BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Research

Cite this article: Davis HR, Sanford HT, Das I, Nashriq I, Leaché AD. 2024 Establishing species boundaries in Bornean geckos. *Biol. Lett.* **20**: 20240157.
https://doi.org/10.1098/rsbl.2024.0157

Received: 26 March 2024
Accepted: 10 July 2024

Subject Category:
Evolutionary biology

Subject Areas:
evolution, taxonomy and systematics

Keywords:
systematics, gene flow, speciation, squamata, Borneo

Author for correspondence:
Hayden R. Davis
e-mail: hrdavis1@uw.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7397758.

THE ROYAL SOCIETY PUBLISHING

Establishing species boundaries in Bornean geckos

Hayden R. Davis[1], Henry T. Sanford[1], Indraneil Das[2], Izneil Nashriq[2] and Adam D. Leaché[1]

[1]Department of Biology, Burke Museum of Natural History and Culture, University of Washington, Seattle, WA 98195, USA
[2]Institute of Biodiversity and Environmental Conservation, Universiti of Malaysia, Sarawak, Kota Samarahan, Malaysia

HRD, 0000-0003-1401-1221; ADL, 0000-0001-8929-6300

Species delimitation using mitochondrial DNA (mtDNA) remains an important and accessible approach for discovering and delimiting species. However, delimiting species with a single locus (e.g. DNA barcoding) is biased towards overestimating species diversity. The highly diverse gecko genus *Cyrtodactylus* is one such group where delimitation using mtDNA remains the paradigm. In this study, we use genomic data to test putative species boundaries established using mtDNA within three recognized species of *Cyrtodactylus* on the island of Borneo. We predict that multi-locus genomic data will estimate fewer species than mtDNA, which could have important ramifications for the species diversity within the genus. We aim to (i) investigate the correspondence between species delimitations using mtDNA and genomic data, (ii) infer species trees for each target species, and (iii) quantify gene flow and identify migration patterns to assess population connectivity. We find that species diversity is overestimated and that species boundaries differ between mtDNA and nuclear data. This underscores the value of using genomic data to reassess mtDNA-based species delimitations for taxa lacking clear species boundaries. We expect the number of recognized species within *Cyrtodactylus* to continue increasing, but, when possible, genomic data should be included to inform more accurate species boundaries.

1. Background

Species delimitation, the process of determining whether two populations belong to the same or distinct species, is an inherently complex task that depends on the operational, methodological and philosophical species definitions. The proliferation of genomic data has reinvigorated the need for detailed investigations into the taxonomy of diverse species spanning the tree of life. This has coincided with the development of coalescent-based approaches for species delimitation that can reconcile genome-wide genealogical discordance [1–6], thus advancing beyond methods that rely on single gene trees [7–9]. Together, these data and methods have advanced species delimitation on the operational and methodological fronts. Yet, identifying the boundary between populations and species for many taxa remains challenging [10], a problem often referred to as the 'grey zone' of speciation [11,12].

For genetically divergent yet morphologically cryptic lineages, species delimitation can prove particularly difficult as the sources of data for integrative approaches are limited [13,14]. Due to a lack of diagnostic morphological characters in cryptic taxa, a heavier reliance is placed on genetic data [13–15]. However, for many researchers, the cost of genomic data remains financially intractable, resulting in most taxonomic studies of cryptic

taxa depending on DNA barcoding or similar approaches, which have lower accuracy due to their reliance on the signal from a single locus [16–22].

To explore whether putative species identified using mtDNA data are supported by genomic data, we focused on a group of related species within the highly diverse gecko genus *Cyrtodactylus*. The genus is the third most diverse among vertebrates with over 350 recognized species, yet few studies have explored evolutionary patterns within the group using genomic data [23–25]. Further, no studies have directly assessed whether the current methods applied to delimit species within gekkonids are effectively delimiting species-level diversity, which is particularly problematic when testing delimitation hypotheses for cryptic taxa. We focus on three independently derived species from the Southeast Asian island of Borneo (*C. consobrinus*, *C. miriensis* and *C. pubisulcus*) that each have evidence for additional unrecognized cryptic species based on mtDNA data [26,27]. We use this study as an opportunity to investigate the correspondence between mtDNA and genomic species delimitations.

Whether explicitly stated or not, the species concept held by a researcher plays a critical role in species delimitation. We view species as independent evolutionary lineages, which is a broadly shared view compatible with most modern species concepts [28–30], and we also see great value in considering species as reproductive communities emerging from the past [10]. In our study, we focus solely on genetic data, as former studies have demonstrated the lack of morphological distinction within the target species [25,26]. We analyse the mtDNA data using approaches often used to identify putative *Cyrtodactylus* species to provide a comparison to current practices and analyse the genomic data using multi-species coalescent methods that incorporate gene flow estimates. Using the demographic parameters from these analyses, we estimate the genealogical divergence index (*gdi*), a metric that helps disentangle species and population boundaries [31,32]. Using these approaches, we demonstrate that mtDNA and genomic-based delimitations often differ, and our results emphasize the value of using genomic data to test species boundaries, especially for cryptic lineages.

2. Methods

Fieldwork was conducted in Borneo in the Malaysian state of Sarawak between 2014 and 2018 (Permits: NPW.907.4.4.(Jld.14)-79; (119)JHS/NCCD/600-7/2/107). We aimed to collect specimens of *C. consobrinus*, *C. miriensis* and *C. pubisulcus* from as many unique localities as possible, but heavy deforestation and minimal road access to field sites have resulted in patchy sampling for these species. We sampled each species from multiple locations and obtained multiple specimens per locality.

