Molecular Phylogenetics and Evolution 198 (2024) 108130

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Genomic SNPs resolve the phylogeny of an ancient amphibian island radiation from the Seychelles

Miranda B. Sherlock [a,b,*], Jeffrey W. Streicher [a], David J. Gower [a,c], Simon T. Maddock [a,c,d], Ronald A. Nussbaum [a,e], Oommen V. Oommen [f], Ana Serra Silva [a,b,g], Julia J. Day [a,b], Mark Wilkinson [a,b]

[a] Natural History Museum, Cromwell Road, London SW7 5BD, UK
[b] Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK
[c] Island Biodiversity and Conservation Centre, University of Seychelles, Mahé, Seychelles
[d] School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK
[e] Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1079, USA
[f] Department of Zoology, University of Kerala, Kariavattom 695 581, Thiruvananthapuram, Kerala, India
[g] School of Earth Sciences, University of Bristol, Queens Road, Bristol BS8 1RL, UK

ARTICLE INFO

Keywords:
ddRADseq
Phylogenomics
Ancient clade
Island biogeography
Caecilian

ABSTRACT

Unusually for oceanic islands, the granitic Seychelles host multiple lineages of endemic amphibians. This includes an ancient (likely ca. 60 million years) radiation of eight caecilian species, most of which occur on multiple islands. These caecilians have a complicated taxonomic history and their phylogenetic inter-species relationships have been difficult to resolve. Double-digest RAD sequencing (ddRADseq) has been applied extensively to phylogeography and increasingly to phylogenetics but its utility for resolving ancient divergences is less well established. To address this, we applied ddRADseq to generate a genome-wide SNP panel for phylogenomic analyses of the Seychelles caecilians, whose phylogeny has so far not been satisfactorily resolved with traditional DNA markers. Based on 129,154 SNPs, we resolved deep and shallow splits, with strong support. Our findings demonstrate the capability of genome-wide SNPs for evolutionary inference at multiple taxonomic levels and support the recently proposed synonymy of *Grandisonia* Taylor, 1968 with *Hypogeophis* Peters, 1879. We revealed three clades of *Hypogeophis* (large-, medium- and short-bodied) and identify a single origin of the diminutive, stocky-bodied and pointy-snouted phenotype.

1. Introduction

The core of the Seychelles archipelago in the Western Indian Ocean is a Gondwanan microcontinent that separated from India ~ 63 *Mya* (million years ago) (Collier et al., 2008; estimates summarised in Gower et al., 2016) and encompasses several montane granitic islands, the largest of which is ~ 150 km^2 (Gardner, 1986; Labisko et al., 2022). Fluctuations in sea level during the Pleistocene changed connectivity between the islands and imposed range restrictions on fauna and flora through changes in exposed island area. The last lowstand on the Seychelles was ~ 21 *Kya* (thousand years ago) during the Last Glacial Maximum (LGM), when sea level was 120 (±5) metres below its present level (Miller et al., 2005), which would have connected all major granitic islands. Prior to this, the Penultimate Glacial Period (PGP) caused the last major drop ~ 140 *Kya* (Colleoni et al., 2016). Such processes make the Seychelles an excellent natural laboratory for studies of biogeography, vicariance and inter-island gene flow. The impacts of these processes are particularly pronounced for terrestrial organisms that cannot easily disperse across sea water, such as amphibians.

The Seychelles are home to thirteen endemic amphibian species, including a radiation of caecilians (Gymnophiona). Distributions of these species vary across the islands, which are of different sizes and distances from each other (Labisko et al., 2022). This provides several opportunities to study the historical biogeography of the Seychelles. Seychellean caecilians comprise a clade of eight species (Labisko et al., 2022; Maddock et al., 2018) in the family Grandisoniidae Lescure, Renous, and Gasc, 1986, which also includes two African species and a clade of peninsular Indian species (Gower et al., 2011; Wilkinson et al.,

* Corresponding author at: Natural History Museum, Cromwell Road, London SW7 5BD, UK.
E-mail address: miranda.sherlock.13@ucl.ac.uk (M.B. Sherlock).

https://doi.org/10.1016/j.ympev.2024.108130
Received 31 October 2023; Received in revised form 5 June 2024; Accepted 15 June 2024
Available online 16 June 2024
1055-7903/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2011). However, a confident understanding of the diversification of these caecilians is lacking because their taxonomic history is complex, and their phylogenetic relationships incompletely resolved. Until recently, Seychelles caecilians were classified in three endemic genera, but molecular phylogenies have repeatedly indicated that the most-recently described genus, *Grandisonia* Taylor, 1968, is paraphyletic with respect to *Hypogeophis* Peters, 1879 (Gower et al., 2011; Gwilt et al., 2021; Loader et al., 2007; Maddock et al., 2016; San Mauro et al., 2014; Wilkinson et al., 2003, 2002), leading to the proposed synonymy of *Grandisonia* with *Hypogeophis* (Dubois et al., 2021), which we follow here. Thus, we recognise the Seychelles caecilians as comprising two endemic genera, *Praslinia* Boulenger, 1908 which is monotypic and *Hypogeophis* which includes seven species. Seychellean endemics featured prominently in early molecular phylogenetic studies of caecilians, including the first mitochondrial ribosomal (rRNA) phylogeny (Hedges et al., 1993). However, subsequent analyses of these markers (Loader et al., 2007; Wilkinson et al., 2003, 2002), of mitochondrial genomes (Gwilt et al., 2021; Maddock et al., 2016; San Mauro et al., 2014; Zhang and Wake, 2009) and of combinations of mitochondrial and nuclear genes (Gower et al., 2016, 2011; Maddock et al., 2018; San Mauro et al., 2012) have produced multiple alternative topologies with low support for conflicting groups. Despite this, some inferred relationships are consistent amongst studies; *Praslinia cooperi* is resolved as the sister taxa to all other Seychelles caecilians, and *Hypogeophis larvatus* and *H. sechellensis* are resolved as sister taxa (Gower et al., 2011; Gwilt et al., 2021; Loader et al., 2007; Maddock et al., 2016; San Mauro et al., 2014; Wilkinson et al., 2003, 2002). The lack of a well-supported phylogeny for the Seychelles caecilians is a barrier to understanding their evolution and historical biogeography. The evolution of Seychelles caecilians is of particular interest given their status as the only extant, truly oceanic-island caecilians, their diversity of life-history modes (San Mauro et al., 2014), and morphological diversity that includes three of the smallest extant caecilian species with the fewest vertebrae (Maddock et al., 2018).

High-resolution, genomic scale, single nucleotide polymorphisms (SNPs) offer hope for resolving Seychellean caecilian phylogeny, but it is unclear how useful such data are for inferring ancient divergences (>40 *Mya*) including any within this radiation. Using methods such as double digest restriction-site associated DNA sequencing (ddRADseq; Peterson et al., 2012), SNP datasets are commonly applied to assess phylogenetic relationships of rapidly radiated or recently diverged groups (e.g., Ambu et al., 2023; Bombonato et al., 2020; Ford et al., 2015; Streicher et al., 2014). At deeper evolutionary timescales, the use of RADseq to collect sufficient numbers of orthologous SNPs for phylogenetic inference is made difficult by allelic dropout, in which loci are unsampled due to mutations occurring in the enzyme recognition site (Leaché and Oaks, 2017). However, RADseq data are being increasingly applied to more deeply diverged groups. For example, alignments from RADseq-derived SNPs have been used to resolve species-level relationships in divergences up to ~55 *Mya* in phrynosomatid lizards (Leaché et al., 2015), and 50 *Mya* in flowering plants (Eaton et al., 2017).

