mouse lemurs of northwestern Madagascar with a description of a new species at Lokobe Special Reserve. *Occas Papers Mus, Texas Tech Univ*. 2006; 259:1–23.
23. Olivieri G, Zimmermann E, Randrianambinina B, et al. The ever-increasing diversity in mouse lemurs: three new species in north and northwestern Madagascar. *Mol Phylogenet Evol*. 2007;43:309–27.
24. Radespiel U, Olivieri G, Rasolofofon DW, et al. Exceptional diversity of mouse lemurs (*Microcebus* spp.) in the Makira region with the description of one new species. *Am J Primatol*. 2008;70:1–14.
25. Radespiel U, Ratsimbazafy JH, Rasoloharijaona S, et al. First indications of a highland specialist among mouse lemurs (*Microcebus* spp.) and evidence for a new mouse lemur species from eastern Madagascar. *Primates*. 2012;53(2):157–70.
26. Tattersall I. Madagascar's lemurs: cryptic diversity or taxonomic inflation? *Evol Anthropol*. 2007;16:12–23.
27. Markolf M, Brameier M, Kappeler PM. On species delimitation: Yet another lemur species or just genetic variation? *BMC Evol Biol*. 2011; 11:216.
28. Heckman KL, Mariani CL, Rasoloarison R, Yoder AD. Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). *Mol Phylogenet Evol*. 2007;43:353–67.
29. Weisrock DW, Rasoloarison RM, Fiorentino I, et al. Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS ONE* 2010;5(3):e9883.
30. Horvath JE, Weisrock DW, Embry SL, et al. Development and application of a phylogenomic toolkit: resolving the evolutionary history of Madagascar's lemurs. *Genome Res*. 2008;18:489–99.
31. Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss AR. Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. *Mol Ecol*. 2006;15:2489–503.
32. Backström N, Fagerberg S, Ellegren H. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol Ecol*. 2008;17(4):964–80.
33. Perelman P, Johnson WE, Roos C, et al. A molecular phylogeny of living primates. *PLoS Genet*. 2011;7(3):e1001342.
34. Lu G, Jiang L, Helikar RM, et al. GenomeBlast: a web tool for small genome comparison. *BMC Bioinformatics*. 2006;7(Suppl 4):S18.
35. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. New York, NY: Cold Spring Harbor Press; 1998.
36. Silva WA Jr, Costa MCR, Valente V, et al. PCR template preparation for capillary DNA sequencing. *Bio Techniques*. 2001;30:537–42.
37. Perry GH, Reeves D, Melsted P, et al. A genome sequence resource for the aye-aye (*Daubentoniadamagascariensis*), a nocturnal lemur from Madagascar. *Genome Biol Evol*. 2012;4(2):126–35.
38. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
39. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24: 1596–9.
40. Swofford DL. *PAUP\* Phylogenetic Analysis Using Parsimony and Other Methods*. Sunderland, MA: Sinauer Associates Inc.; 2001.
41. Huelsenbeck JP, Ronquist F. MRBAYES v.3.0b4: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001;17:754–5.
42. Ronquist F, Huelsenbeck JP. Mr Bayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
43. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics*. 1998;14:817–8.
44. Nylander JAA. *MrModeltest v2*. Program distributed by the author; 2004.
45. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol*. 2010;27:570–80.
46. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214.
47. Rambaut A, Drummond A. Tracer, version 1.5. Available online: <http://beast.bio.ed.ac.uk/Tracer>. Accessed on Nov 30, 2009.
48. Rambaut A. FigTree, version 1.3.1. Available online: <http://tree.bio.ed.ac.uk/software/figtree>. Accessed on Dec 21, 2009.



49. Tattersall I. *The Primates of Madagascar*. New York: Columbia University Press; 1982.
50. Mittermeier RA, Louis EE Jr, Richardson M, et al. *Lemurs of Madagascar*. 3rd ed. Washington, DC.: Conservation International; 2010.
51. Knowles LL. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Syst Biol*. 2009;58:463–7.
52. Liu L, Yu L, Pearl DK. et al. Estimating species phylogenies using coalescence times among sequences. *Syst Biol*. 2009;58:468–77.