To delimit species using mtDNA, we assembled and aligned NADH dehydrogenase subunit 2 (ND2) data. We generated an assembly for *Cyrtodactylus* species in Borneo, comprising two evolutionarily distinct clades referred to herein as the small-bodied (*C. cavernicolus*, *C. miriensis*, *C. pubisulcus*, *C. hantu*) and large-bodied (*C. consobrinus*, *C. hutan*, *C. kapitensis*, *C. malayanus*) clades. Sample sizes for the target species are as follows: *C. miriensis* = 19, *C. pubisulcus* = 20 and *C. consobrinus* = 9. We used the assembly to apply commonly used species delimitation methods (bPTP; mPTP; GMYC; ASAP), calculated pairwise distances (p-distance) using SPDEL [33], and estimated genealogies using BEAST$_2$ [34].

For genomic data, we prepared double digest restriction site-associated DNA sequencing (ddRADseq) [35] libraries. Sample sizes for the target species are as follows: *C. miriensis* = 26, *C. pubisulcus* = 32 and *C. consobrinus* = 15. To maximize the number of loci and reduce missing data, we produced separate assemblies for the small and large-bodied clades (electronic supplementary material, table S1). From the small- and large-bodied assemblies, we used branching in IPYRAD [36] to generate species-specific assemblies from their respective clades (electronic supplementary material, table S1). For population genetic analyses, we filtered each dataset using VCFTOOLS [37] to allow only one single nucleotide polymorphism (SNP) per locus (--thin 50) and filtered out variable sites present in <5% of individuals (--maf 0.05).

Using a reduced dataset for computational efficiency, we inferred time-calibrated species trees using SNAPP within the BEAST$_2$ [34] framework for both the small- and large-bodied datasets. We applied secondary calibrations to the roots of each tree using the snapp_prep ruby script, which allows the generation of an .xml file with a molecular clock [38]. For the large-bodied dataset, we constrained the crown age of the tree to 18.83 Ma with a normal distribution ($\sigma = 2$), and for the small-bodied clade, we constrained the crown age to 25.80 Ma with a normal distribution ($\sigma = 2$). We applied dates inferred from a previous Bornean *Cyrtodactylus* study [27]. To test for phylogenetic structure within populations, we concatenated the RAD loci and inferred a phylogeny using IQ-TREE [39].

For the population-based analyses, we analysed the SNP data using three analytical frameworks to explore the processes driving population divergence and to delimit species. These approaches included structure inference, estimation of migration surfaces, and species delimitation and gene flow estimation using the MSC-M model [40]. We estimated population structure using principal component analysis (PCA) and ADMIXTURE [41] and assessed models with varying population numbers (K) for biological reality.

To identify geographic areas with increased migration (corridors) and reduced migration (barriers), we used the estimated effective migration surface (EEMS) method [42]. Quantifying migration and population connectivity are pertinent to the exploration of spatial patterns of genetic diversity. EEMS identifies deviations from population structure expected under a model of isolation-by-distance (IBD). Additionally, we directly tested for IBD using the R package ADEGENET [43]. By considering the combined results from mtDNA genealogies, population structure and migration surfaces, we established population boundaries used to conduct demographic modelling and phylogenetic tests for gene flow.

To conduct joint species delimitation and gene flow quantification, we used the Python wrapper HHSD [44] to run BPP with the MSC-M model [40] and subsequently used the demographic values to calculate *gdi* values [31,32]. We started with species guide trees obtained by running the A01 step in BPP (when more than two tips were tested), and then used HHSD to estimate population sizes (Θ), species divergence times (τ) and gene flow rate (M; where $M = mN$). Conducting these tests in BPP

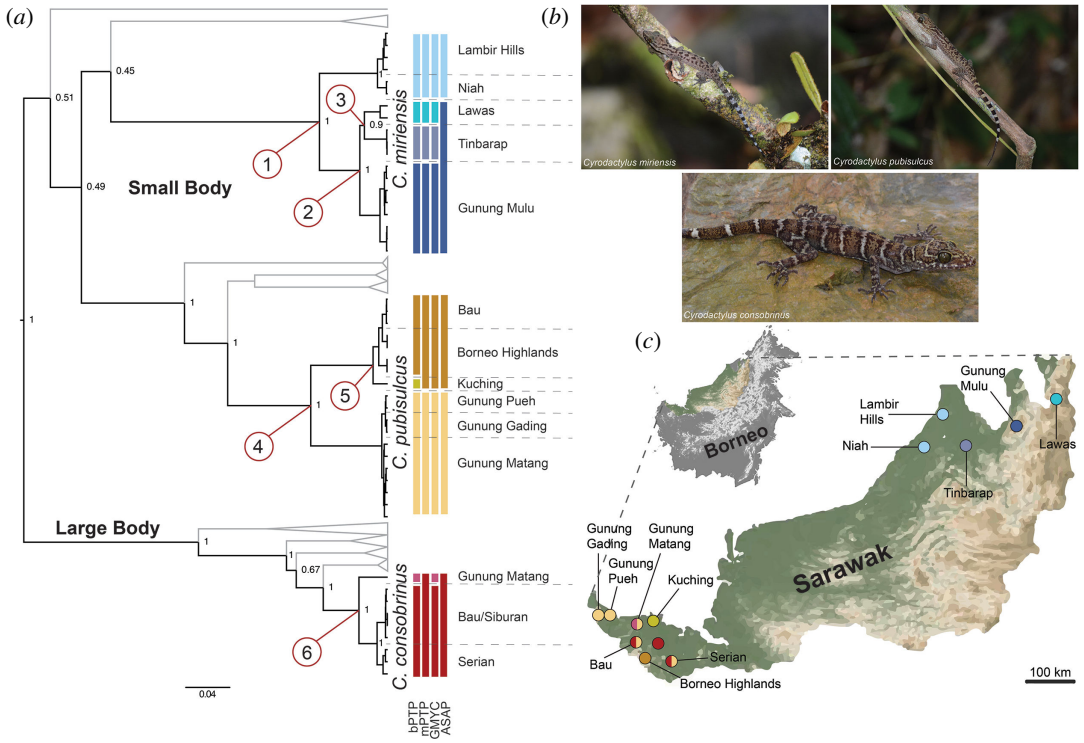

Figure 1. (*a*) mtDNA gene tree highlighting the three *Cyrtodactylus* species of focus in this study. Values on the nodes are posterior probabilities. Vertical bars on the tips of the nodes correspond to the delimitations estimated by the single-locus delimitation programs: bPTP, mPTP, GMYC and ASAP, where the colours of the bars denote the delimitation hypothesis. Node numbers 1–6 correspond to putative species of the most species-rich delimitation hypotheses. Branches for species not discussed herein are greyed out. (*b*) Photographs of live individuals of each of the three species of focus. (*c*) Map of localities from which we have sequence data, with the coloured dots corresponding to delimitations highlighted by nodes 1–6.