For amphibians, these challenges of applying RADseq-derived SNPs are compounded by their large genomes (e.g., Liedtke et al., 2018), which may make recovering enough homologous loci for phylogenetic inference at deeper evolutionary time scales difficult (Rodríguez et al., 2017). These issues can be mitigated by selecting a narrower size range for loci, sequencing at greater depth and by using double-digest RAD library preparation to increase the likelihood that the same locus is sequenced in multiple individuals (Rodríguez et al., 2017; Salas-Lizana and Oono, 2018). Despite the large genomes of amphibians, RADseq phylogenomic studies have resolved genus- to species-level relationships, including up to 10 *Mya* in *Alytes* midwife toads (Ambu et al., 2023) and 16 *Mya* in *Leptopelis* tree frogs (Reyes-Velasco et al., 2018). Diversity in divergence times, significant mito-nuclear discordance, large-genome size and recalcitrant nodes make the Seychelles caecilians an interesting case to further challenge the effectiveness of RADseq-derived SNPs in resolving species-level relationships, which motivated our current study. Here, we aimed to resolve the species-level relationships of the caecilians of the Seychelles while also testing the efficacy of RADseq-derived SNPs in resolving phylogenetic relationships in a group with wide-ranging divergences (<63 *Mya*; Gower et al., 2016; Maddock et al., 2018).

2. Materials and methods

2.1. Sample collection, DNA extraction and ddRAD library construction

A total of 38 samples were sequenced: two individuals of *Hypogeophis montanus*, three individuals of *H. brevis*, five individuals of *H. sechellensis* plus four individuals of each of the other five Seychelles caecilian species and two outgroup species (Table 1). The two outgroup species, *Gegeneophis ramaswamii* (endemic to the Western Ghats of peninsular India) and *Idiocranium russeli* (endemic to Cameroon) have varying inferred divergence times from the ingroup of approximately 79–103 *Mya* (Roelants et al., 2007; Wilkinson et al., 2002; Zhang and Wake, 2009) and 130–191 *Mya* (Zhang and Wake, 2009), respectively. DNA was extracted using SeraPure paramagnetic beads (Rohland and Reich, 2012) following a protocol detailed in Lambert et al. (2019) or with a DNeasy Blood & Tissue kit (Qiagen) following the manufacturers protocol. Using a protocol modified from Peterson et al. (2012) samples were cleaned and digested using the restriction enzymes MspI and Sbf1, indexed (by ligating an adapter with a 5′ barcode) and PCR amplified (with a 5′ barcoded PCR-primer). Changes from this protocol are detailed in the methods and supplementary material of Lambert et al. (2019). Libraries were sequenced on a single 300 cycle Illumina NextSeq 500 run using a high-output kit at the Natural History Museum, London. All demultiplexed illumina read data generated for this study has been deposited at the NCBI Sequence Read Archive (SRA) with the BioProject number PRJNA1034088. Raw illumina read data and matrices used in this study are deposited at the UCL Research Data Repository under the project DOI: 10.5522/04/24459628.

2.2. SNP genotyping and filtering

Resulting fastQ data files of sequenced libraries were checked in FastQC v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and demultiplexed in STACKS v.2.52 (Rochette and Catchen, 2017). Individual fastQ files were trimmed and filtered in Trimmomatic v.0.39 (Bolger et al., 2014). Reads were assembled, mapped and SNP genotyped in dDocent v.2.94 (Puritz et al., 2014) with paired-end assembly, a sequence identity threshold of 0.9 and BWA mapping parameters 1, 4 and 6 for the match score value, mismatch score, and gap opening penalty respectively.

Although there are potentially adverse consequences for improper clustering of putatively orthologous loci, empirical testing using other amphibian RAD datasets found that inferred topologies and their resolution were robust to changes of the intra-sample clustering threshold (Rancilhac et al., 2023). Among the taxa included in tests by Rancilhac et al. (2023), a clustering threshold of 0.89–0.91 maximised the number of parsimony informative sites in the assembly. We therefore applied a clustering threshold of 0.9 to maximise the number of loci recovered without over-splitting, although it should be noted that the crown age of Seychelles caecilians is much older than of the taxa tested by Rancilhac et al. (2023) and so these parameters may not be optimal for our taxa.

For reference contig assembly, we set a minimum coverage threshold of three, both for sequences within individuals and for the number of individuals representing unique sequences. To appropriately balance the possibility of including erroneous SNPs with that of losing variants unique to single species (and therefore a few individuals only) we filtered using VCFtools v.0.1.16 (Danecek et al., 2011) for a minor allele count of two, minimum read quality of 30, minimum genotype call frequency 0.5 (i.e. only sites that are present in 50 % of individuals) and

Table 1
Table of sample information including the proportion of unsampled SNPs present in the final alignment. BMNH = Natural History Museum; UK = University of Kerala; UMMZ = University of Michigan Museum of Zoology. sf = sequencing failed.

