Data Article

Molecular and phylogenetic datasets for the Asterophryinae frogs of New Guinea with additional data on lifestyle, geography, and elevation

Ethan C. Hill[a], Mary J. Jarman[a], Claire J. Fraser[a], Diana F. Gao[b], Elizabeth R. Henry[a], Allison R. Fisher[a], Bulisa Iova[c], Allen Allison[d], Marguerite A. Butler[a],*

[a] School of Life Sciences, University of Hawaiʻi at Mānoa, Honolulu, HI 96822, USA
[b] University of San Francisco, San Francisco, CA 94116, USA
[c] National Museum and Art Gallery, P. O. Box 5560 Boroko, National Capital District, Papua New Guinea
[d] Bernice P. Bishop Museum, Honolulu, HI 96817, USA

A R T I C L E   I N F O

Dataset link: Molecular and phylogenetic datasets for the Asterophryinae frogs of New Guinea with additional data on lifestyle, geography, and elevation (Original data)

A B S T R A C T

The data provided here are related to the article "Resolving the Deep Phylogeny: Implications for Early Adaptive Radiation, Cryptic, and Present-day Ecological Diversity of Papuan Microhylid Frogs" [1]. The dataset is based on 233 tissue samples of the subfamily Asterophryinae, with representatives from all recognized genera, in addition to three outgroup taxa. The sequence dataset contains over 2400 characters per sample for five genes: three nuclear (Seventh in Absentia (SIA), Brain Derived Neurotrophic Factor (BDNF), Sodium Calcium Exchange subunit-1 (NXC-1)), and two mitochondrial loci (Cytochrome oxidase b (CYTB), and NADH dehydrogenase subunit 4 (ND4)); and is 99% complete. New primers were designed for all loci and accession numbers for the raw sequence data are provided. The sequences are

used with geological time calibrations to produce time-calibrated Bayesian inference (BI) and Maximum Likelihood (ML) phylogenetic reconstructions using BEAST2 and IQ-TREE. Lifestyle data (arboreal, scansorial, terrestrial, fossorial, semi-aquatic) were collected from the literature and field notes and used to infer ancestral character states for each lineage. Collection location and elevation data were used to verify sites where multiple species or candidate species co-occur. All sequence data, alignments, and associated metadata (voucher specimen number, species identification, type locality status, global positioning system [GPS] coordinates, elevation, site with species list, and lifestyle) as well as the code to produce all analyses and figures are provided.

Specifications Table

Subject	Biological Sciences, Phylogeny and Evolution
Specific subject area	Phylogenetics and biodiversity studies of a hyperdiverse clade of frogs, anuran lifestyle and diversification, herpetology of the South Pacific.
Type of data	Tables
	Figures
	Supplementary Materials (code, input files, Table 1 containing metadata)
How the data were acquired	Deoxyribonucleic acid (DNA) sequences were acquired via Sanger sequencing for three nuclear loci: brain-derived neurotrophic factor (BDNF), sodium calcium exchange subunit-1 (NXC-1); and seventh in absentia (SIA); and two mitochondrial loci: NADH dehydrogenase-4 (ND4) and cytochrome b (CYTB).
	Primers were designed using the National Center for Biotechnology Information software Primer-Blast [2] and HYDEN [3].
	Sequences were aligned with MUSCLEv3 [4]. The best evolutionary model was fit using PartitionFinder2 [5].
	Phylogenetic trees were inferred with BEAST v2.6.0 [6] and IQ-TREE2 [7].
	Ancestral character state reconstruction on the lifestyle data was conducted using the fitDiscrete in the geiger package [8].
	Probabilities of ancestral states were simulated using the make.simmap function and diversification rate was modeled using the ltt function, both in phytools [9].
	Geographic distances between collection sites was computed from gps coordinates with the geodist function in package geodist [10].
	Phylogenetic plots were created using the R packages ggtree, ouch and treeio [11,12,13,19], and with a custom script based on modifications of ouch. All analyses were conducted in the R statistical computing environment unless otherwise mentioned [14].
Data format	Raw
	Analyzed
Description of data collection	GPS and elevation data were collected in the field and from museum databases at the Bernice Pauahi Bishop Museum (Table 1 in the supplementary materials).
	Microhabitat use and lifestyle data were gathered from field notes or from the literature, citations are listed in Table 1 in the supplementary materials.
	DNA was extracted from liver tissue using Qiagen DNeasy blood and tissue kits or NEB Monarch Genomic DNA purification kits following manufacturer's protocols. Polymerase chain reaction (PCR) protocols and primers (Table 2) are described below. Sanger sequencing followed standard protocols using Applied Biosystems BigDye terminator chemistry on an ABI 3730XL sequencer at the University of Hawaii at Manoa's Advanced Studies of Genomics, Proteomics and Bioinformatics facility (https://www.hawaii.edu/microbiology/asgpb/).