Table 1. Number of species estimated by each of the species delimitation programs utilized in this study for the single-locus mtDNA data, we use bPTP, mPTP, GMYC and ASAP, and for genome-wide data, we use the *gdi* metric. The min and max p-distances are for the mtDNA dataset. The min values correspond with the lowest p-distance among populations which resulted in a delimitation for any of the species. The maximum p-distance is the highest p-distance among populations that resulted in a delimitation for any of the species.

	max p-distance	min p-distance	bPTP	mPTP	GMYC	ASAP	*gdi*
C. miriensis	7.3%	3.1%	4	4	4	2	≤3
C. pubisulcus	10.0%	2.1%	3	2	2	2	≤2[a]
C. consobrinus	4.6%	4.6%	2	1	2	1	1

[a]*Cyrtodactylus pubisulcus gdi* corresponds to a different species boundary between mtDNA and SNP data.

requires *a priori* populations to be defined, which we set using the ADMIXTURE and PCA results. Because HHSD iterates through populations and merges those that do not meet the *gdi* cut-off, we tested models where *K* was one higher than the optimal number, which also enabled us to test for gene flow among disjunct geographic regions. The *gdi* value is the probability that two alleles from a population will coalesce with one another before reaching the ancestral population. The *gdi* metric spans the speciation continuum, ranging from 0 (panmictic) to 1 (genetically distinct) [32]. Values >0.2 and <0.7 are considered ambiguous. Using the combination of these methods, we identify putative species that may be representative of species-level diversity. Lastly, we converted the τ values to years using a germline mutation rate of 4.46×10^{-9} [45].

Detailed methods are available in the electronic supplementary material.

3. Results

Using mtDNA species delimitation approaches, *C. miriensis* is estimated to contain two to four species, *C. pubisulcus* from two to three species and *C. consobrinus* from one to two. Within each of these species, a high amount of diversity is present (figure 1; table 1). The pairwise distances (p-distance) within major lineages of *C. miriensis* range from 3.15 to 7.32% (min. to max.), 2.07 to 10.3% for lineages within *C. pubisulcus* and 4.31% for the single major phylogenetic division within *C. consobrinus* (figure 1). These p-distances are within the current standards for what constitutes a species by other studies using similar types of data and methods. For example, recent taxonomic work within the highly diverse *C. pulchellus*, *C. intermedius* and *C. khasiensis* species complexes has recognized new species with p-distances of ≥6%, ≥3.5% and ≥4.0%, respectively, for the ND2 locus [46–48].

Both the species trees inferred using SNAPP and the concatenated phylogeny have strong support for all species-level relationships (PP = 1; UFBoot ≥ 95; electronic supplementary material, figure S1). The divergence dating analysis supports *C.*

