Journal of Zoological Systematics and Evolutionary Research

WILEY

RESEARCH ARTICLE OPEN ACCESS

SNPs Resolve the Phylogenetic History of Chinese Geckos Where mtDNA and Morphology can Mislead

Jun Xu[1] | Yubin Wo[1] | Hua Feng[1] | Qianjing Lin[1] | Haojie Tong[1] | Zhengjun Wu[2] | Richard P. Brown[3] | Yuanting Jin[1]

[1]College of Life Sciences, China Jiliang University, Hangzhou 310018, China | [2]Guangxi Key Laboratory of Rare and Endangered Animal Ecology, College of Life Science, Guangxi Normal University, Guilin 541004, China | [3]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, L3 3AF, UK

Correspondence: Zhengjun Wu (wu_zhengjun@aliyun.com) | Richard P. Brown (r.p.brown@ljmu.ac.uk) | Yuanting Jin (jinyuanting@126.com)

Received: 8 December 2025 | **Revised:** 11 March 2026 | **Accepted:** 12 March 2026

Academic Editor: Savel Daniels

Keywords: *Gekko* | molecular systematics | population divergence | Sichuan Basin

ABSTRACT

Discordance between mitochondrial DNA (mtDNA) trees and species history is not uncommon and caused by several well-known processes. Differences in phylogenetic signal between loci within the mtDNA genome have received less attention. Large amounts of nuclear genomic data provide the opportunity to resolve uncertain phylogenetic inference in cases of discordance. We examined relationships among three putative divergent lineages of Chinese webbed-toed gecko: *Gekko subpalmatus*, *G. cib* (formerly *G. subpalmatus*) *and G. melli*. MtDNA sequences (*ND2*/*COXI*) were analyzed and an mtDNA tree that supported ((*G. melli* and *G. subpalmatus*) and *G. cib*) was obtained. This differed from a recent well-supported mtDNA–based phylogeny. To resolve this, we estimated species trees using genomic SNPs (from genotyping-by-sequencing), which supported our *ND2*/*COXI* tree. We investigated how the mtDNA loci used in these studies might provide discordant phylogenetic signals by comparison of 16 trees obtained from individual mtDNA loci. Well-supported trees, but with two distinct topologies, were obtained for six of the loci, while the remaining 10 loci provided weakly-supported topologies. Only three mtDNA loci provided a well-supported tree that was concordant with the genomic SNPs tree. This suggests that differences in phylogenetic signal across the mitochondrial genome misled previous phylogenetic inference. Divergence time dating revealed quite similar internal node ages, which likely exacerbated this effect. Finally, we analyzed generalized divergence in body dimensions: the two nonsister taxa *G. subpalmatus* and *G. cib* were the least divergent lineage pair, potentially explaining previous incorrect conclusions. We conclude that different well-supported phylogenetic histories can be supported by different mtDNA loci while large numbers of genomic SNPs provide more reliable topologies.

1 | Introduction

Numerous studies have used mitochondrial DNA (mtDNA) to establish the phylogenetic relationships among taxa [1–3]. These findings have been of particular relevance in systematics, especially since phylogenetic relationships became widely accepted as the basis for taxonomy [4, 5]. The rise of phylogenetic systematics paralleled the introduction and subsequent surge in mtDNA sequencing by phylogeneticists. Over the years, different authors have both urged caution and promoted its use for determining historical relationships, but the overwhelming attitude towards

Jun Xu, Yubin Wo, and Hua Feng contributed equally to this work.

mtDNA has remained positive [6–8]. The greater substitution rate of mtDNA has been a key factor in its use at lower taxonomic levels, although this advantage is now diminished due to increased availability of genomic methodologies that provide large amounts of DNA sequence from across the nuclear genome. These approaches not only overcome the lower phylogenetically informative content per base of nuclear DNA but also circumvent other potential limitations of mtDNA.

Mito-nuclear discordance is increasingly detected as the number of nuclear genomic studies increase, while limits to the usefulness of mtDNA in species delimitation are also recognized [9]. Any single locus such as mtDNA will be less reliable for inferring speciation events because of potential stochastic effects, with perhaps the best-known effect being incomplete lineage sorting due to lack of coalescence of mtDNA lineages within species. In addition, cross-species hybridization can facilitate mtDNA capture from other (potentially extinct) species/lineages and mislead detection of speciation, while the maternal inheritance pattern of mtDNA can also affect phylogenetic inference [10]. For example, lack of female dispersal with male vagility can lead to incorrect species inference [11]. Nonetheless, one effect that has not been thoroughly explored is discordant phylogenetic signal between different loci within the mtDNA genome.