Sample ID	Species	Locality	Raw forward read no.	Missing data %	Voucher prefix	Voucher code	NCBI SRA no.
RAN 31756	*Hypogeophis rostratus*	Praslin, Seychelles	4,054,922	17.1	UMMZ	200500	SRR26876954
RAN 31329	*Hypogeophis rostratus*	Mahé, Seychelles	8,597,062	13.3	UMMZ	193026	SRR26876955
SM 326	*Hypogeophis rostratus*	Praslin, Seychelles	6,171,580	14.8	BMNH	2005.1814	SRR26876953
SM 376	*Hypogeophis rostratus*	Mahé, Seychelles	924,721	32.0	BMNH	2005.1810	SRR26876952
SM 296	*Hypogeophis alternans*	Praslin Seychelles	2,104,230	15.6	BMNH	2005.1700	SRR26876974
SM 318	*Hypogeophis alternans*	La Digue, Seychelles	9,486,954	9.2	BMNH	2005.1676	SRR26876962
SM 397	*Hypogeophis alternans*	Silhouette, Seychelles	7,530,203	10.7	BMNH	2005.1709	SRR26876951
SM 303	*Hypogeophis alternans*	Praslin, Seychelles	3,792,203	17.0	BMNH	2005.1701	SRR26876973
SM 238	*Hypogeophis larvatus*	Silhouette, Seychelles	4,589,528	17.5	BMNH	2005.1732	SRR26876942
SM 473	*Hypogeophis larvatus*	Praslin, Seychelles	16,625,293	14.5	BMNH	2005.1730	SRR26876939
SM 351	*Hypogeophis larvatus*	Mahé, Seychelles	3,328,155	22.0	BMNH	2005.1717	SRR26876940
SM 316	*Hypogeophis larvatus*	La Digue, Seychelles	3,164,344	20.9	BMNH	2005.1715	SRR26876941
SM 379	*Hypogeophis sechellensis*	Mahé, Seychelles	3,757,841	16.0	BMNH	2005.1752	SRR26876967
SM 278	*Hypogeophis sechellensis*	Praslin, Seychelles	7,355	sf	BMNH	2005.1755	SRR26876969
SM 249	*Hypogeophis sechellensis*	Silhouette, Seychelles	2,071,655	22.7	BMNH	2005.1758	SRR26876970
SM 338	*Hypogeophis sechellensis*	Mahé, Seychelles	6,663,689	12.5	BMNH	2005.1749	SRR26876968
SM 495	*Hypogeophis sechellensis*	Mahé, Seychelles	7,608,685	11.9	BMNH	2005.1691	SRR26876966
SM 472	*Praslinia cooperi*	Mahé, Seychelles	10,709,450	35.9	BMNH	2005.1842	SRR26876943
SM 470	*Praslinia cooperi*	Mahé, Seychelles	8,261,022	36.9	BMNH	2005.1840	SRR26876944
SM 237	*Praslinia cooperi*	Silhouette, Seychelles	5,417,870	42.2	BMNH	2005.1843	SRR26876946
SM 469	*Praslinia cooperi*	Mahé, Seychelles	9,353,962	36.0	BMNH	2005.1839	SRR26876945
SM 354	*Hypogeophis brevis*	Mahé, Seychelles	7,134,082	17.6	BMNH	2005.1784	SRR26876965
SM 355	*Hypogeophis brevis*	Mahé, Seychelles	11,583,956	15.8	BMNH	2005.1785	SRR26876964
SM 356	*Hypogeophis brevis*	Mahé, Seychelles	38,959	90.3	BMNH	2005.1786	SRR26876963
SM 284	*Hypogeophis montanus*	Mahé, Seychelles	6,067,507	18.2	BMNH	2005.1820	SRR26876961
SM 285	*Hypogeophis montanus*	Mahé, Seychelles	7,090,710	17.1	BMNH	2005.1821	SRR26876960
SM 490	*Hypogeophis pti*	Praslin, Seychelles	3,376,578	20.1	BMNH	2005.1838	SRR26876956
SM 295	*Hypogeophis pti*	Praslin, Seychelles	3,070,575	20.2	BMNH	2005.1827	SRR26876958
SM 302	*Hypogeophis pti*	Praslin, Seychelles	4,239,994	17.3	BMNH	2005.1831	SRR26876957
SM 282	*Hypogeophis pti*	Praslin, Seychelles	5,380,432	14.2	BMNH	2005.1826	SRR26876959
MW 01293	*Gegeneophis ramaswamii*	Shonlode, India	1,463,154	87.2	UK	MW 01292	SRR26876938
MW 01522	*Gegeneophis ramaswamii*	Vanchuvam, India	1,404,878	87.4	UK	MW 01521	SRR26876937
MW 01565	*Gegeneophis ramaswamii*	Pathanapuram, India	2,404,680	85.0	UK	MW 01561	SRR26876972
MW 02170	*Gegeneophis ramaswamii*	Maraimalai, India	66,440	97.5	UK	MW 02169	SRR26876971
MW 08365	*Idiocranium russeli*	Tinta, Cameroon	2,557,735	94.4	BMNH	2008.688 (8364)	SRR26876950
MW 08481	*Idiocranium russeli*	Tinta, Cameroon	986,191	96.0	BMNH	2008.688 (8480)	SRR26876949
MW 08508	*Idiocranium russeli*	Makamune, Cameroon	3,556,272	93.5	BMNH	2008.688 (8350)	SRR26876948
MW 08608	*Idiocranium russeli*	Makamune, Cameroon	5,499,169	93.1	BMNH	2008.688 (8607)	SRR26876947

minimum mean read depth of two. There was no maximum threshold for missing data per individual.

2.3. Phylogenomic analyses

The Python script *VCF2phylip* v.2.6 (Ortiz, 2019) was used to convert the filtered VCF to PHYLIP format for phylogenetic analyses. IQ-TREE v. 2.2.0.7 (Minh et al., 2020b) was then used to output an additional alignment without constant sites to apply the ascertainment bias correction (+ASC option). IQ-TREE considers sites that contain only a non-ambiguous code (e.g. C) and any ambiguity code also containing that base (e.g. Y or S) to be constant and therefore removes these from the alignment. Maximum likelihood (ML) analyses were run in IQ-TREE. We used the inbuilt ModelFinder (Kalyaanamoorthy et al., 2017) to run model selection in IQ-TREE on PHYLIP files with variable sites only (MFP + ASC models) and the alignment including both variable and invariant (according to IQ-TREE) sites (MFP + I models). The ultra-fast bootstrap 2 (Thi Hoang et al., 2017) was applied with 10,000 iterations and with hill-climbing nearest neighbour interchange (option −bnni) and Shimodaira-Hasegawa approximate likelihood ratio testing (option −alrt) with 1000 iterations to avoid over-estimation of branch support values (Guindon et al., 2010). Site concordance factors (sCF) were calculated with 10,000 quartets to measure discordance in our dataset by the percentage of decisive sites supporting the resultant topology (Minh et al., 2020a). These analyses were replicated for different permitted levels of missing data per locus (0 %, 5 %, 10 % and 80 %) to determine whether analyses converged on the same topology regardless of the amount of missing data.

BEAST 2 v.2.7.1 (Bouckaert et al., 2014) was also applied to generate a maximum clade credibility (MCC) tree using Bayesian inference, to assess whether our inferred trees were robust to methodological differences. BEAST 2 was run on the unprocessed PHYLIP file produced from the filtered VCF (with no restrictions for ambiguity codes), with three independent runs of the Markov chain each with 20 million generations. We assessed any possible incongruence in the IQ-TREE bootstrap trees by visualising the alternate splits in SplitsTree v.5.0.0 (Dress and Huson, 2004; Huson and Bryant, 2006).

To account for the possibility that any discordance in our trees is due to hybridisation (given the possibility for historic migration between islands) or incomplete lineage sorting (ILS), we applied the species tree reconstruction method SVDQuartets (Chifman and Kubatko, 2014) implemented in PAUP* v.4.0a (Swofford, 2003). Analysis was run on a NEXUS formatted alignment (with no restrictions on ambiguity codes), using exhaustive sampling (38,528 quartets) and with 100 bootstrap replicates. This tree was used as the starting tree for network analysis with SNaQ implemented in PhyloNetworks v.0.16.3 (Solís-Lemus et al., 2017). The table of quartet concordance factors (CF) used as the input was produced using the R function SNPs2CF (www.github.com/melisaolave/SNPs2CF, see Olave and Meyer, 2020). Concordance factors were calculated between species and subsampling 50 alleles per species quartet (n.quartets = 50), with 100 bootstrap replicates. The resulting 10,500 quartets were used to infer the network, testing up to three reticulation events (h_{max} = 0:3) with 10 runs each.

3. Results

3.1. RADseq data processing and filtering

For the 38 individuals sequenced, a total of 191,494,502 reads with a phred score greater than 30 were obtained. The number of forward reads per individual ranged from 38,959 to 16,625,293, excluding one *H. sechellensis* individual (SM278) with only 7355 (forward) reads which was removed from further analyses. After coverage cut-offs were applied, 580,565 sequences remained that were assembled into 130,195 contigs. Across all individuals 1,581,926 variants were called, which following filters applied in VCFtools left 257,353 sites (170,379 were SNPs and the remainder indels). Of the 257,353 sites, 99.97 % had a mean read depth > 5 and 95.58 % >10. The average percentage of sites unsampled per individual was 22.4 % for ingroup samples and 91.8 % for outgroups, but this varied greatly from 9.2–90.3 % for ingroups and 85.0–96.0 % for outgroups (Table 1).