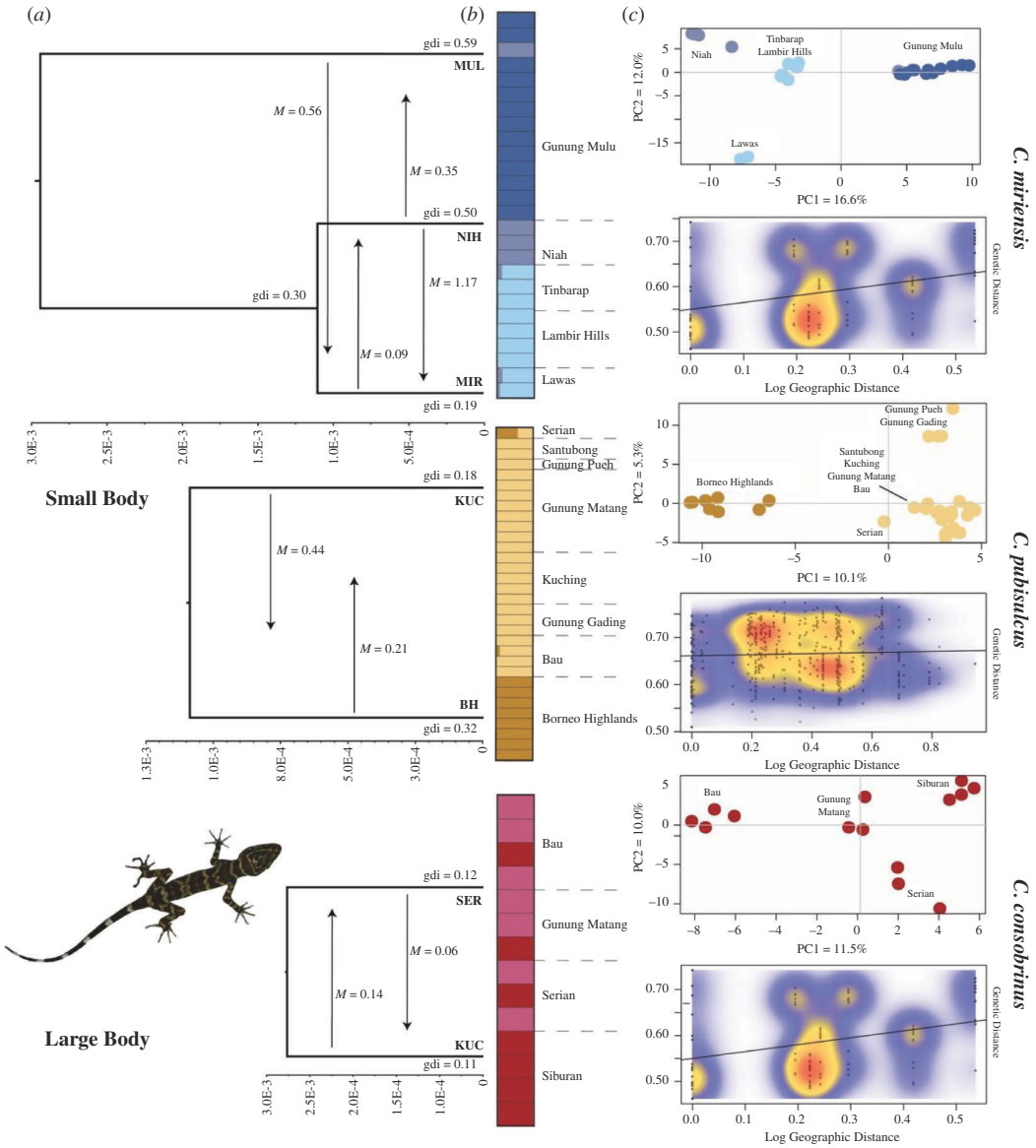

Figure 2. (*a*) Species trees from the three target species of *Cyrtodactylus* in this study. Arrows indicate the direction of gene flow, with the quantity of gene flow (*m*). Scale bars for respective species show the *tau* values. The population acronyms are shown on the tips and are referred to throughout the text. (*b*) Population structure analyses, showing the results from the *K*-value that corresponds with the number of populations used for the species tree inference. (*c*) Top: PCA of SNP data with colours corresponding to the results of the population structure analyses. The colours on the PCA for *C. consobrinus* do not correspond with the ADMIXTURE, as both mtDNA and genomic data clearly infer a single species. Two populations are shown for the species tree to highlight the *gdi* and gene flow metrics. Percentages show the per cent variation explained by the given principal component. Bottom: isolation-by-distance plots. Colours represent sampling density (white: low; red: high). The line shows the correlation between geographic and genetic distance.

miriensis as the oldest lineage (25.6 Ma (21.5–29.5 Myr)), followed by *C. pubisulcus* (14.7 Ma (12.2–16.9 Myr)) and then *C. consobrinus* (5.1 Ma (3.9–6.2 Myr)). Of note, these dates are highly reliant on commonly used fossil calibrations, which are substantially older than the dates inferred using the demographic values and a molecular substitution rate established for geckos (see below).

The PCA of the SNP data reveals that each species clusters primarily by geography (figure 2). The optimal number of populations (*K*-values) estimated using population structure inference supports fewer distinct clusters than revealed in the PCA plots. The optimal *K*-value for *C. miriensis* is uncertain with different analyses supporting $K = 1$–4. The optimal models for *C. consobrinus* and *C. pubisulcus* are $K = 1$ (electronic supplementary material, figure S2). The $K = 3$ model for *C. miriensis* structures populations geographically and closely matches the PCA groupings. Assuming a two-population model for *C. pubisulcus* also provides a clear geographic division that is reflected in the PCA (figure 2). For *C. consobrinus*, a population model with $K = 1$ is optimal, and assuming $K = 2$ demonstrates a lack of geographic structure (figure 2). The population structure and admixture results provide evidence of admixed samples in *C. pubisulcus* and *C. miriensis*. For example, within *C. pubisulcus*, one sample from the region northeast of Borneo Highlands is admixed between the Borneo Highlands and surrounding populations (Serian; figure 2).

For each species, significant levels of gene flow are detected between populations, although in some cases symmetric gene flow can be rejected (table 2). Within *C. miriensis*, one instance of bidirectional gene flow is detected ($M = 0.09$ (5.0×10^{-2} to 0.11) from MIR and NIH; $M = 1.17$ (0.93 to 1.48) from NIH and MIR)) and two instances of unidirectional gene flow ($M = 0.56$ (0.37 to 0.72) from MUL to MIR; $M = 0.35$ (0.29 to 0.41) from NIH and MUL; electronic supplementary material, table

Table 2. Species delimitation results for *Cyrtodactylus* based on genomic data. Lineages with a *gdi* ≤ 0.2 are considered conspecific and ≥ 0.7 are considered distinct species; the species status for lineages > 0.2 and < 0.7 is ambiguous. *M* 1 is the gene flow estimate from population 1 into population 2; *M* 2 is the gene flow estimate from population 2 into population 1, with the 95% confidence intervals in parentheses. MIR = Tinbarap + Lambir Hills + Lawas; NIH = Niah; MUL = Gunung Mulu; BH = Borneo Highlands.

species	population 1	population 2	*gdi* 1	*gdi* 2	*M* 1	*M* 2
C. miriensis	MIR/NIH	MUL	0.30	0.59	1.9×10^{-2} (0.0 to 0.11)	5.7×10^{-3} (0.0 to 3.5×10^{-2})
	MIR	NIH	0.19	0.50	0.09 (5.0×10^{-2} to 0.11)	1.17 (0.93 to 1.48)
	MIR	MUL	NA	NA	0.04 (0.0 to 9.5×10^{-2})	0.56 (0.37 to 0.72)
	MUL	NIH	NA	NA	5.3×10^{-4} (0.0 to 3.8×10^{-3})	0.35 (0.29 to 0.41)
C. pubisulcus	KUC	BH	0.18	0.32	0.44 (0.33 to 0.57)	0.21 (3.0×10^{-3} to 0.34)
C. consobrinus	KUC	SER	0.12	0.11	0.06 (5.8×10^{-5} to 0.37)	0.14 (8.3×10^{-3} to 1.02)

S3)). The remaining gene flow estimates include zero in the 95% credible interval and are therefore considered not significant (electronic supplementary material, table S3). Gene flow in *C. pubisulcus* (M = 0.44 (0.33 to 0.57) from KUC to BH; M = 0.21 (0.003 to 0.34) from BH to KUC)) and *C. consobrinus* (M = 0.06 (5.8×10^{-5} to 0.37) from KUC to SER; M = 0.14 (8.3×10^{-4} to 1.02) from SER to KUC)) is bidirectional. Signatures of IBD are significant in *C. miriensis* and *C. consobrinus* (p-value = 0.004 and 0.025, respectively). The geographic pattern of diversity within *C. pubisulcus* is not driven by IBD (p = 0.43; figure 2). The EEMS analyses for *C. miriensis* and *C. pubisulcus* identify migration barriers in Lawas and Borneo Highlands, respectively, but the gene flow estimates suggest that these barriers are porous (electronic supplementary material, figure S3).