| Data source location	Samples were collected by us or provided through loans from throughout New Guinea and its satellite islands of Normanby, Fergusson, New Britain, Misima, Rossel, Sudest, as well as the Philippines and Sulawesi. Specific sample locations including GPS and elevation are provided in Table 1 in the supplementary materials.
Data accessibility	Data is available within this article and at the public repository at DOI: 10.5281/zenodo.7063167 Genbank Accession Numbers: MZ634561:MZ634796 (SIA) MZ634797:MZ635031 (BDNF) MZ635032:MZ635268 (NXC) MZ635269:MZ635501 (CYTB) MZ647717:MZ647945 (ND4)
Related research article	E.C. Hill, C.J. Fraser, D.F. Gao, M.J. Jarman, E.R. Henry, B. Iova, A. Allison, M.A. Butler, Resolving the deep phylogeny: Implications for early adaptive radiation, cryptic, and present-day ecological diversity of Papuan microhylid frogs. *Molecular Phylogenetics and Evolution.* 177 (2022) 107618, 10.1016/j.ympev.2022.107618

Value of the Data

- Molecular phylogenetic studies of the hyperdiverse Asterophryinae frogs have proved insufficient for resolving the deep nodes of this radiation. The nuclear and mitochondrial Sanger sequence data presented here for 205 species, combined with large geographic sampling and a 99% complete sequence dataset produce a nearly completely resolved and strongly supported phylogeny, clarifying intergeneric relationships, identifying many cryptic species complexes, and resolving many taxonomic questions.
- The phylogenetic data is complemented by associated GPS and lifestyle data from 77 sites across all major geological sectors of New Guinea and several satellite islands. These data will be foundational for future studies of biogeography, spatial analyses through time, pacific biodiversity, ecological diversification, community assembly, character evolution, and exploring the tempo and mode of evolution.
- New primers are presented for three nuclear and five mitochondrial loci which resulted in 99% completeness of the molecular dataset, and will be useful for sequencing other vertebrates (nuclear loci) or anurans (mitochondrial loci).
- The dataset includes complete species sampling in 18 multi-species communities ranging from two to 15 species, which will be useful for exploring the impact of data completeness for sequence or geographical sampling, and rigorous model fitting on resolving large phylogenies for clades that have undergone rapid diversification.

1. Data Description

Asterophryinae is a hyperdiverse subfamily of microhylid frogs that has remained poorly resolved despite decades of phylogenetic work [reviewed in 1]. Given the ∼20 MY age of the clade and the poor resolution of deeper nodes in previous efforts, we suspected that long-branch attraction was a problem preventing resolution of nodes along the backbone of the phylogeny, and that mitochondrial loci alone were not sufficient to clarify intergeneric relationships. In addition, previous studies included only a small fraction of the clade's diversity. In this study, we present a well resolved phylogeny for 205 species (233 taxa) of Papuan Asterophryinae using both Bayesian inference (Fig. 1) or Maximum Likelihood (Fig. 2) reconstruction methods, that importantly expands geographic sampling and tests whether widely-distributed species instead represent cryptic diversity. These trees show well-defined genera and clarified the majority of intergeneric relationships, and also clarified cryptic species complexes for many taxa that were once believed to be widespread. Instead, the data reveal a mosaic of species with small geographic ranges and a tendency toward fragmentation. Geographic distance between tips in the