3.2. Tree topology, support and convergence

The PHYLIP formatted alignment for phylogenetic inference in IQ-TREE comprised 129,154 variable sites. The best model according to the Bayesian Information Criterion (BIC) was TVM + F + ASC + R2. The ML bootstrap majority-rule consensus tree was fully resolved and generally highly supported (Fig. 1) with all species-level branches (i.e., those subtending species) having maximal bootstrap proportions and SH-aLRT values. Site concordance factors showed good support, with no branches having an sCF value lower than the 33 % expected by chance alone, and most branches having sCF values over 50 % (where the majority of sites support that topology) (Fig. 1). The ML tree placed *P. cooperi* as the sister taxa to all other ingroup species. Three main clades within *Hypogeophis* were recovered: (*H. rostratus* + *H. alternans*), (*H. sechellensis* + *H. larvatus*) and (*H. pti* + *H. montanus* + *H. brevis*). Additionally, *H. brevis* and *H. montanus* were inferred to be sister taxa. The sCF support values for clades within *Hypogeophis* ranged from 36.6 % to 84.2 %. Within *H. rostratus*, there were well supported internal branches comprising samples from the northern island of Praslin (100 for UFBoot and sh-aLRT, 82.2 % for sCF) and the southern island of Mahé (100 for UFBoot and sh-aLRT, 66.0 % for sCF). The corresponding splits network (Fig. 2) further illustrates the maximal support for tree-like interspecies relationships with reticulation restricted to relationships within species. The alignment containing both variable and invariant (as per IQ-TREE's consideration of ambiguity codes) sites comprised 170,379 sites. Phylogenetic inference in IQ-TREE with this alignment yielded the same topology as with the variable sites only alignment. Apart from the lower value for the branch leading to the ingroup taxa, support was also similar (Fig. S1).

In the per-locus missing data threshold analyses, the 10 % (666 characters), 50 % (129,154 characters) and 80 % (415,334 characters) permitted missing data matrices recovered the same (inter-specific) topology. Support was similar across these trees (marginally lower for 10 % and marginally higher for 80 %). However, the lowest supported *Hypogeophis* clade for the 50 % permitted missing data tree (*H. larvatus* + *H. sechellensis* + *H. pti* + *H. montanus* + *H. brevis*) had lower support in the 80 % matrix (sCF value of 34.9 vs 36.6) (Fig. S2). Analysis of the 5 % matrix (161 characters) yielded a mostly resolved tree, but with low support. Unsurprisingly, the data matrix with no missing samples per locus (0 %, 21 characters) yielded a generally poorly resolved and supported tree. The outgroup samples had exceptionally high levels of missing SNPs (>85 %: Table 1) and intraspecific internal branches for these taxa were moderately well-supported (mean bootstrap proportions of 68.5 % for *I. russeli* and 72.5 % for *G. ramaswamii*).

The PHYLIP file without restriction for ambiguity codes for Bayesian inference in BEAST 2 contained 170,379 SNPs. The ML topology for the full dataset was largely congruent with the 95 % MCC tree generated in BEAST 2 (Fig. S3), except that the diminutive, short-bodied Seychelles species (*H. pti* + *H. brevis* + *H. montanus*) were sister to (*H. rostratus* + *H. alternans*) in the BEAST 2 tree rather than to (*H. sechellensis* + *H. larvatus*) in the ML tree, but this had lower support (posterior probability of 0.92) than other relationships upon which the two trees agree. The multi-species coalescent tree from SVDQuartets inferred the same species-level topology as the ML trees, with maximum support, except for the branch subtending the clade (*H. larvatus* + *H. sechellensis* + *H. pti* + *H. montanus* + *H. brevis*) which had a bootstrap value of 97 % (Fig. S4). The phylogenetic network identified the optimal number of hybridisation events to be two ($h = 2$) (Fig. S5, Table S1), (*G. ramaswamii*: $\gamma = 0.153$; (*H. brevis* + *H. montanus*): $\gamma = 0.0729$; Fig. 3).

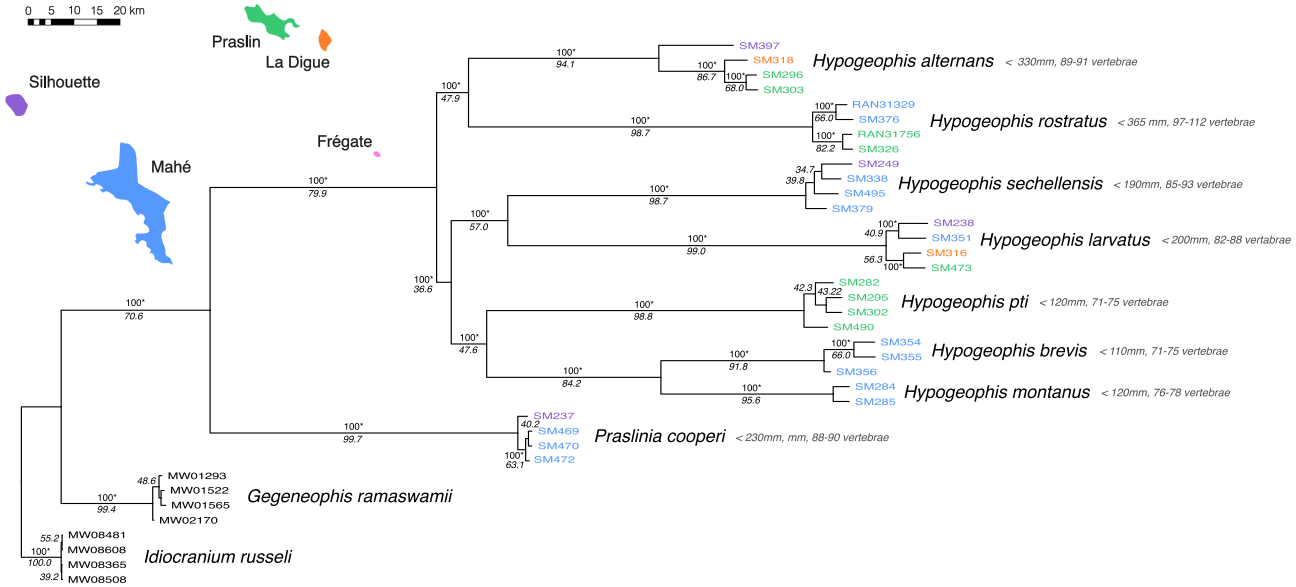

Fig. 1. Majority-rule consensus of the maximum likelihood tree inferred with IQ-TREE. Full UFBoot and SH-aLRT branch support is labelled with '100*' (lower support not shown), sCF values are in italics (values less than 33.3 not shown). Inset map of the Seychelles granitic islands that are occupied by multiple caecilian species. Coloured sample IDs represent the island where the sample was collected. Approximate maximum length and vertebrae numbers given in italics.

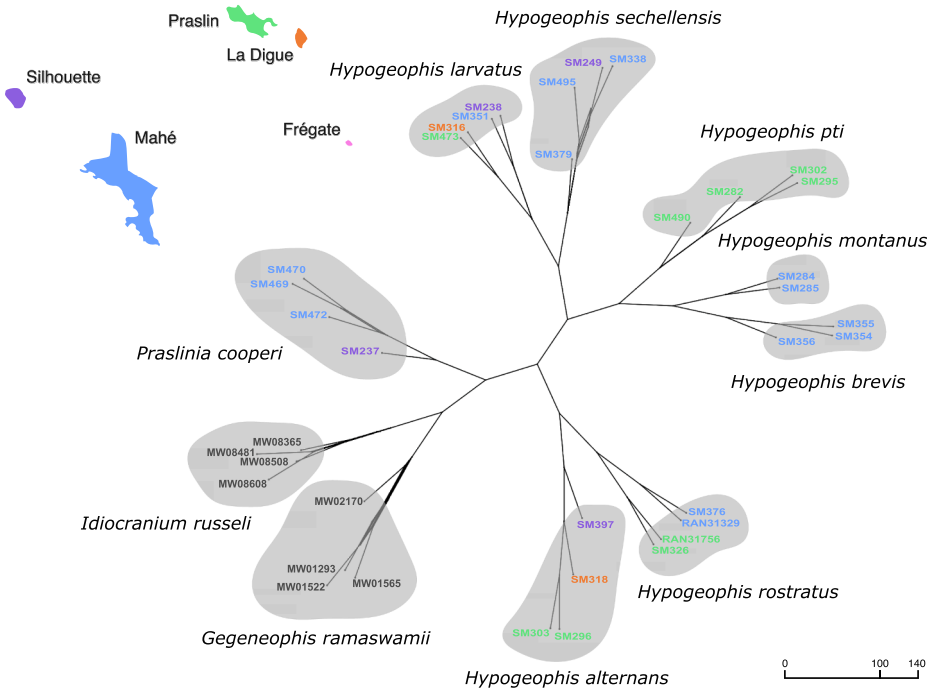

Fig. 2. Splits network of IQ-TREE bootstrap trees. Scale bar indicates substitutions per site. Sample IDs are coloured to indicate the island where the sample was collected (see inset map of the Seychelles).