Divergence dates estimated from the genomic data using a germline-based mutation rate are younger than fossil-based estimates (electronic supplementary material, table S3). For *C. miriensis*, the earliest divergence occurs at 665 ka (95% highest posterior density (HPD): 409–817 kyr), followed by a more recent divergence between MIR and NIH at 146 ka (95% HPD: 108–188 kyr). For *C. pubisulcus*, the earliest divergence occurs at 252 ka (95% HPD: 217–289 kyr). Lastly, for *C. consobrinus*, the earliest divergence occurs at 60 ka, albeit with broad confidence intervals (95% HPD: 44–330 kyr). Of note, as mentioned above, the best model for *C. consobrinus* is as a single population.

Even when mtDNA and nuclear data support the same number of species, we found species boundaries that differed. Species delimitation using the *gdi* metric indicates that *C. miriensis* may contain two unrecognized species (*gdi* ≥ 0.50; table 2). The *gdi* values for *C. pubisulcus* are low (0.18) to intermediate (0.32), resulting in dubious support for splitting these populations into two species. The two-species model for *C. pubisulcus* differs from the two species delimited by mtDNA data by placing the population from Borneo Highlands as the only distinct genetic group, whereas mtDNA forms two groups: (i) Borneo Highlands + Bau + Kuching and (ii) Gunung Pueh + Gunung Gading + Gunung Matang. Finally, there is weak evidence for splitting *C. consobrinus* into two species, with *gdi* values definitively within the single-species range (0.12 and 0.11; table 2).

4. Discussion

Accurately delimiting species is an inherently complicated process, especially when dealing with cryptic taxa. Yet, species serve as fundamental units for conservation and other usages; thus, there is a clear rationale for applying rigorous approaches and relevant data for species delimitation [22]. For taxonomically complicated groups, it is advantageous to leverage genomic data to explore evolutionary patterns and the processes giving rise to them. We recognize the immense value of using mtDNA loci in systematic studies, yet it is important to acknowledge the limitations imposed by the reliance on a single genetic locus when delimiting species [16–21,49].

Although previous studies have shown that mtDNA tends to delimit more species than nuclear data, we find three distinct outcomes in our comparisons of mtDNA versus genomic species delimitation in *Cyrtodactylus* geckos: (i) oversplitting with mtDNA; (ii) agreement on a number of species, but differing population boundaries; and (iii) concordant species hypotheses. In the case of *C. miriensis*, genomic data provide support for recognizing a new species from Gunung Mulu based on a combination of a high *gdi* value and minimal gene flow (tables 1 and 2; electronic supplementary material, figure S3). However, we advise sampling of additional localities in the region prior to making any taxonomic changes to determine if this result is potentially biased by the small effective population sizes, which can skew *gdi* values towards supporting species-level designations [50]. Furthermore, the strong signature of IBD in this group indicates that comprehensive sampling could reveal a geographic gradient of diversity. For *C. pubisulcus*, mtDNA and genomic data largely agree on a two-population model, yet the composition of the populations differs. Such disagreement can be caused by incomplete lineage sorting, mtDNA introgression, or gene flow [51–53]. The species delimitation comparison in *C. pubisulcus* is illustrative of the different conclusions that are reached when using a single locus versus genomic data; specifically, the single-locus methods clearly support two species (>10% p-distance), whereas the genomic data support two populations connected by gene flow with weak support for a two-species model. Lastly, *C. consobrinus* is supported by both datasets as a single lineage, although some single-locus delimitation methods support a two-species model (figure 1).

The divergence dates estimated using a genomic rate calibration rather than a fossil calibration are substantially younger (electronic supplementary material, table S3). Fossil-based approaches estimate the earliest divergence time at 7.7 Ma for *C.*

miriensis and 8.3 Ma for *C. pubisulcus* [27]. Using a rate calibration, these same divergence events are estimated at 665 and 252 ka, respectively. There are several sources of error that need to be considered for the fossil calibration approach. First, there are no crown or stem Gekkonidae fossils that can confidently be placed on the phylogeny [54]. Therefore, secondary calibrations from other studies must be used, which propagates errors across studies. More importantly, the fossil calibrations that are frequently applied to *Cyrtodactylus* depend on fossils for *Sphaerodactylus* which shares an MRCA with Gekkonidae >100 Ma. Alternatively, the rate calibration we used is estimated from germline mutations quantified from high-coverage genomes using parent–offspring trios [45]. Therefore, we suspect that the rate-based divergence times are more accurate than the fossil-derived dates.

A prominent issue facing many *Cyrtodactylus*—and other taxa with similar geographic ranges—is an inability to comprehensively sample between known populations, which can overestimate species diversity [55]. Extensive sampling from intermediate or contact zones can better inform taxonomic decisions by revealing the patterns and processes of population divergence and/or population merging [56–59], whereas sampling isolated populations can inflate species numbers by accentuating the distinctiveness of populations, even under the MSC model [60]. In regions such as Borneo, where extensive land is inaccessible due to private ownership, a lack of public access and/or having undergone extensive deforestation, sampling gaps often reflect the true absence of data. However, these sampling gaps, whether biologically real or a byproduct of the landscape, can impact how we define a species boundary [60,61]. Sampling for *Cyrtodactylus* on Borneo for genetic studies has only just begun. Using genetic samples available to us, we aimed to generate genomic data for as many unique localities as possible, while utilizing enough samples per locality to accurately infer demographic patterns under an MSC model. Although genomic data do not definitively resolve these issues, they provide an avenue for inferring demographic histories in a manner that is not possible with mtDNA data. This demography can translate into more accurate taxonomic inferences, especially in complex systems containing cryptic diversity [62]. Despite our incomplete sampling, we demonstrate the utility of using genomic data to interrogate mtDNA-based delimitations.