Bayesian Inference Phylogeny (BEAST2)
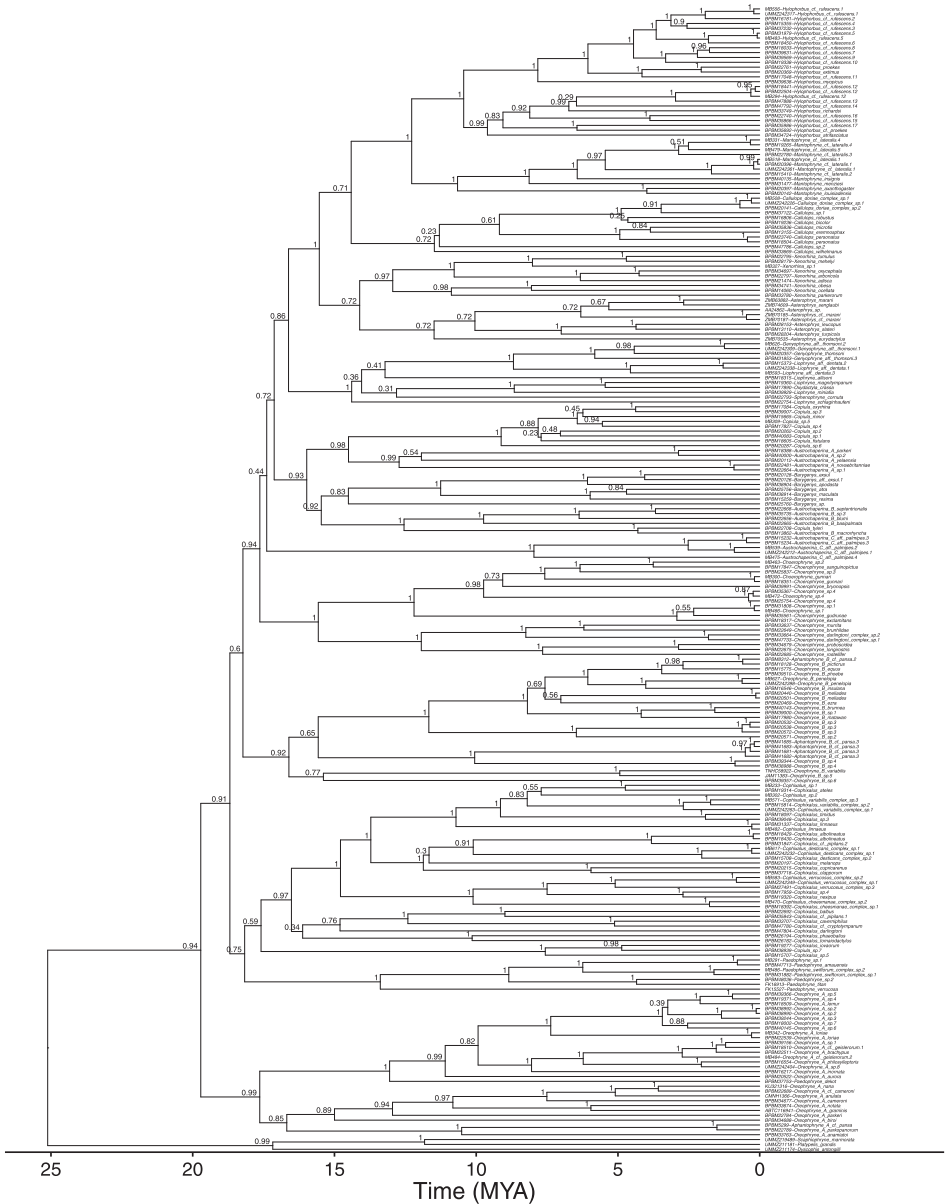
Fig. 1. Time calibrated Bayesian inference phylogeny for 233 samples of Asterophryinae and three outgroups generated using BEAST2. Posterior probabilities are provided above nodes. All trees contain the same set of tips.

tree can be obtained from GPS coordinates (Table 1 in the supplementary materials), and used in analyses of geographic range evolution.