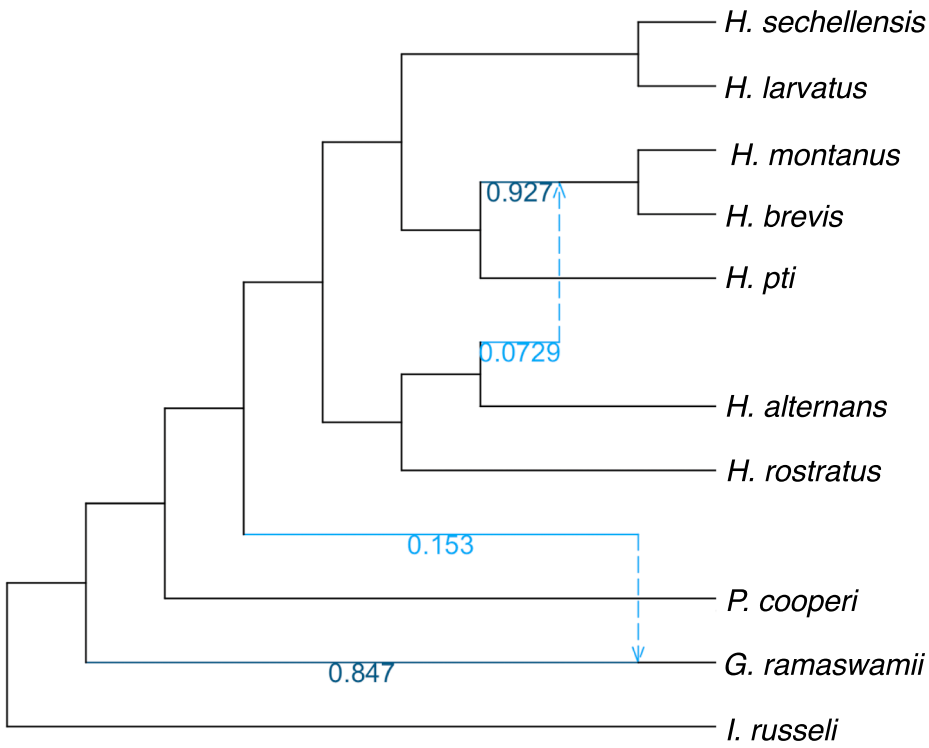

Fig. 3. The optimal network inferred in the PhyloNetworks analysis. Blue edges indicated inferred hybridisation events. Numbers adjacent to blue lines are γ inheritance values representing the proportion of the genome inherited from the ancestor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Our results add to a growing number of studies (e.g., Ambu et al., 2023; Bombonato et al., 2020; Secci-Petretto et al., 2023) that demonstrate the usefulness of RADseq SNP data for reconstructing relationships across different temporal and taxonomic scales. Our analyses of RADseq-derived SNPs were able to infer well-supported phylogenetic relationships of grandisoniid caecilians ranging from young (within species) to much older (between ingroup and outgroups > 60 *Mya*) divergences. The ML SNP tree is a fully resolved and highly

supported topology that overcomes previous persistent difficulties with incongruence, low-support and incomplete sampling of taxa in inferring interspecific relationships among Seychelles caecilians. Some degree of discordance within *Hypogeophis* is indicated by sCF values. The lowest sCF (36.6, Fig. 1) is for the branch that subtends the clade (*H. larvatus* + *H. sechellensis* + *H. pti* + *H. montanus* + *H. brevis*). We consider this lower concordance to be likely explained by the correlation between branch length and sCF (Arcila et al., 2021; Burbrink et al., 2020) and thus the higher probability of ILS in this short branch.

The two putative hybridising edges inferred by the network analysis (Fig. 3) explain some of the discordant signal, but both are unexpected and merit further investigation. Hybridisation between the common ancestor of *Hypogeophis* and the *Gegeneophis ramaswamii* lineage implies transoceanic dispersal between Seychelles and India. Taken at face value, this seems implausible but may have occurred when these land masses were closer together. Understanding the timing and plausibility of this event would benefit from analyses including additional Indian grandisoniids. Hybridisation between the lineage leading to *H. alternans* and the common ancestor of *H. brevis* + *H. montanus* is also perhaps implausible given substantial differences in body size, but further basic knowledge of ecology and life history would be helpful in assessing plausibility in this case. Taxa involved in the putative hybridisation events had some of the highest missing data levels in our study with *G. ramaswamii* having a mean of 89.3 % missing data (n = 4) and one of the *H. brevis* (SM356) samples having 90.3 % missing data (Table 1) suggesting that the impact of missing data on the detection of introgression merits further exploration.

Consistent with almost all previous phylogenetic analyses of Seychelles caecilians, *P. cooperi* is inferred to be the sister taxon to all other Seychelles species, and *H. larvatus* and *H. sechellensis* are inferred to be sister species (e.g., Gower et al., 2011; Gwilt et al., 2021; Loader et al., 2007; Maddock et al., 2016; San Mauro et al., 2014; Wilkinson et al., 2003, 2002). The ingroup topology includes three well-supported main clades within *Hypogeophis,* (*H. alternans* + *H. rostratus*), (*H. larvatus* + *H. sechellensis*), and (*H. pti* + *H. montanus* + *H. brevis*), which correspond to large, medium, and small-bodied phenotypes, respectively. The interrelationships of these three clades is the least well supported part of the phylogeny, evident from the low sCFs observed on the branches connecting these three clades (Fig. 1) and the alternative topology recovered in the BEAST 2 analysis where *H. rostratus* + *H. alternans* is the sister clade of *H. brevis* + *H. montanus* (Fig. S3). Additionally, a divide between north and south island lineages that has previously been reported for *H. rostratus* based on mitochondrial and nuclear loci (Maddock et al., 2020) is further supported by the Mahé and Praslin branches on the ML SNP tree, although this is based on only four individuals (Fig. 1).

4.1. Hypogeophis alternans and H. rostratus as sister taxa

We resolved *H. alternans* (the type species of *Grandisonia*) and *H. rostratus* (the type species of *Hypogeophis*) as sister taxa, with full support from UFBoot and SH-aLRT (100 %) and with the greatest share of support (47.9 %) from sCF (alternate topologies were supported by 30.3 and 21.8 % of sites). This grouping is further supported by morphology (these species have the largest body sizes among Seychelles caecilians: e.g., Taylor, 1968). The paraphyly of *Grandisonia* supports the proposal of Dubois et al. (2021) to transfer all *Grandisonia* species to *Hypogeophis*. Our results contrast with previous phylogenetic studies that have not found a well-supported consensus on the relationships of these species (e.g., Gwilt et al., 2021; Loader et al., 2007; San Mauro et al., 2014).