In this study we reexamined phylogenetic relationships among three closely related lineages of Chinese gecko, *Gekko subpalmatus*, *G. cib* and *G. melli*, using mtDNA, nuclear SNPs, and morphology. The long taxonomic history of this group has been recently summarized in detail by Lyu et al. [12] and so will not be fully reviewed again here. The main relevant publications that preceded Lyu et al. [12] were: Rösler and Tiedemann [13]; Yang et al. [14]; Zhao and Adler [15]. We also note that the focal species (along with others) have recently been assigned to the subgenus *Japonigekko* [16], but here we will refer to the different species simply using the traditional genus name, *Gekko*.

Several older studies addressed the systematics of the group but did not consider phylogenetic relationships, which provided the initial motivation for this study. Furthermore, a recently published mtDNA tree did not show a sister relationship between geographically neighboring *Gekko* lineages [12]. These results suggested that the underlying species phylogeny remained unresolved.

Prior to Lyu et al. [12], two species were recognized across three allopatric groups within China: (i) *G. subpalmatus* from the Sichuan Basin and adjacent low-elevation mountainous regions in central China and a geographically disjunct area of eastern China comprising parts of Zhejiang province and the Zhoushan archipelago and (ii) *G. melli* which is geographically adjacent to and found directly to the south of the Zhejiang *G. subpalmatus* populations, mainly in northern Guangdong province but also neighboring areas of southern Jiangxi [12]. Lyu et al.'s [12] analysis of short 16S rRNA and cytochrome b (Ctyb) sequences identified three deep mitochondrial lineages across these three groups: *G. subpalmatus* Sichuan was inferred as the sister lineage to *G. melli* (bootstrap support, 77%; Bayesian posterior probability, 0.99) which together formed the sister lineage to *G. subpalmatus* Zhejiang/Zhoushan. Hence, they elevated *G. subpalmatus* from Sichuan to the species level, describing *Gekko* (*Japonigekko*) *cib* sp.nov. Lyu et al. [12] also provided a detailed analysis of body dimensions, scalation and hemipenes that showed clear differentiation that allowed morphological diagnosis of the three lineages. Examination of geographical proximity could lead to the prima facie prediction that *G. melli* and *G. subpalamatus* represent sister taxa, in contrast with Lyu et al.'s [12] findings.

Here, *G. cib*, *G. subpalmatus* and *G. melli* were sampled to provide genetic and morphological datasets to reassess the phylogeny and systematics of this group. We analyzed long mtDNA sequences from individuals from all areas and also analyzed representative mitochondrial genomes from each of the three allopatric groups mentioned above. We measured linear body dimensions to gain further insights into patterns of morphological divergence among putative species. Most significantly, we set out to resolve uncertainties by performing genotyping by sequencing (GBS) to allow analysis of nuclear markers from across the genome. Through replication and extension of previous analyses, we were able to examine how different markers provided different evolutionary inferences and how this had misled previous studies. The ultimate aim was to unequivocally determine the species history of these geckos to provide a solid foundation for the species taxonomy.

2 | Materials and Methods

2.1 | Sampling

A total of 116 *G. subpalmatus*, *G. cib*, and *G. melli* were sampled in 2017. While there is only sparse information on their respective distributions, we believed that the samples covered the major parts of their known range (despite a lack of *G. subpalmatus* samples from mainland Zhejiang; Figure 1). Specifically, geckos were measured and tail-tips were retained from: (1) *G. cib* from three localities within the Sichuan Basin in western China ($n = 18$), (2) *G. melli* from five localities from Guandong and Jiangxi provinces in Southeastern China ($n = 29$), and (3) *G. subpalmatus* from 15 localities within the Zhoushan Archipelago in eastern China ($n = 69$; see Supporting Information 1: Table S1 for more details). Tail tips were stored in 95% ethanol and individuals were released at the site of capture following morphological measurements.