The high diversity of *Cyrtodactylus* and its rank as the third most species-rich vertebrate genus is not an artefact of oversplitting with mtDNA but rather reflects an old origin, broad geographic distribution, high ecological diversity or other factors [63–65]. Yet, applying genomic data to delimit species within *Cyrtodactylus* more broadly may reduce the number of species in the group, especially for some of the recently described cryptic species based primarily on mtDNA. We demonstrate how mtDNA alone may be insufficient for accurately inferring species boundaries for cryptic lineages. However, dependence on mtDNA will presumably continue for many years, as it provides a comparatively easy and inexpensive approach to documenting life's biodiversity. Further, mtDNA allows for rapid species discovery, hypothesis development and delimitation which are crucial in an era of substantial global biodiversity loss. For these reasons, these data will remain important for biodiversity assessments across the tree of life, but how we define species continues to change as new data and methods become available [66]. Thus, the increasing accessibility of genomic data will provide more rigorous approaches for estimating and redefining species boundaries. For *Cyrtodactylus*, we recommend testing mtDNA species hypotheses as genomic data become more readily available and ultimately utilizing genomic data to explore the impressive diversification within the genus.

Ethics. Specimens were collected using a collection permits from the Sarawak Forestry Department and the Sarawak Forestry Corporation (NPW.907.4.4.(Jld.14)-79; (119)JHS/NCCD/600-7/2/107) and export permits were provided by the Sarawak Forestry Department (no. 18160).

Data accessibility. Sequence data are available at: BioProject ID: PRJNA1117503 and GenBank nos. PP973933–PP973940. Files used to perform analyses in this manuscript are available from the Dryad Digital Repository [67].

Supplementary material is available online [68].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. H.R.D.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, validation, visualization, writing—original draft, writing—review and editing; H.T.S.: data curation, formal analysis, methodology, writing—original draft, writing—review and editing; I.D.: investigation, resources, writing—original draft, writing—review and editing; I.N.: investigation, resources, writing—original draft, writing—review and editing; A.D.L.: conceptualization, formal analysis, investigation, methodology, project administration, resources, supervision, validation, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. H.R.D. received funding from The Society of Systematic Biologists (Graduate Research Award), The Biology Department at UW (Heerensperger Award & Orians Award for Tropical Studies), and from the Association for Tropical Biology and Conservation (ATBC Seed Research Grant) for this research.

Acknowledgements. We thank two anonymous reviewers for their insightful suggestions that greatly improved the final production of this manuscript. We thank the Sarawak Forestry Department and the Sarawak Forestry Corporation for providing support and permitting to conduct this study. We are grateful to Aaron M. Bauer, Ian G. Brennan, and Benjamin R. Karin for project organization and data collection. We are also grateful to Brett O. Butler and André L. G. Carvalho for providing comments and insight prior to submission.

References

1. Leaché AD, Fujita MK, Minin VN, Bouckaert RR. 2014 Species delimitation using genome-wide SNP data. *Syst. Biol.* **63**, 534–542. (doi:10.1093/sysbio/syu018)
2. O'Meara BC. 2010 New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* **59**, 59–73. (doi:10.1093/sysbio/syp077)
3. Grummer JA, Bryson RW, Reeder TW. 2014 Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Syst. Biol.* **63**, 119–133. (doi:10.1093/sysbio/syt069)
4. Yang Z. 2015 The BPP program for species tree estimation and species delimitation. *Curr. Zool.* **61**, 854–865. (doi:10.1093/czoolo/61.5.854)