4.2. Monophyly of diminutive and abbreviated Hypogeophis

RADseq SNP data strongly support the monophyly of a group of diminutive, short-bodied and pointy-snouted species (*H. brevis*, *H. montanus* and *H. pti*), consistent with a single origin of this distinctive phenotype. This contradicts findings from analyses of mitochondrial 16S rRNA (Maddock et al., 2018) and mitochondrial genomes (Gwilt et al., 2021) in which the sampled diminutive *Hypogeophis* are not monophyletic. Within this clade, *H. brevis* and *H. montanus* are resolved as sister species supporting findings from 16S rRNA data (Maddock et al., 2018) and reflecting their adjacent distribution. *Hypogeophis brevis* and *H. montanus* occur only on the large and high southern island of Mahé and presumably diverged from each other within this island, whereas *H. pti* is found only on the northern island of Praslin (Maddock et al., 2017). On Mahé, *H. montanus* is found at higher elevations (718–731 m) than *H. brevis* (350–650 m) (Maddock et al., 2018). *Hypogeophis brevis* and *H. montanus* are difficult to distinguish based on external morphology, with *H. montanus* possessing a slightly smaller head and more primary annuli (Maddock et al., 2018). However, there is very little discordance in this part of our new tree (Fig. 2) and strong support for the reciprocal monophyly of these two species. In addition, branch lengths in the ML SNP tree indicate substantial genetic distances between *H. montanus* and *H. brevis,* as was found from limited mitochondrial data (Maddock et al., 2018).

4.3. RADseq phylogenomic inference is resistant to missing data

High levels of topological congruence together with high branch support across multiple thresholds of missing data (10 %, 50 %, and 80 %) suggest that the presence of missing data was not problematic (Fig. S2). In contrast, the repeatedly inferred sister-species relationship between *H. larvatus* and *H. sechellensis* (Gower et al., 2011; Gwilt et al., 2021; Loader et al., 2007; Maddock et al., 2016; San Mauro et al., 2014; Wilkinson et al., 2003, 2002) is not recovered at the 0 % and 5 % missing data thresholds (Fig. S2). This corroborates previous investigations with empirical (Crotti et al., 2019), simulated (Huang and Knowles, 2016) and combined (Rick et al., 2022) datasets that reveal potentially negative consequences of minimising or excluding missing data in RADseq phylogenetics. That our analyses were able to infer strongly supported relationships using a ddRADseq dataset with substantial amounts of missing data extends this observation to evolutionary timescales that are seldom studied with RADseq data. Our results are similar to those from a recent study by Ambu et al. (2023) on *Alytes* midwife toads, whose phylogeny has also been difficult to resolve with mitochondrial or nuclear genes alone or in combination but was resolved with RADseq data. As in our dataset, Ambu et al.'s (2023) result was robust in analyses of larger but more incomplete datasets, but in more stringently filtered smaller datasets topological uncertainty increased.

5. Conclusions

Our ML analyses of SNPs support commonly inferred relationships among Seychelles caecilians (*Praslinia* as the sister taxa to the other Seychelles caecilians, *H. larvatus* and *H. sechellensis* as sister species), but also resolve previously more recalcitrant ones, revealing the close relationship of the diminutive species (*H. brevis, H. montanus, H. pti*) on the one hand and the larger species (*H. alternans, H. rostratus*) on the other and confirming the paraphyly of '*Grandisonia*'. Sampling widely across the genome with RADseq phylogenomics may avoid some of the issues associated with mitochondrial gene studies, including uniparental inheritance and low representation of the entire genome, both of which can warp estimates of a species' ancestry and/or demographic history (Ballard and Whitlock, 2004). Moreover, cytonuclear discordance is prevalent in studies using both nuclear and mitochondrial genes due to, for example, introgression (see Secci-Petretto et al., 2023; Streicher and Day, 2020). Here, sensitively applied filtering has allowed relationships of disparate age to be resolved from the same dataset, which sets the scene for biogeographic and additional evolutionary investigations that rely on an understanding of how Seychelles caecilians are related. Further work on this system should investigate evolutionary processes

within Mahé, including the recentness and/or completeness of the split between *H. brevis* and *H. montanus*. In addition, comparative phylogeographic studies using fine-scale genomic data could reveal the causes of the disparate distributions of species in the Seychelles including incomplete dispersal, local extinction, and possible intra-island speciation events.

This study used reduced-representation genomic SNP data to produce the best-supported phylogeny of Seychelles caecilians to date, including well-supported clades of large-, medium- and small-bodied *Hypogeophis*. However the interrelationships of these three clades remains less certain, likely due to a short internal branch. The use of whole-genome sequencing (WGS) could dramatically increase the number of phylogenetically informative markers, although given the large genome sizes of Seychelles caecilians (Liedtke et al., 2018), assembly and analysis of these genomes could be problematic (Zhang et al., 2019). There are also some interesting biogeographical patterns in the distributions of the Seychelles caecilians (e.g. single-island species) that could be better understood with molecular dating of the phylogeny, however the requisite nuclear mutation rate is not yet known for caecilians. In addition, morphological data should be investigated for possible synapomorphies of the large-, medium- and small-bodied *Hypogeophis* clades identified here.

CRediT authorship contribution statement

Miranda B. Sherlock: Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization. **Jeffrey W. Streicher:** Writing – review & editing, Resources, Methodology, Conceptualization. **David J. Gower:** Writing – review & editing, Resources. **Simon T. Maddock:** Writing – review & editing, Resources, Conceptualization. **Ronald A. Nussbaum:** Resources. **Oommen V. Oommen:** Resources. **Ana Serra Silva:** Writing – review & editing, Methodology, Funding acquisition. **Julia J. Day:** Writing – review & editing, Supervision, Resources, Conceptualization. **Mark Wilkinson:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

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.

Acknowledgments

This work was funded by the Natural Environment Research Council (NE/S007229/1 and NE/L002434/1), which supported MBS and ASS, respectively. We thank R. Bristol, L. Chong-Seng, R. Kamei, C. Morel, M. Pfrender and C. Raxworthy for sample collection. We would also like to thank G. Schneider for access to specimens stored at UMMZ and the Seychelles Bureau of Standards for providing field collection and export permits. We thank C. Griffin, A. Hall, E. Lugli and S. Walkington for assistance in the NHM molecular laboratory. We thank the Associate Editor Melisa Olave, and two anonymous reviewers for their helpful comments on this manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108130.

References

Ambu, J., Martínez-Solano, Í., Suchan, T., Hernandez, A., Wielstra, B., Crochet, P.-A., Dufresnes, C., 2023. Genomic phylogeography illuminates deep cyto-nuclear discordances in midwife toads (*Alytes*). Mol. Phylogenet. Evol. 183, 107783 https://doi.org/10.1016/J.YMPEV.2023.107783.

Arcila, D., Hughes, L.C., Meléndez-Vazquez, B., Baldwin, C.C., White, W.T., Carpenter, K. E., Williams, J.T., Santos, M.D., Pogonoski, J.J., Miya, M., Ortí, G., Betancur-R, R., 2021. Testing the Utility of Alternative Metrics of Branch Support to Address the Ancient Evolutionary Radiation of Tunas, Stromateoids, and Allies (Teleostei: Pelagiaria). Syst. Biol. 70, 1123–1144. https://doi.org/10.1093/SYSBIO/SYAB018.

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744. https://doi.org/10.1046/J.1365-294X.2003.02063.X.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170.