2.2 | DNA Extraction, PCR, and Sequencing

Genomic DNA was extracted from tail muscle tissue of all sampled individuals using a standard protocol of proteinase K digestion and DNeasy Blood & Tissue Kit (QIAGEN). DNA of 106 individuals was analyzed as described in the next two sections (10 individuals were not analyzed due to degradation of DNA).

2.3 | MtDNA Sequences and Trees

Primer pairs (Supporting Information 2: Table S2) were designed to amplify two mtDNA fragments (2289 bp) from the *ND2* and *COXI* genes for all 106 specimens sequenced. PCR reactions were performed with 1.2 μL of DNA template, 1.2 μL of each primer, 11.4 μL of dd H_2O, and 15 μL of PrimeSTAR Max DNA Polymerase. The PCR cycle was: 95°C for 2 min, 35 cycles of 98°C for 10 s, 55°C for 5 s, and 72°C for 7 s, with a final extension at 72°C for 10 min. The PCR amplification products were sent to Sangon Biotech (Shanghai) Co., Ltd. for product purification and ABI 3770 Sanger sequencing in both directions using PCR primers.

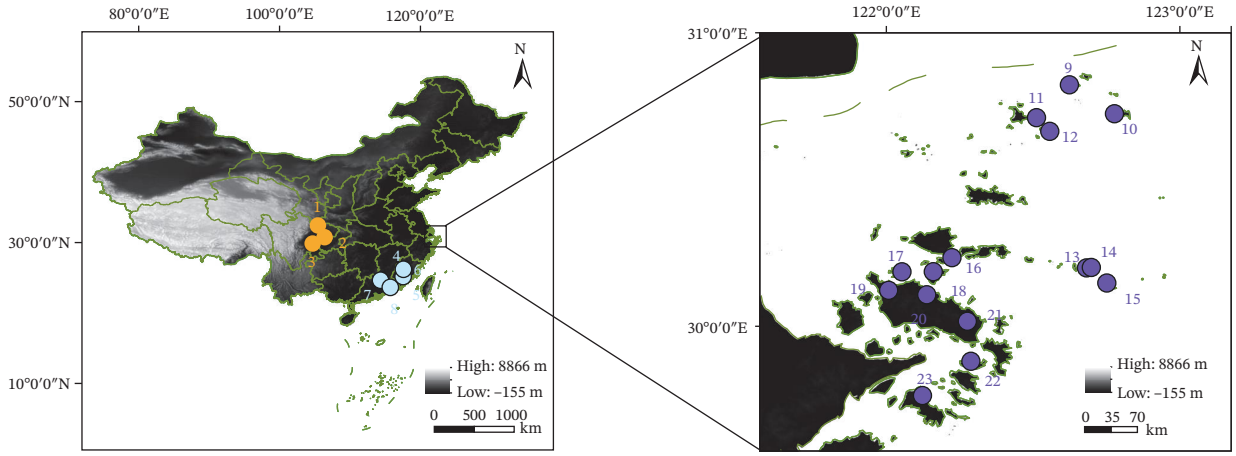

FIGURE 1 | Sampling distribution of the *Gekko* species complex used in this study. Dots represented sampled localities of the taxa as named here: *G. cib* (Sites 1–3), *G. subpalmatus* from Zhejiang and Zhoushan (Sites 9–23), *G. melli* from Fujiang, Guangdong, and Jiangxi in Southeast of China (Sites 4–8).

MtDNA sequences were aligned using ClustalW in MEGA11 [17], manually checked and then collapsed into haplotypes using DnaSP v. 6.12.03 [18]. Homogenous partitions of sequence data were first identified using PartitionFinder v. 2.1.1 [19]. Bayesian inference was used to obtain mtDNA trees, using MrBayes v. 3.2.7 [20]. Two simultaneous runs of four MCMC chains were performed for 10 million steps (sampled every 5000 steps). Convergence was assessed from MrBayes trace plots, examination of potential differences between runs and from the standard deviation of split frequencies (<0.01 was used as indicative of convergence).

Maximum likelihood (ML) analyses were computed using RAxML-NG v. 1.1.0 [21]. The GTR+G model was used for *COXI* and the HKY+G model was used for *ND2*, as determined by PartitionFinder v. 2.1.1 [19]. Parsimony trees were used as starting points for the ML analyses. Statistical support was obtained from 1000 bootstrap replicates. We used *Gekko japonicus* (GenBank Accession: KT005800.1; [22]) as an outgroup, which was justified following previous studies [16, 23]. During preliminary analyses we also used another outgroup (*G. swinhonis*), but this had no impact on the results and so is not reported here.