We also show nuclear-only (Fig. 3) and mitochondrial-only (Fig. 4) Bayesian inference trees, which are highly concordant in topology, although suffering a loss of resolution in relation to the nuclear+mitochondrial combined data tree. Therefore, the resolution of this phylogeny can

Maximum Likelihood Phylogeny (IQTREE)
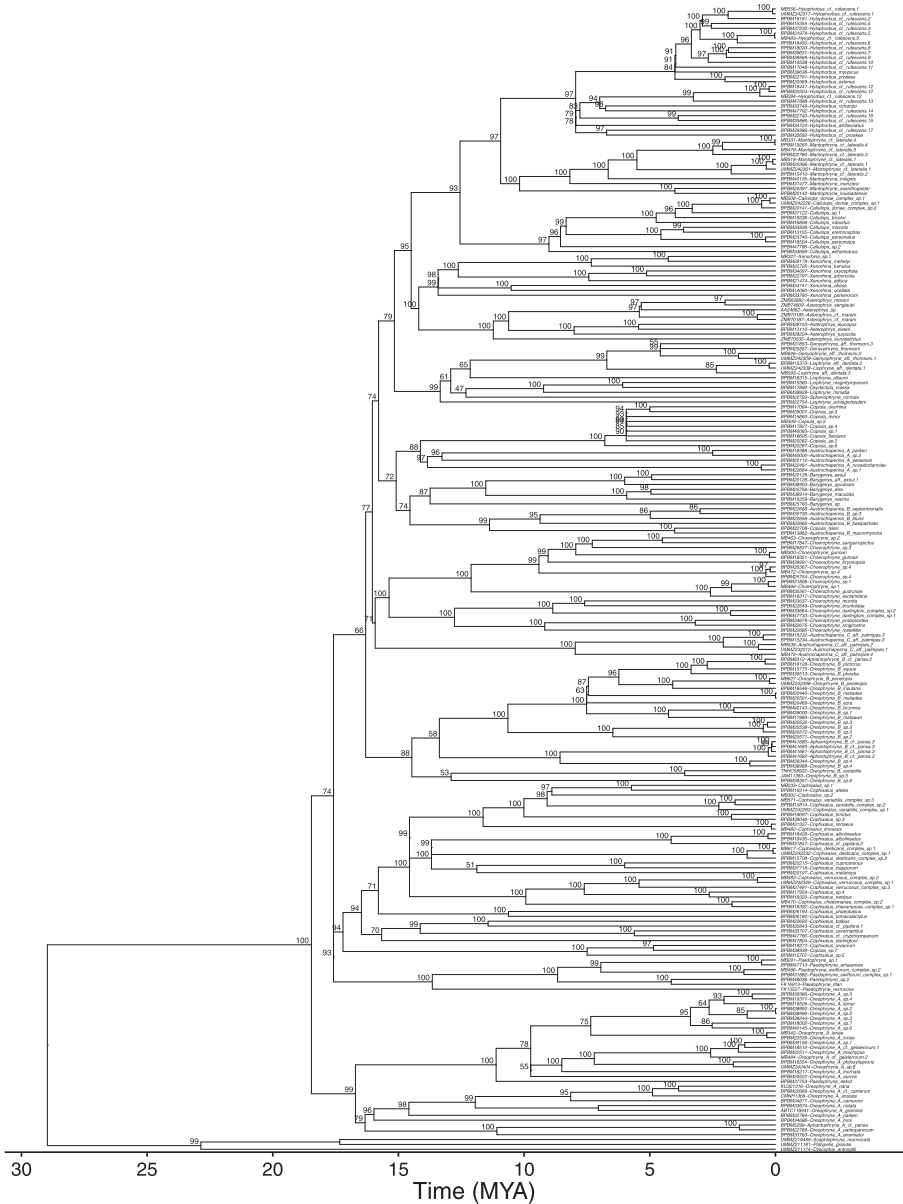
Fig. 2. Time calibrated maximum likelihood phylogeny generated using IQTREE with bootstrap support indicated above nodes.

only be achieved through the combination of nuclear and mitochondrial datasets (Figs. 1–4). The nuclear-only tree recovered an age estimate for the clade of 30 MY, compared to 17 MY from the mitochondrial-only tree, and and age of 20 MY for the combined dataset, which are concordant with other estimates.

Bayesian Inference Phylogeny (BEAST2)
Nuclear Loci Only

Fig. 3. Nuclear-DNA-ooutgropnly time calibrated Bayesian inference phylogeny.