Bombonato, J.R., do Amaral, D.T., Silva, G.A.R., Khan, G., Moraes, E.M., da Silva Andrade, S.C., Eaton, D.A.R., Alonso, D.P., Ribolla, P.E.M., Taylor, N., Zappi, D., Franco, F.F., 2020. The potential of genome-wide RAD sequences for resolving rapid radiations: a case study in Cactaceae. Mol. Phylogenet. Evol. 151, 106896. doi: 10.1016/J.YMPEV.2020.106896.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for bayesian evolutionary analysis. PLoS Comput. Biol. 10 https://doi.org/10.1371/JOURNAL.PCBI.1003537.

Burbrink, F.T., Grazziotin, F.G., Pyron, R.A., Cundall, D., Donnellan, S., Irish, F., Keogh, J.S., Kraus, F., Murphy, R.W., Noonan, B., Raxworthy, C.J., Ruane, S., Lemmon, A.R., Lemmon, E.M., Zaher, H., 2020. Interrogating Genomic-Scale Data for Squamata (Lizards, Snakes, and Amphisbaenians) Shows no Support for Key Traditional Morphological Relationships. Syst. Biol. 69, 502–520. https://doi.org/10.1093/sysbio/syz062.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30 (23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Colleoni, F., Wekerle, C., Näslund, J.O., Brandefelt, J., Masina, S., 2016. Constraint on the penultimate glacial maximum Northern Hemisphere ice topography (≈140 kyrs BP). Quat. Sci. Rev. 137, 97–112. https://doi.org/10.1016/J.QUASCIREV.2016.01.024.

Collier, J.S., Sansom, V., Ishizuka, O., Taylor, R.N., Minshull, T.A., Whitmarsh, R.B., 2008. Age of Seychelles-India break-up. Earth Planet Sci. Lett. 272, 264–277. https://doi.org/10.1016/J.EPSL.2008.04.045.

Crotti, M., Barratt, C.D., Loader, S.P., Gower, D.J., Streicher, J.W., 2019. Causes and analytical impacts of missing data in RADseq phylogenetics: Insights from an African frog (*Afrixalus*). Zool. Scr. 48, 157–167. https://doi.org/10.1111/zsc.12335.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/BIOINFORMATICS/BTR330.

Dress, A.W.M., Huson, D.H., 2004. Constructing splits graphs. IEEE/ACM Trans. Comput. Biol. Bioinform. 1, 109–115. https://doi.org/10.1109/TCBB.2004.27.

Dubois, A., Ohley, A., Pyron, R.A., 2021. New concepts and methods for phylogenetic taxonomy and nomenclature in zoology, exemplified by a new ranked cladonomy of recent amphibians (Lissamphibia). Megataxa 5. https://doi.org/10.11646/megataxa.5.1.1.

Eaton, D.A.R., Spriggs, E.L., Park, B., Donoghue, M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. 66, 399–412. https://doi.org/10.1093/sysbio/syw092.

Ford, A.G.P., Dasmahapatra, K.K., Rüber, L., Gharbi, K., Cezard, T., Day, J.J., 2015. High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. Mol. Ecol. 24, 3421–3440. https://doi.org/10.1111/MEC.13247.

Gardner, A.S., 1986. The Biogeography of the Lizards of the Seychelles Islands. J. Biogeogr. 13, 237–253. https://doi.org/10.2307/2844923.

Gower, D.J., Mauro, D.S., Giri, V., Bhatta, G., Govindappa, V., Kotharambath, R., Oommen, O.V., Fatih, F.A., Mackenzie-Dodds, J.A., Nussbaum, R.A., Biju, S.D., Shouche, Y.S., Wilkinson, M., 2011. Molecular systematics of caeciliid caecilians (Amphibia: Gymnophiona) of the Western Ghats. India. Mol. Phylogenet. Evol. 59, 698–707. https://doi.org/10.1016/j.ympev.2011.03.002.

Gower, D.J., Agarwal, I., Karanth, K.P., Datta-Roy, A., Giri, V.B., Wilkinson, M., San Mauro, D., 2016. The role of wet-zone fragmentation in shaping biodiversity patterns in peninsular India: Insights from the caecilian amphibian *Gegeneophis*. J. Biogeogr. 43, 1091–1102. https://doi.org/10.1111/jbi.12710.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/SYSBIO/SYQ010.

Gwilt, A.J., Streicher, J.W., Maddock, S.T., 2021. Mitogenomic analyses provide further evidence for multiple miniaturization events during the evolution of Seychelles caecilian amphibians. bioRxiv 2021.03.24.436854. doi: 10.1101/2021.03.24.436854.

Hedges, S.B., Nussbaum, R.A., Maxson, L.R., 1993. Caecilian Phylogeny and Biogeography Inferred from Mitochondrial DNA Sequences of the 12S rRNA and 16S rRNA Genes (Amphibia: Gymnophiona). Herpetol. Monogr. 7, 64–76. https://doi.org/10.2307/1466952.

Huang, H., Knowles, L.L., 2016. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of rad sequences. Syst. Biol. 65 (3), 357–365. https://doi.org/10.1093/sysbio/syu046.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol. Evol. 23, 254–267. https://doi.org/10.1093/MOLBEV/MSJ030.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 2017 14:6 14, 587–589. doi: 10.1038/nmeth.4285.

Labisko, J., Maddock, S.T., Rocha, S., Gower, D.J., 2022. Cryptic Lineages among Seychelles Herpetofauna. In: Cryptic Species. Cambridge University Press, pp. 242–280. https://doi.org/10.1017/9781009070553.010.

Lambert, S.M., Streicher, J.W., Fisher-Reid, M.C., Méndez de la Cruz, F.R., Martínez-Méndez, N., García-Vázquez, U.O., Montes, N., de Oca, A., Wiens, J.J., 2019. Inferring introgression using RADseq and DFOIL: Power and pitfalls revealed in a case study of spiny lizards (*Sceloporus*). Mol. Ecol. Resour 19, 818–837. https://doi.org/10.1111/1755-0998.12972.

Leaché, A.D., Oaks, J.R., 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 48, 69–84. doi: 10.1146/annurev-ecolsys-110316-022645.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D., Linkem, C.W., 2015. Phylogenomics of Phrynosomatid Lizards: Conflicting Signals from Sequence Capture versus Restriction Site Associated DNA Sequencing. Genome Biol. Evol. 7, 706–719. https://doi.org/10.1093/GBE/EVV026.

Liedtke, H.C., Gower, D.J., Wilkinson, M., Gomez-Mestre, I., 2018. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat. Ecol. Evol. 2, 1792–1799. https://doi.org/10.1038/s41559-018-0674-4.

Loader, S.P., Pisani, D., Cotton, J.A., Gower, D.J., Day, J.J., Wilkinson, M., 2007. Relative time scales reveal multiple origins of parallel disjunct distributions of African caecilian amphibians. Biol. Lett. 3, 505–508. https://doi.org/10.1098/RSBL.2007.0266.

Maddock, S.T., Briscoe, A.G., Wilkinson, M., Waeschenbach, A., San Mauro, D., Day, J.J., Littlewood, D.T.J., Foster, P.G., Nussbaum, R.A., Gower, D.J., 2016. Next-generation mitogenomics: A comparison of approaches applied to caecilian amphibian phylogeny. PLoS One 11, e0156757.

Maddock, S.T., Wilkinson, M., Nussbaum, R.A., Gower, D.J., 2017. A new species of small and highly abbreviated caecilian (Gymnophiona: Indotyphlidae) from the Seychelles island of Praslin, and a recharacterization of *Hypogeophis brevis* Boulenger, 1911. Zootaxa 4329, 301–326. https://doi.org/10.11646/ZOOTAXA.4329.4.1.