To allow direct comparison with the Lyu et al. [12] study, we retrieved their 16S rRNA and Cytb sequences (see "MtDNA analyses previous data" file within Supporting Information 3) from Genbank and, as described above, applied PartitionFinder and then RAxML-NG v. 1.1.0 [21] to infer the mtDNA tree.

Indices of substitution saturation (ISS) were assessed using DAMBE [24], for both our *COXI–ND2* dataset and the 16S rRNA/*Cytb* dataset from Lyu et al. [12]. In addition, we obtained a bootstrapped ML tree for each of 16 mtDNA loci representing all protein coding genes, ribosomal RNA, and the control region using a full mtDNA genome retrieved for *G. cib* (GenBank Accession: MK680534, [25]) and genomes from a *G. subpalmatus* and a *G. melli* specimen that we recently sequenced for another project (available within "MtDNA analyses previous data" file within Supporting Information 3). Sequences from each of these three individuals and the outgroup *G. japonicus* were individually aligned for each of the 16 mtDNA loci using the ClustalW algorithm in MEGA11 [17]. The program jModelTest 2v. 2.1.10 [26, 27] was used to identify the best-fit model for each mtDNA locus (Supporting Information 4: Table S4). We used the RAXML-NG procedure described previously to obtain boostrap trees for each locus. Topologies of all boostrap replicate trees were recorded.

Finally, we concatenated the mtDNA loci sequences to obtain a complete mitochondrial genome tree using the same approach. Here, partitions and models were determined by PartitionFinder v. 2.1.1 [19] (Supporting Information 5: Table S5).

2.4 | Divergence Times From mtDNA

COXI and *ND2* haplotype sequences were analyzed using BEAST2 [28] to analyze mtDNA divergence times. The topology was monophyly-constrained to ensure that posterior samples comprised only trees in which relationships between the three major lineages were those identified by SNP analyses (Section 3). Sequences from each gene were partitioned into (i) the first two and (ii) the third codon positions, providing four partitions across the *COXI* and *ND2* sequences. The HKY model was employed as the site model for all partitions, with site rate variation within *ND2* being accommodated by a gamma distribution. Two strict clock models were specified, one for each fragment. We used a tight prior on ND2 rates, which was specified from the Normal distribution N(0.00762, 1.6×10^{-7}), providing a 95% prior density from 0.00684 to 0.00840. The rationale for this choice was that previous studies on the lizard *Phrynocephalus erythrurus* had shown a rate of 0.762% per million years for a similar mtDNA fragment (albeit extending from the *ND2* gene to $tRNA^{Ala}$) [29]. Relatively few studies have estimated the *ND2* substitution rate in reptiles and amphibians but there is a general consensus of approximately 0.0065–0.0100 subs/site/Ma, with the lower end of this range appearing most credible ([30] and references therein). The *COXI* rate was expected to be quite similar, or slightly greater but, as we had little prior information on expected rates in geckos, we specified a diffuse prior from a log Normal distribution with a geometric mean of 0.008 and a standard deviation of 0.5 (in real space). This allowed the *ND2* rate prior to primarily determine estimation of divergence times. Preliminary runs with a relaxed clock indicated little rate variation and so a strict molecular clock was used. The tree prior was specified from a Yule process. The MCMC chain contained 20 million steps, following a pre-burn-in period of 500,000 steps. The final chain burn-in period was determined by examination of the MCMC traces of posteriors and likelihoods against generation using Tracer v. 1.7

[31]. This same program was used to determine whether lengths of runs were sufficient, using the effective sample sizes (ESSs; ESS >200 was used to diagnose this). MCMC chains were run three times, from different starting positions.