In addition, we present the specimen vouchers used in the phylogenetic analysis (Table 1 in the supplementary materials), the species identification, collection site and locality, noting whether it was the type locality, species citation, GPS, elevation, site, geological terrane, lifestyle based on microhabitat use with their citations, and GenBank accession numbers by locus.

Bayesian Inference Phylogeny (BEAST2)
Mitochondrial Loci Only

Fig. 4. Mitochondrial-DNA-only time calibrated Bayesian inference phylogeny.

We developed new primers three nuclear (Seventh in Absentia (SIA), Brain Derived Neurotrophic Factor (BDNF), Sodium Calcium Exchange subunit-1 (NXC-1)), and two mitochondrial loci (Cytochrome oxidase b (CYTB), and NADH dehydrogenase subunit 4 (ND4)) loci (Table 2). We present the aligned dataset of 2474 base pairs (as input datasets for BEAST2 or IQ-TREE2) that is 99% complete along with the data partitions, evolutionary models, and geological timings used to reconstruct the phylogeny ([1]; Fig. 3; Table 3).

Table 2

Primer sequences redesigned for this study, based on Rivera et al. [15]. Product length reported in base pairs, and Tm in°C. Starting touchdown PCR temperature range in°C is reported in TD range. Outer indicates the outermost most forward/reverse primer for a locus.

Name	Microhylid Specific	Outer	Primer Sequence - 5'-3'	Length	Tm	TD PCR Range
SIA 1F*		x	TCGAGTGCCCCGTGTGYTTYGAYTA	-400-450	63	66-56
SIA 455R*		x	GAAGTGGAAGCCGAAGCAGSWYTGCATCAT		64.7	
SIA 401R	x		AGGTTAATGTCCGTGGCGA		56.5	
SIA m731R**	x		GTTCCTATTAGYTGYACRATKGCRAAGAAC	-700	58.7	50-60
BDNF 1F*		x	ACCATCCTTTTCCTTACTATGG	-630-750	52	63-53
BDNF 750R*		x	CTATCTTCCCCTTTTAATGGTC		50.6	
BDNF 634R	x		CTTTTAATGGTCAGTGTACATAC		55.6	
NXC-1 1F*		x	GACCTTGGTCCMAGNACCATT	-600-700	57.2	55-45
NXC-1 64F	x		CCTGATGGGGAAATAAGGAGGATT		56.1	
NXC-1 690R*		x	TSACTGCTTTCCTTGCYTG		54.4	
NXC-1 649R	x		GATCTGCGGCATGTCTCTTTA		54.8	
NXC-1 617R	x		GCACCAGTCATCAGGCGA		57.8	
CytB 1F*		x	ACMGGHYTMTTYYTRGCHATRCAYTA	-600-650	57.3	55-45
CytB 27F	x		ACAGCHGAYACWTCAATAGC		52	
CytB 652R*		x	TADGCRAAWAGRAARTAYCAYTCNGG		55.6	
CytB 623R	x		GGGTTKGCTGGKGTGAARTT		57	
ND4 -13F	x	x	CHAAAGCYCAYGTWGAAGCCC	-550-650	57.2	55-45
ND4 1F*			CACCTTATGACTACCAAAAGCTCATGTAGAAGC		60	
ND4 26F	x		TAGAAGCCCCAATTGCCGG		57.9	
ND4 650R*		x	TATTAGGAGATGTTCTCG		44.8	
ND4 642R	x		GGTGTTCTCGAATTTGAAAGGG		52	
ND4 557R	x		CGTTGTGTKGATCAWAGRGTRTAKA		54.2	

Table 3

Data partitions and their best-fit evolutionary models for sequence evolution using PartitionFinder2 allowing models to vary by locus and codon position (15 possible partitions).