5. Jackson ND, Morales AE, Carstens BC, O'Meara BC. 2017 PHRAPL: phylogeographic inference using approximate likelihoods. *Syst. Biol.* **66**, 1045–1053. (doi:10.1093/sysbio/syx001)
6. Sukumaran J, Holder MT, Knowles LL. 2021 Incorporating the speciation process into species delimitation. *PLoS Comput. Biol.* **17**, e1008924. (doi:10.1371/journal.pcbi.1008924)
7. Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013 A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876. (doi:10.1093/bioinformatics/btt499)
8. Fujisawa T, Barraclough TG. 2013 Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* **62**, 707–724. (doi:10.1093/sysbio/syt033)
9. Puillandre N, Brouillet S, Achaz G. 2021 ASAP: assemble species by automatic partitioning. *Mol. Ecol. Resour.* **21**, 609–620. (doi:10.1111/1755-0998.13281)
10. Maddison WP, Whitton J. 2023 The species as a reproductive community emerging from the past. *Bull. Soc. Syst. Biol.* **2**, 1–35. (doi:10.18061/bssb.v2i1.9358)
11. Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016 Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* **14**, e2000234. (doi:10.1371/journal.pbio.2000234)
12. Stankowski S, Ravinet M. 2021 Quantifying the use of species concepts. *Curr. Biol.* **31**, R428–R429. (doi:10.1016/j.cub.2021.03.060)
13. Struck TH *et al*. 2018 Finding evolutionary processes hidden in cryptic species. *Trends Ecol. Evol.* **33**, 153–163. (doi:10.1016/j.tree.2017.11.007)
14. Singhal S, Hoskin CJ, Couper P, Potter S, Moritz C. 2018 A framework for resolving cryptic species: a case study from the lizards of the Australian wet tropics. *Syst. Biol.* **67**, 1061–1075. (doi:10.1093/sysbio/syy026)
15. Fišer C, Robinson CT, Malard F. 2018 Cryptic species as a window into the paradigm shift of the species concept. *Mol. Ecol.* **27**, 613–635. (doi:10.1111/mec.14486)
16. Chan KO, Alexander AM, Grismer LL, Su YC, Grismer JL, Quah ESH, Brown RM. 2017 Species delimitation with gene flow: a methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian torrent frogs. *Mol. Ecol.* **26**, 5435–5450. (doi:10.1111/mec.14296)
17. Burriel-Carranza B, Estarellas M, Riaño G, Talavera A, Tejero-Cicuéndez H, Els J, Carranza S. 2023 Species boundaries to the limit: integrating species delimitation methods is critical to avoid taxonomic inflation in the case of the Hajar banded ground gecko (*Trachydactylus hajarensis*). *Mol. Phylogenet. Evol.* **186**, 107834. (doi:10.1016/j.ympev.2023.107834)
18. Hickerson MJ, Meyer CP, Moritz C. 2006 DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst. Biol.* **55**, 729–739. (doi:10.1080/10635150600969898)
19. Yang Z, Rannala B. 2017 Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Mol. Ecol.* **26**, 3028–3036. (doi:10.1111/mec.14093)
20. Pedraza-Marrón CDR *et al*. 2019 Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. *Proc. R. Soc. B* **286**, 20182924. (doi:10.1098/rspb.2018.2924)
21. Barley AJ, Thomson RC. 2016 Assessing the performance of DNA barcoding using posterior predictive simulations. *Mol. Ecol.* **25**, 1944–1957. (doi:10.1111/mec.13590)
22. Derkarabetian S, Starrett J, Hedin M. 2022 Using natural history to guide supervised machine learning for cryptic species delimitation with genetic data. *Front. Zool.* **19**, 8. (doi:10.1186/s12983-022-00453-0)
23. Oaks JR, Siler CD, Brown RM. 2019 The comparative biogeography of Philippine geckos challenges predictions from a paradigm of climate-driven vicariant diversification across an island archipelago. *Evolution* **73**, 1151–1167. (doi:10.1111/evo.13754)
24. Reilly SB *et al*. 2023 Bewildering biogeography: waves of dispersal and diversification across southern Wallacea by bent-toed geckos (genus: *Cyrtodactylus*). *Mol. Phylogenet. Evol.* **186**, 107853. (doi:10.1016/j.ympev.2023.107853)
25. Davis HR, Nashriq I, Woytek KS, Wikramanayake SA, Bauer AM, Karin BR, Brennan IG, Iskandar DT, Das I. 2023 Genomic analysis of Bornean geckos (Gekkonidae: *Cyrtodactylus*) reveals need for updated taxonomy. *Zool. Scr.* **52**, 249–263. (doi:10.1111/zsc.12575)
26. Davis HR, Das I, Leaché AD, Karin BR, Brennan IG, Jackman TR, Nashriq I, Onn Chan K, Bauer AM. 2021 Genetically diverse yet morphologically conserved: hidden diversity revealed among Bornean geckos (Gekkonidae: *Cyrtodactylus*). *J. Zool. Syst. Evol. Res.* **59**, 1113–1135. (doi:10.1111/jzs.12470)
27. Davis HR *et al*. 2020 Multilocus phylogeny of bornean bent-toed geckos (Gekkonidae*: Cyrtodactylus*) reveals hidden diversity, taxonomic disarray, and novel biogeographic patterns. *Mol. Phylogenet. Evol.* **147**, 106785. (doi:10.1016/j.ympev.2020.106785)
28. Queiroz de K. 1998 The general lineage concept of species, species criteria, and the process of speciation. In *Endless forms: species speciation*, pp. 57–75. New York, NY: Oxford University Press.
29. Mayden R. 1997 A hierarchy of species concepts: the denouncement in the saga of the species problem. In *Species: the units of biodiversity* (eds MF Claridge, HA Dawah, MR Wilson), pp. 381–423. London, UK: Chapman and Hall.
30. Naomi SI. 2011 On the integrated frameworks of species concepts: Mayden's hierarchy of species concepts and de Queiroz's unified concept of species. *J. Zool. Syst. Evol. Res.* **49**, 177–184. (doi:10.1111/j.1439-0469.2011.00618.x)
31. Leaché AD, Zhu T, Rannala B, Yang Z. 2019 The spectre of too many species. *Syst. Biol.* **68**, 168–181. (doi:10.1093/sysbio/syy051)
32. Jackson ND, Carstens BC, Morales AE, O'Meara BC. 2017 Species delimitation with gene flow. *Syst. Biol.* **66**, 799–812. (doi:10.1093/sysbio/syw117)
33. Ramirez JL, Valdivia P, Rosas-Puchuri U, Valdivia NL. 2023 SPdel: a pipeline to compare and visualize species delimitation methods for single-locus datasets. *Mol. Ecol. Resour.* **23**, 1959–1965. (doi:10.1111/1755-0998.13864)
34. Bouckaert R *et al*. 2019 BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650. (doi:10.1371/journal.pcbi.1006650)
35. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135. (doi:10.1371/journal.pone.0037135)
36. Eaton DAR. 2014 PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844–1849. (doi:10.1093/bioinformatics/btu121)
37. Danecek P *et al*. 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
38. Matschiner M. 2022 Species tree inference with SNP data. *Methods Mol. Biol.* **2512**, 23–44. (doi:10.1007/978-1-0716-2429-6_2)
39. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
40. Flouri T, Jiao X, Huang J, Rannala B, Yang Z. 2023 Efficient Bayesian inference under the multispecies coalescent with migration. *Proc. Natl Acad. Sci. USA* **120**, e2310708120. (doi:10.1073/pnas.2310708120)
41. Alexander DH, Lange K. 2011 Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246. (doi:10.1186/1471-2105-12-246)
42. Petkova D, Novembre J, Stephens M. 2016 Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100. (doi:10.1038/ng.3464)
43. Jombart T. 2008 adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)
44. Kornai D, Flouri T, Yang Z. 2023 Hierarchical heuristic species delimitation under the multispecies coalescent model with migration. *bioRxiv* 2023.09.10.557025. (doi:10.1101/2023.09.10.557025)