2.5 | Genotyping-by-Sequencing

A 43-sample subset of the 106 individuals described above was used to represent the three main groups (*G. cib*, $n = 15$; *G. melli*, $n = 12$; *G. subpalmatus*, $n = 16$; see Supporting Information 1: Table S1 for site details). Extracted DNA was submitted to Beijing Novogene Bioinformatics Technology Co., Ltd. for GBS analysis using the following protocol. Step one was incubation of 50–100 ng of extracted genomic DNA with the restriction enzymes *NlaIII* and *EcoRI*. After purification, DNA was amplified by PCR, and the amplification product was purified with magnetic AMPure XP. Purified samples were subject to electrophoresis, and then DNA fragments measuring 300–400 bp in size were excised from the gel to construct the library. The Qubit 2.0 Fluorometer was used for preliminary quantification of the DNA library. The insert size of the library was detected using an Agilent 2100 bioanalyzer, and when the insert size met expectations, qRT-PCR was used to quantify the library effective concentration (library effective concentration is higher than 2 nmol/L) to ensure library quality. Paired-end sequencing was performed on the NovaSeq 6000 platform (read length was 150 bp).

AdaptorRemoval v. 2 [32] was used to remove adapters in the raw data and FastQC v. 0.12.1 [33] was used to remove low-quality reads. When the number of low-quality ($Q \leq 5$) bases contained in a single-ended sequencing read exceeded 50% of the length ratio of the read, the paired reads were removed.

Alignment of reads and SNP calling were performed using GBS-SNP-CROP v. 4.1 [34] with the *G. japonicus* genome (GenBank Accession: GCF_001447785.1) [35] as the reference genome. A matrix containing all SNPs was generated and then filtered using the final perl script in GBS-SNP-CROP (i.e., GBS-SNP-CROP-7.pl) with the following specifications: a minimum read depth (-mnHetDepth) for heterozygote calling of 3; a minimum required proportion of secondary reads to all non-primary reads (-altStrength) of 0.8; proportion of individuals required for a variant to be called (-mnCall) of 0.75; maximum average depth of an acceptable variant (-mxAvgDepth) of 50; minimum depth for calling a homozygote when the alternative allele depth is 0 (-mnHoDepth0) of 5; minimum depth for calling a homozygote when the alternative allele depth is 1 (-mnHoDepth1) of 20; minimum ratio of less frequent allele depth to more frequent allele depth (-mnAlleleRatio) of 0.25; minimum average depth of an acceptable variant (-mnAvgDepth) of 3. This pipeline was used to provide only biallelic SNPs; indels were discarded.

2.6 | Trees From Genomic SNP Data

We obtained a ML SNP tree using RAxML-NG v. 1.1.0 [21] and a Bayesian SNP tree using MrBayes v. 3.2.7 [20] to gain preliminary insight into the relationships between individuals. All SNPs were concatenated for these analyses and then filtered using a published script (raxml_ascbias/ascbias.py at master · btmartin721/raxml_ascbias · GitHub). The GTR+G model was identified using jModelTest 2v. 2.1.10 [26, 27] and used in both ML and Bayesian analyses. In the ML analysis, the conditional likelihood method was used to correct for ascertainment bias [36], that is, +ASC_LEWIS in RAxML-NG v. 1.1.0 [21], and the model for unphased diploid genotypes (GTGTR) was specified. Heuristic searches were performed 10 times to identify trees, with 1000 bootstrap replicates used to assess node support. For the Bayesian analysis, the Lewis correction was used but we also note the lack of a phasing function, so the two bases at heterozygous sites (coded by an ambiguity symbol) are each assigned a probability of 0.5. The MCMC chain length was 10 million iterations, with trees sampled every 1000 steps. ESSs, traces, and standard deviations of split frequencies (obtained from the analysis itself) were used to determine appropriate burn-in.

The species tree and species delimitation analyses were performed using the SNP model implemented in SNAPPER v. 1.1.2 [37] in conjunction with SPEEDEMON [38], within the program BEAST 2 [28]. To mitigate the effects of linkage, the SNPs were thinned to include only one SNP per sequence tag. These analyses provided comparisons of different species division models based on genome-wide SNP data within a multispecies coalescent framework. The three defined species in these analyses were the three groups within the study (*G. melli*, *G. subpalmatus*, and *G. cib*). The tree prior was specified by the Yule skyline collapse model. Branches that did not reach a threshold length were collapsed. Here the threshold (ε) used was 10^{-6} (we tested alternative ε values of 10^{-5} and 10^{-7}, but this had no impact on our results). The weights of the Gamma Mover and Rate Mixer operators were set to 10. The MCMC chain length was 1 million generations with a sample interval of 1000 generations, and a pre-burn-in of 100,000, ensuring an ample number of iterations. Other parameters were kept at their default values. As described previously, convergence of posteriors was assessed using Tracer v. 1.7 [31]. The post-burn-in sample of trees was then used to construct a maximum clade credibility consensus tree using TreeAnnotator [39] and DensiTree [40] used to examine the posterior tree sample.