Partition	Locus & Codon Position	Model
1	SIA_pos1 + BDNF_pos2	GTR+I+G
2	SIA_pos2	JC69
3	SIA_pos3	GTR+I+G
4	BDNF_pos1	GTR+I+G
5	BDNF_pos3	HKY
6	NXC_pos1	GTR+I+G
7	NXC_pos2	GTR+I+G
8	NXC_pos3	GTR+G
9	CYTB_pos1	GTR+I+G
10	CYTB_pos2	GTR+I+G
11	CYTB_pos3 + ND4_pos3	GTR+G
12	ND4_pos1	GTR+I+G
13	ND4_pos2	GTR+I+G

Furthermore, we mapped anuran lifestyles (Table 1 supplementary materials; one of arboreal, scansorial, terrestrial, fossorial, and semi-aquatic) onto the phylogeny (Fig. 5), which indicated a general trend of niche conservatism throughout the clade [1]. Code to reproduce all analyses with all input files are provided in the GitHub repository.

Fig. 5. The evolution of Asterophryinae lifestyle. Lifestyles were mapped onto the 218 taxa time calibrated Bayesian inference phylogeny with ancestral states reconstructed using maximum likelihood. Pies on each node show the probabilities of lifestyles at ancestral nodes generated from 2000 stochastic mappings.

2. Experimental Design, Materials and Methods

2.1. Sampling Design

We built upon a previous sequencing dataset [15] adding an additional 62 taxa to produce the largest dataset to date of 205 candidate taxa of Asterophryinae and three outgroup taxa [1]. We conducted new field collections of four New Guinea island sites and two offshore island sites (Normanby Island; site detail provided in Table 1 in the supplementary materials). Collections from the Central Highlands added samples to a region with low sampling. Furthermore, we included multiple samples of nominal species with "widespread distributions", including at least one sample per site. We collected locality data from museum databases and identified 19 sites (18 multispecies) with reasonably complete species sampling.

2.2. Primer Design

The previous dataset had significant missing data for some of the loci, in particular ND4 and NXC-1. We redesigned all primers (Table 2) for improved PCR performance in order to complete the dataset for all species. Nuclear primers were based on anuran and vertebrate sequences downloaded from GenBank [16] using NCBI Primer-BLAST software [2]. Mitochondrial sequences for Asterophryinae were downloaded from GenBank and aligned with previous sequencing efforts to design degenerate primers using HYDEN [17] with default settings.

2.3. PCR Strategies to Improve Data Completeness

Standard PCR protocols were described in [1]. Several modifications were used to improve recovery of difficult targets. Touchdown PCR used annealing temperatures starting at 5 °C above Tm (Table 2) and decreasing by 1 °C per cycle for the first 10 cycles, followed by 25 cycles at the lowest annealing temperature. Some primer combinations for NXC-1, ND4, and CYTB required a touchdown window of 55–45 °C, whereas SIA and BDNF primers were improved by windows of 66–56 °C or 63–53 °C, respectively. Difficult targets were temperature optimized. A 15° touchdown window was used for CytB, as necessary, to improve yield likely due to the degeneracy or the greater Tm disparity between forward and reverse primers. Yield was improved for faint PCR bands with five additional amplification cycles. In other cases, a nested PCR strategy was effective with an initial PCR reaction using outer primers as the template for a nested or semi-nested reaction. PCR products were cleaned using Monarch spin columns (NEB). Several samples produced two bands for ND4, in these cases, bands of the correct length were excised from the gel and purified prior to Sanger sequencing. Sanger sequencing was conducted at the University of Hawaii at Manoa's Advanced Studies of Genomics, Proteomics and Bioinformatics facility (https://www.hawaii.edu/microbiology/asgpb/).

2.4. Increased Data Quality

We improved data quality by increasing data vetting beyond standard quality score criteria (>40%) and BLAST confirmation of loci. To eliminate the possibility of amplifying non-homologous targets such as pseudogenes, we furthermore instituted a practice of confirmation of suspect sequences by resequencing any data that resulted unexpected placement of taxa in single-locus phylogenies, sometimes with greater PCR stringency, as well as scrutinizing pairwise alignments of the data resulting in unexpected distance scores.

Sequence data were aligned after translation into amino acid sequences using MUSCLE [4], then back translated to nucleotides for phylogenetic reconstruction. Alignments provided in the Supplementary Materials.