Maddock, S.T., Wilkinson, M., Gower, D.J., 2018. A new species of small, long-snouted Hypogeophis Peters, 1880 (Amphibia: Gymnophiona: Indotyphlidae) from the highest elevations of the Seychelles island of Mahé. Zootaxa 4450, 359–375. https://doi.org/10.11646/ZOOTAXA.4450.3.3.

Maddock, S.T., Nussbaum, R.A., Day, J.J., Latta, L., Miller, M., Fisk, D.L., Wilkinson, M., Rocha, S., Gower, D.J., Pfrender, M.E., 2020. The roles of vicariance and isolation by distance in shaping biotic diversification across an ancient archipelago: Evidence from a Seychelles caecilian amphibian. BMC Evol. Biol. 20, 1–20. https://doi.org/10.1186/S12862-020-01673-W/TABLES/3.

Miller, K.G., Kominz, M.A., Browning, J.V., Wright, J.D., Mountain, G.S., Katz, M.E., Sugarman, P.J., Cramer, B.S., Christie-Blick, N., Pekar, S.F., 2005. The phanerozoic record of global sea-level change. Science 310, 1293–1298. https://doi.org/10.1126/SCIENCE.1116412/SUPPL_FILE/MILLER_SOM.PDF.

Minh, B.Q., Hahn, M.W., Lanfear, R., 2020a. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol. Evol. 37, 2727–2733. https://doi.org/10.1093/MOLBEV/MSAA106.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., Teeling, E., 2020b. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/MOLBEV/MSAA015.

Olave, M., Meyer, A., 2020. Implementing large genomic single nucleotide polymorphism data sets in phylogenetic network reconstructions: a case study of particularly rapid radiations of cichlid fish. Syst. Biol. 69, 848–862. https://doi.org/10.1093/sysbio/syaa005.

Ortiz, E.M., 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. doi: 10.5281/ZENODO.2540861.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7, e37135.

Puritz, J.B., Hollenbeck, C.M., Gold, J.R., 2014. dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2014, e431.

Rancilhac, L., Sylvestre, F., Hutter, C.R., Arntzen, J.W., Babik, W., Crochet, P.A., Deso, G., Duguet, R., Galan, P., Pabijan, M. and Policain, M., 2023. Exploring the impact of read clustering thresholds on RADseq-based systematics: an empirical example from European amphibians. bioRxiv, 2023-04.

Reyes-Velasco, J., Manthey, J.D., Freilich, X., Boissinot, S., 2018. Diversification of African tree frogs (genus *Leptopelis*) in the highlands of Ethiopia. Mol. Ecol. 27, 2256–2270. https://doi.org/10.1111/MEC.14573.

Rick, J.A., Brock, C.D., Lewanski, A.L., Golcher-Benavides, J., Wagner, C.E., 2022. Reference genome choice and filtering thresholds jointly influence phylogenomic analyses. Syst. Biol. syad065 https://doi.org/10.1093/sysbio/syad065.

Rochette, N.C., Catchen, J.M., 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nat. Protoc. 12, 2640–2659. https://doi.org/10.1038/nprot.2017.123.

Rodríguez, A., Burgon, J.D., Lyra, M., Irisarri, I., Baurain, D., Blaustein, L., Göçmen, B., Künzel, S., Mable, B.K., Nolte, A.W., Veith, M., Steinfartz, S., Elmer, K.R., Philippe, H., Vences, M., 2017. Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. Mol. Phylogenet. Evol. 115, 16–26. https://doi.org/10.1016/J.YMPEV.2017.07.009.

Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Biju, S.D., Guillaume, K., Moriau, L., Bossuyt, F., 2007. Global patterns of diversification in the history of modern amphibians. PNAS. 104, 887–892. https://doi.org/10.1073/PNAS.0608378104/SUPPL_FILE/08378TABLE4.XLS.

Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946. https://doi.org/10.1101/GR.128124.111.

Salas-Lizana, R., Oono, R., 2018. Double-digest RADseq loci using standard Illumina indexes improve deep and shallow phylogenetic resolution of *Lophodermium*, a widespread fungal endophyte of pine needles. Ecol. Evol. 8, 6638–6651. https://doi.org/10.1002/ECE3.4147.

San Mauro, D., Gower, D.J., Cotton, J.A., Zardoya, R., Wilkinson, M., Massingham, T., 2012. Experimental Design in Phylogenetics: Testing Predictions from Expected Information. Syst. Biol. 61, 661–674. https://doi.org/10.1093/sysbio/sys028.

San Mauro, D., Gower, D.J., Müller, H., Loader, S.P., Zardoya, R., Nussbaum, R.A., Wilkinson, M., 2014. Life-history evolution and mitogenomic phylogeny of caecilian amphibians. Mol. Phylogenet. Evol. 73, 177–189. https://doi.org/10.1016/J.YMPEV.2014.01.009.

Secci-Petretto, G., Englmaier, G.K., Weiss, S.J., Antonov, A., Persat, H., Denys, G.P.J., Schenekar, T., Romanov, V.I., Taylor, E.B., Froufe, E., 2023. Evaluating a species phylogeny using ddRAD SNPs: Cyto-nuclear discordance and introgression in the salmonid genus *Thymallus* (Salmonidae). Mol. Phylogenet. Evol. 178, 107654.

Solís-Lemus, C., Bastide, P., Ané, C., 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34 (12), 3292–3298. doi: 10.1093/molbev/msx235 doi: 10.1016/J.YMPEV.2022.107654.

Streicher, J.W., Day, J.J., 2020. The toad's warts: Discordance creates bumpy expectations of mitochondrial-nuclear evolution between species. Mol. Ecol. 29, 3400–3402. https://doi.org/10.1111/mec.15568.

Streicher, J.W., Devitt, T.J., Goldberg, C.S., Malone, J.H., Blackmon, H., Fujita, M.K., 2014. Diversification and asymmetrical gene flow across time and space: Lineage sorting and hybridization in polytypic barking frogs. Mol. Ecol. 23, 3273–3291. https://doi.org/10.1111/mec.12814.

Swofford, D., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachussets.

Taylor, E.H., 1968. The Caecilians of the World. University of Kansas Press, Lawrence.

Thi Hoang, D., Chernomor, O., von Haeseler, A., Quang Minh, B., Sy Vinh, L., Rosenberg, M.S., 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol 35, 518–522. https://doi.org/10.5281/zenodo.854445.

Wilkinson, M., Sheps, J.A., Oommen, O.V., Cohen, B.L., 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol. Phylogenet. Evol. 23, 401–407.

Wilkinson, M., Loader, S.P., Gower, D.J., Sheps, J.A., Cohen, B.L., 2003. Phylogenetic relationships of african caecilians (amphibia: gymnophiona): Insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. 52, 83–92. https://doi.org/10.1080/21564574.2003.9635483.

Wilkinson, M., Mauro, D.S., Sherratt, E., Gower, D.J., 2011. A nine-family classification of caecilians (Amphibia: Gymnophiona). Zootaxa 2874, 41–64.

Zhang, F., Ding, Y.H., Zhu, C.D., Zhou, X., Orr, M.C., Scheu, S., Luan, Y.X., 2019. Phylogenomics from low-coverage whole-genome sequencing. Methods Ecol. Evol. 10, 507–517. https://doi.org/10.1111/2041-210X.13145.

Zhang, P., Wake, M.H., 2009. A mitogenomic perspective on the phylogeny and biogeography of living caecilians (Amphibia: Gymnophiona). Mol. Phylogenet. Evol. 53, 479–491. https://doi.org/10.1016/j.ympev.2009.06.018.