2.7 | Morphological Divergence

The following measurements were recorded with a vernier caliper on 115 sampled individuals: snout-vent length (SVL), head length (HL), head width (HW), anterior limb length (AL, distance between axilla and wrist), posterior limb length (LL, distance between groin and ankle), distance between axillae (DBA), and distance between iliac crests (DBIs). All traits were measured to 0.1 mm by the first author. Four *G. melli* specimens were preserved and retained in the laboratory of animal evolution, College of Life Sciences, China Jiliang University.

We performed individual ANCOVAs on each of the six (excluding SVL) $\log_{10}$-transformed linear body dimensions across the three defined groups, with $\log_{10}$ SVL as the covariate, to check for significance of individual among-group variation and estimate pooled within-group regression slopes (we previously tested for sexual size dimorphism using a two-way ANCOVA but did not detect this for any character). This analysis was performed using IBM SPSS Statistics v. 28.0. All six $\log_{10}$-transformed characters were adjusted to their expected values at the overall mean $\log_{10}$ SVL using the pooled within-group regression slopes from the ANCOVA. This is a standard allometric correction (see for example, [41]). Next we performed a one-way multivariate analysis of

variance between the three gecko groups using the npmv package [42] in R (ver.4.2.3) [43]. This package provides a permutation test that can test for a difference in location between groups without the restrictive assumptions of a normal MANOVA.

Discriminant function analysis (DFA) was applied to the six $\log_{10}$ SVL–adjusted characters to investigate generalized morphological divergence. Mahalanobis distances were computed between the three groups, with confidence intervals for these distances being obtained using nonparametric bootstrapping (this was used rather than the noncentral F distribution because of low Mahalanobis D^2 values: [44]). This analysis was performed using the R packages HDMD [45] and boot [46] and have provided our R scripts for this procedure in the supporting information.

3 | Results

3.1 | MtDNA Tree

MtDNA sequences were obtained from 106 of the sampled specimens. GenBank accessions and other information are given in Supporting Information 1: Table S1. The *ND2* alignment consisted of 975 bp (310 variable sites), while the *COXI* alignment comprised 1314 bp (306 variable sites). Pairwise uncorrected divergence between pairs of species was 16.6%–27.4%. PartitionFinder analyses favored use of a single partition for both genes, with the HKY +G model favored for *ND2* and the GTR+G model for *COXI*.

The three focal groups were monophyletic in trees obtained using the *COXI* + *ND2* sequences (Figure 2). The most recent node