2.5. Phylogenetic Analysis

A more complex evolutionary model was used for phylogenetic analysis as compared to previous efforts. The best model was identified by simultaneously optimizing evolutionary models and partitions (by locus and codon) using PartitionFinder2 [5], and contained 13 partitions for the 5 locus dataset [(Table 3). Trees were time-calibrated using the timing of six geological events separating sister taxa (see [1] and file asterophryinae_dates.txt in Supplementary Materials). The maximum likelihood (ML) tree was estimated using IQTREE [7] with nodal support based on 2000 bootstrap replicates. The Bayesian inference (BI) tree was estimated using BEAST2 [6] with two independent Monte Carlo (MCMC) runs performed for 100,000,000 generations with sampling every 10,000 generations and 30% burn-in. Input files for these analyses are provided in Supplementary Materials. Further details of phylogenetic reconstruction are given in [1]. We checked for the possibility of morphologically cryptic species by examining mitochondrial divergence and phylogenetic distance [see 1].

2.6. Assembly of Metadata

Collection site data including GPS and elevation information were collected using museum databases for our voucher specimens as well as data collected from our own fieldwork. Samples were mapped to terranes (major geological sectors) using their GPS coordinates. Lifestyle data were collected based on microhabitat use (perch type) and behavior data from literature research and our own fieldwork. Briefly, frogs were classified as: arboreal if found on tree trunks or canopy, more than 2m above the ground; scansorial if found on shrubs less than 2m above the ground; terrestrial if generally found on the ground on the forest floor or among leaf litter; fossorial if found underground in holes or burrows; and semi-aquatic if found associated with streams and swimming for mode of escape. All metadata, including collection site, museum number and locus-specific GenBank numbers are assembled in Table 1 of the supplementary materials.

We evaluated the designation of a "collection site" empirically. Traditionally, field workers in New Guinea designate sites or "field camps" at various distances based on their perceptions of variation in the species under study. We analyzed spatial clustering of samples by calculating the distances between collection sites from their GPS coordinates using the geodist function in the R package geodist [10]. A spatial cutoff of 1.1 km corresponded well with collection sites designated by field workers, as well as with distinct species in our phylogeny (with the exception of cryptic species complexes).

2.7. Ancestral Reconstruction of Lifestyle

We reconstructed ancestral states for lifestyle along the phylogeny using Maximum Likelihood (ML) with the fitDiscrete function in the R package geiger [8]. The best fit ML model assumed a symmetric transition rate between lifestyles (AICc=309) and was used to generate 2000 stochastic mappings of lifestyle onto the phylogeny using SIMMAP [18] with the make.simmap function in the R package phytools [9]. A majority of nodes (excepting 10) were reconstructed with >= 70% support for a single lifestyle (Fig. 5).

Phylogenetic plots were created using the R packages ggtree, ouch and treeio [11–13], and with a custom script based on modifications of ouch. All analyses were conducted in the R statistical computing environment unless otherwise noted [14]. Code for all analyses are provided in the supplementary file "asterophryinae_phylogeny_analysis_figures.R". All supplementary materials are available in the repository (https://doi.org/10.5281/zenodo.7063167).

Ethics Statements

No human subjects were used in this study. Animals used in this study were euthanized using MS-222 following UH IACUC protocol 12–1458 issued to M. Butler.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Data Availability

Molecular and phylogenetic datasets for the Asterophryinae frogs of New Guinea with additional data on lifestyle, geography, and elevation (Original data) (DOI: 10.5281/zenodo.7063167).

CRediT Author Statement

Ethan C. Hill: Formal analysis, Data curation, Methodology, Project administration, Writing – original draft, Investigation, Writing – review & editing, Visualization; **Mary J. Jarman:** Investigation, Writing – review & editing; **Claire J. Fraser:** Investigation, Data curation, Writing – original draft, Writing – review & editing; **Diana F. Gao:** Investigation, Writing – review & editing, Visualization; **Elizabeth R. Henry:** Investigation, Resources, Data curation; **Allison R. Fisher:** Data curation, Writing – review & editing; **Bulisa Iova:** Investigation, Resources; **Allen Allison:** Conceptualization, Methodology, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing; **Marguerite A. Butler:** Conceptualization, Methodology, Formal analysis, Data curation, Supervision, Investigation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition.

Acknowledgments

Director of the PNG National Museum and Art Gallery, for use of facilities. We are grateful for excellent critical comments from Alex Slavenko, Paul Oliver, Dave Carlon, and three anonymous reviewers that greatly improved the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2023.108987.