45. Bergeron LA *et al*. 2023 Evolution of the germline mutation rate across vertebrates. *Nature* **615**, 285–291. (doi:10.1038/s41586-023-05752-y)
46. Wood P, Grismer LL, Muin MA, Anuar S, Oaks JR. 2020 A new potentially endangered limestone-associated bent-toed gecko of the *Cyrtodactylus pulchellus* (Squamata: Gekkonidae) complex from northern peninsular Malaysia. *Zootaxa* **4751**. (doi:10.11646/zootaxa.4751.3.2)
47. Termprayoon K, Rujirawan A, Grismer LL, Wood PL Jr, Aowphol A. 2023 Two new karst-adapted species in the C*yrtodactylus pulchellus* group (Reptilia, Gekkonidae) from southern Thailand. *Zookeys* **1179**, 313–352. (doi:10.3897/zookeys.1179.109712)
48. Lalremsanga HT, Colney Z, Vabeiryureilai M, Malsawmdawngliana F, Bohra SC, Biakzuala L, Muansanga L, DAS M, Purkayastha J. 2023 Its all in the name: another new *Cyrtodactylus* Gray (Squamata: Gekkonidae) from northern Mizoram, north-east India. *Zootaxa* **5369**, 553–575. (doi:10.11646/zootaxa.5369.4.5)
49. Vences M, Miralles A, Dufresnes C. 2024 Next-generation species delimitation and taxonomy: implications for biogeography. *J. Biogeogr.* (doi:10.1111/jbi.14807)
50. Leaché AD, Davis HR, Singhal S, Fujita MK, Lahti ME, Zamudio KR. 2021 Phylogenomic assessment of biodiversity using a reference-based taxonomy: an example with horned lizards (*Phrynosoma*). *Front. Ecol. Evol.* **9**, 678110. (doi:10.3389/fevo.2021.678110)
51. DeRaad DA, McCullough JM, DeCicco LH, Hime PM, Joseph L, Andersen MJ, Moyle RG. 2023 Mitonuclear discordance results from incomplete lineage sorting, with no detectable evidence for gene flow, in a rapid radiation of *Todiramphus* kingfishers. *Mol. Ecol.* **32**, 4844–4862. (doi:10.1111/mec.17080)
52. Toews DPL, Brelsford A. 2012 The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **21**, 3907–3930. (doi:10.1111/j.1365-294X.2012.05664.x)
53. Firneno TJ Jr, O'Neill JR, Portik DM, Emery AH, Townsend JH, Fujita MK. 2020 Finding complexity in complexes: assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. *Mol. Ecol.* **29**, 3543–3559. (doi:10.1111/mec.15496)
54. Daza JD, Bauer AM, Snively ED. 2014 On the fossil record of the Gekkota. *Anat. Rec.* **297**, 433–462. (doi:10.1002/ar.22856)
55. Opatova V, Bourguignon K, Bond JE. 2024 Species delimitation with limited sampling: an example from rare trapdoor spider genus *Cyclocosmia* (Mygalomorphae, Halonoproctidae). *Mol. Ecol. Resour.* **24**, e13894. (doi:10.1111/1755-0998.13894)
56. Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013 How to fail at species delimitation. *Mol. Ecol.* **22**, 4369–4383. (doi:10.1111/mec.12413)
57. Chambers EA, Marshall TL, Hillis DM. 2023 The importance of contact zones for distinguishing interspecific from intraspecific geographic variation. *Syst. Biol.* **72**, 357–371. (doi:10.1093/sysbio/syac056)
58. Nabhan AR, Sarkar IN. 2012 The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Brief. Bioinform.* **13**, 122–134. (doi:10.1093/bib/bbr014)
59. Heath TA, Hedtke SM, Hillis DM. 2008 Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* **46**, 239. (doi:10.3724/SP.J.1002.2008.08016)
60. Hedin M, Carlson D, Coyle F. 2015 Sky island diversification meets the multispecies coalescent—divergence in the spruce-fir moss spider (Microhexura montivaga, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Mol. Ecol.* **24**, 3467–3484. (doi:10.1111/mec.13248)
61. Mason NA, Fletcher NK, Gill BA, Funk WC, Zamudio KR. 2020 Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Syst. Biodivers.* **18**, 269–280. (doi:10.1080/14772000.2020.1730475)
62. Pyron RA, Kakkera A, Beamer DA, O'Connell KA. 2024 Discerning structure versus speciation in phylogeographic analysis of seepage salamanders (*Desmognathus aeneus*) using demography, environment, geography, and phenotype. *Mol. Ecol.* **33**, e17219. (doi:10.1111/mec.17219)
63. Grismer LL, Poyarkov NA, Quah ESH, Grismer JL, Wood PL Jr. 2022 The biogeography of bent-toed geckos, *Cyrtodactylus* (Squamata: Gekkonidae). *PeerJ* **10**, e13153. (doi:10.7717/peerj.13153)
64. Grismer LL, Wood PL, Le MD, Quah ESH, Grismer JL. 2020 Evolution of habitat preference in 243 species of bent-toed geckos (genus *Cyrtodactylus* Gray, 1827) with a discussion of karst habitat conservation. *Ecol. Evol.* **10**, 13717–13730. (doi:10.1002/ece3.6961)
65. O'Connell KA, Smart U, Sidik I, Riyanto A, Kurniawan N, Smith EN. 2019 Diversification of bent-toed geckos (*Cyrtodactylus*) on Sumatra and west Java. *Mol. Phylogenet. Evol.* **134**, 1–11. (doi:10.1016/j.ympev.2019.01.021)
66. Pante E *et al*. 2015 Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Mol. Ecol.* **10**, 525–544. (doi:10.1111/mec.13048)
67. Davis HR, Sanford HT, Das I, Nashriq I, Leaché AD. 2024 Data from: establishing species boundaries in bornean geckos. Dryad Digital Repository. (doi:10.5061/dryad.w0vt4b90w)
68. Davis HR, Sanford HT, Das I, Nashriq I, Leaché AD. 2024 Data from: Establishing species boundaries in Bornean geckos. Figshare. (doi:10.6084/m9.figshare.c.7397758)