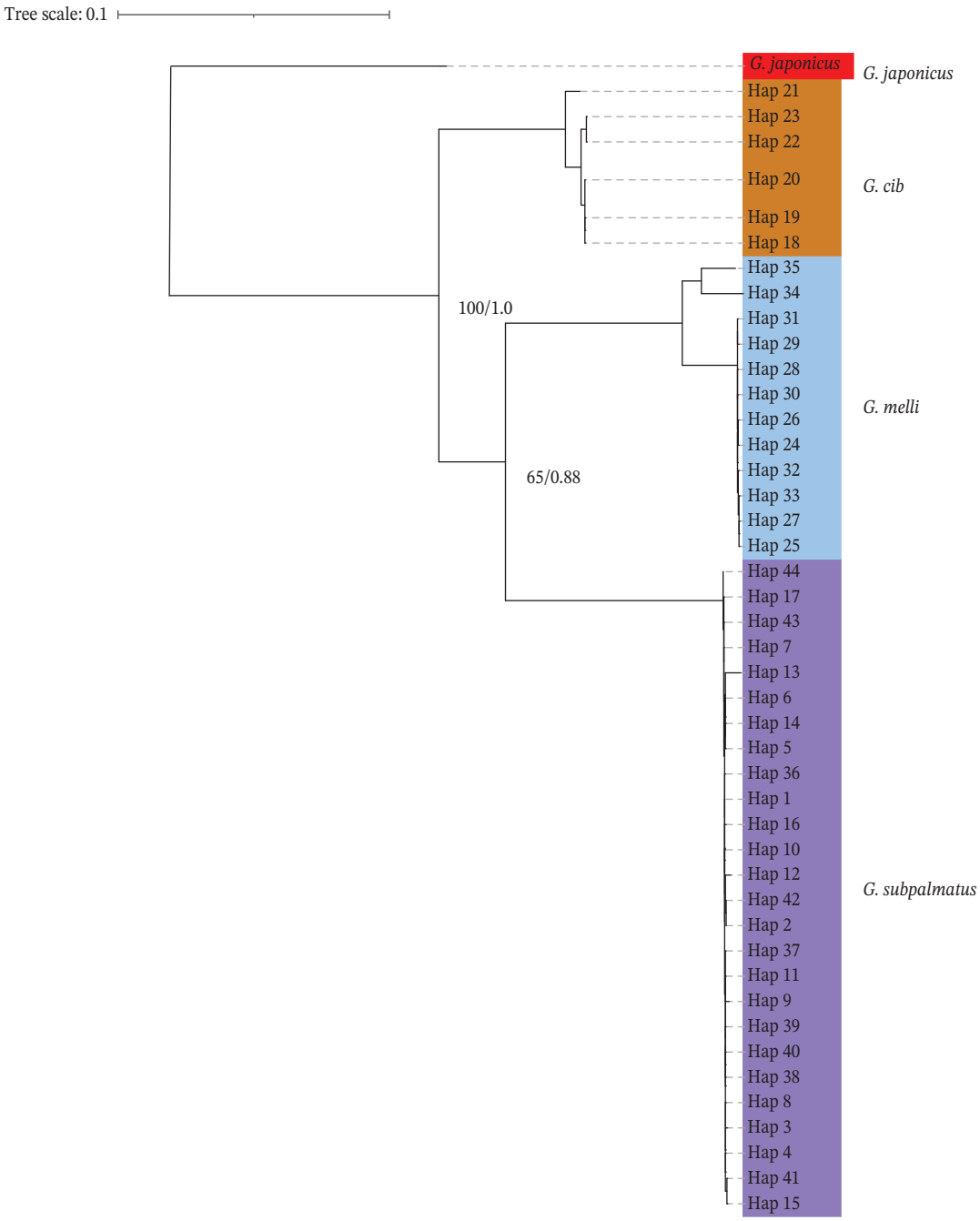

FIGURE 2 | The topologies obtained for RAxML and MrBayes phylogenetic trees estimated from mitochondrial *ND2* and *COXI* sequences. Node labels represent RAxML bootstrap values/MrBayes posterior probabilities.

represented divergence of *G. melli* and *G. subpalmatus* mtDNA, although support was weak (bootstrap branch support = 65%, posterior probability = 0.88). This group forms a sister group to *G. cib*.

The HKY+G and GTR+G, substitution models were used to infer trees using Lyu et al.'s [12] published *Cytb* and 16S rRNA sequence data, respectively. They provided exactly the same topology as that published by the original authors, that is, they differed from our *ND2*/*COXI* mtDNA tree, with *G. melli* and *G. cib* forming a monophyletic group. Bootstrap values and posterior probabilities for the previous branch (preceding divergence of the *G. subpalmatus* mitochondrial lineage) were high (bootstrap support = 97% and posterior probability = 1.0) while respective values for the (*G. melli* and *G. cib*) branch were lower (bootstrap support = 77% and posterior probability = 0.98).

For our *COXI*/*ND2* dataset, analysis across all OTU subsets (4–32) revealed ISS values (0.270–0.292) that were significantly lower than critical ISS values (the threshold above which saturation is deemed to have led to a loss of phylogenetic signal) under both symmetrical (0.796–0.843) and asymmetrical (0.528–0.820) tree topologies ($p < 0.0001$). For the 16S rRNA/*Cytb* dataset of Lyu et al. [12], the observed ISS (0.2754) was also significantly lower than critical ISS values (0.7960 for symmetrical trees, 0.6166 for extremely asymmetrical trees; $p < 0.0001$).