References

[1] E.C. Hill, C.J. Fraser, D.F. Gao, M.J. Jarman, E.R. Henry, B. Iova, A. Allison, Resolving the deep phylogeny: Implications for early adaptive radiation, cryptic, and present-day ecological diversity of papuan microhylid frogs, Mol. Phylogenetics Evol. 177 (2022) 107618, doi:10.1016/j.ympev.2022.107618.

[2] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-blast: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinform. 13 (134) (2012), doi:10.1186/1471-2105-13-134.

[3] C. Linhart, R. Shamir, Degenerate primer design: theoretical analysis and the hyden program, Methods Mol. Biol. 402 (2007) 221–244, doi:10.1007/978-1-59745-528-2_11.

[4] R.C. Edgar, Muscle: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinform. 5 (113) (2004), doi:10.1186/1471-2105-5-113.

[5] R. Lanfear, P.B. Frandsen, A.M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses, Mol. Biol. Evol. (2016) 10.1093/molbev/msw260.

[6] R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.H. Wu, D. Xie, M.A. Suchard, A. Rambaut, A.J. Drummond, Beast 2: a software platform for bayesian evolutionary analysis, PLoS Comput. Biol. 10 (4) (2014) e1003537, doi:10.1371/journal.pcbi.1003537.

[7] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. Von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol. 37 (2020) 1530–1534, doi:10.1093/molbev/msaa015.

[8] M.W. Pennell, J.M. Eastman, G.J. Slater, J.W. Brown, J.C. Uyeda, R.G. FitzJohn, M.E. Alfaro, L.J. Harmon, Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees, Bioinformatics 30 (15) (2014) 2216–2218.

[9] L.J. Revell, Phytools: an r package for phylogenetic comparative biology (and other things), Methods Ecol. Evol. 3 (2) (2012) 217–223, doi:10.1111/j.2041-210X.2011.00169.x.

[10] M. Padgham, M.D. Sumner, Geodist: fast, dependency-free geodesic distance calculations, 2021, R package version 0.0.7. https://CRAN.R-project.org/package=geodist.

[11] A.A. King, M.A. Butler, Ouch: ornstein-uhlenbeck models for phylogenetic comparative hypotheses. R package, version 2.18. https://kingaa.github.io/ouch/.

[12] L.G. Wang, T.T.Y. Lam, S. Xu, Z. Dai, L. Zhou, T. Feng, P. Guo, C.W. Dunn, B.R. Jones, T. Bradley, H. Zhu, Y. Guan, Y. Jiang, G. Yu, Treeio: an r package for phylogenetic tree input and output with richly annotated and associated data, Mol. Biol. Evol. (2019), doi:10.1093/molbev/msz240.

[13] G. Yu, Using ggtree to visualize data on tree-like structures, Curr. Protoc. Bioinform. 69 (1) (2020) e96, doi:10.1002/cpbi.96.

[14] R Core TeamR: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022 https://www.R-project.org/.

[15] J.A. Rivera, F. Kraus, A. Allison, M.A. Butler, Molecular phylogenetics and dating of the problematic new guinea microhylid frogs (amphibia: Anura) reveals elevated speciation rates and need for taxonomic reclassification, Mol. Phylogenet. Evol. 112 (2017) 1–11, doi:10.1016/j.ympev.2017.04.008.

[16] E.W. Sayers, E.E. Bolton, J.R. Brister, K. Canese, J. Chan, D.C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B.W. Trawick, K.D. Pruitt, S.T. Sherry, Database resources of the national center for biotechnology information, Nucleic Acids Res. 50 (D1) (2022) D20–D26, doi:10.1093/nar/gkab1112.

[17] C. Linhart, R. Shamir, The degenerate primer design problem, Bioinformatics 18 (Suppl 1) (2002) S172–S181, doi:10.1093/bioinformatics/18.suppl_1.s172.

[18] J.P. Bollback, Simmap: stochastic character mapping of discrete traits on phylogenies, BMC Bioinform. 7 (1) (2006) 88 –88.

[19] MA Butler, AA King, Phylogenetic comparative analysis: a modeling approach for adaptive evolution, American Naturalist 164 (2004) 683–695, doi:10.1086/426002.