Of the 16 mtDNA ML trees corresponding to individual mtDNA genes, six provided bootstrap support with values for sister lineages ≥70%, which represents the minimum level that is considered indicative of support [47]. However, the topologies of these six trees differed. Three of them (16S rRNA, *ND4L*, and *ND6*) supported monophyly of the *G. melli* and *G. subpalmatus* lineages, while the other three (*ATP6*, *COXII*, and *Cytb*) favored monophyly of *G. melli* and *G. cib* (relationships across bootstrap replicates are summarized in Figure 3). The remaining 10 trees (12S rRNA, *ATP8*, *COXI*, *COXIII*, *ND1*, *ND2*, *ND3*, *ND4*, *ND5*, and control region) showed lower bootstrap support (42%–68% for the two sister mtDNA lineages) with seven of these showing a sister-lineage relationship for *G. melli* and *G. subpalmatus*. The complete mitochondrial genome tree also supported *G. melli* and *G. subpalmatus* as sister groups with strong bootstrap support value (100%: see "MtDNA analyses previous data" file within Supporting Information 3).

3.2 | Divergence Times From mtDNA

Following removal of the first 30% of the MCMC chain, mean posterior divergence time between the sister lineages *G. subpalmatus* and *G. melli* was 15.40 million years (Ma) ago with a 95% highest posterior density (95% HPD) of 9.73–22.43 Myr (Figure 4). Estimated divergence time between *G. cib* and these lineages was just slightly older: 15.85 Myr ago (95% HPD: 9.88–22.98 Ma). The estimated clock rate for the *COXI* gene was 0.010 subs/site/Ma.

3.3 | Reduced Representation Genome Tree

GBS-SNP-CROP identified 12,650 SNPs from the GBS reads. Genbank accessions and other information are given in Supporting Information 1: Table S1. Further filtering reduced this to 10,059 concatenated SNPs (Section 2) across all individuals for the RAxML-NG v. 1.1.0 [21] analysis. The ML and Bayesian (25% of posterior sample removed as burnin) consensus trees had essentially the same topology as our *ND2*/*COXI* mtDNA tree although the main branches were better-supported than for the

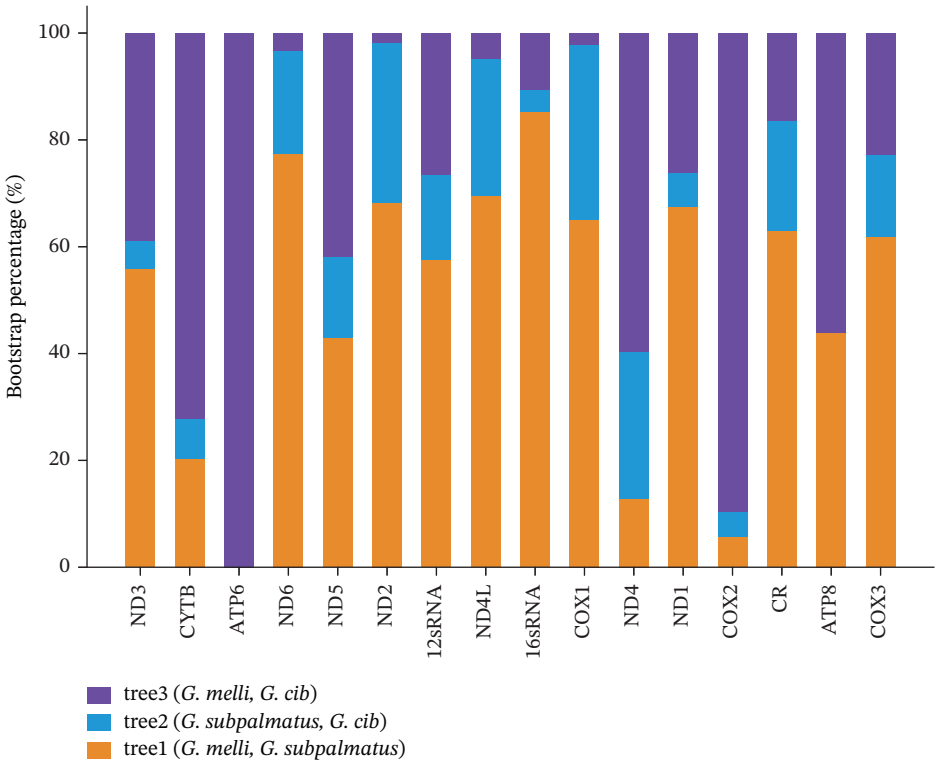

FIGURE 3 | Proportions of bootstrap replicates in ML analyses that support each of the three topologies for the 16 mtDNA loci analyzed. The topology that is supported by analyses of our mtDNA data and the SNP analyses is denoted "tree1."

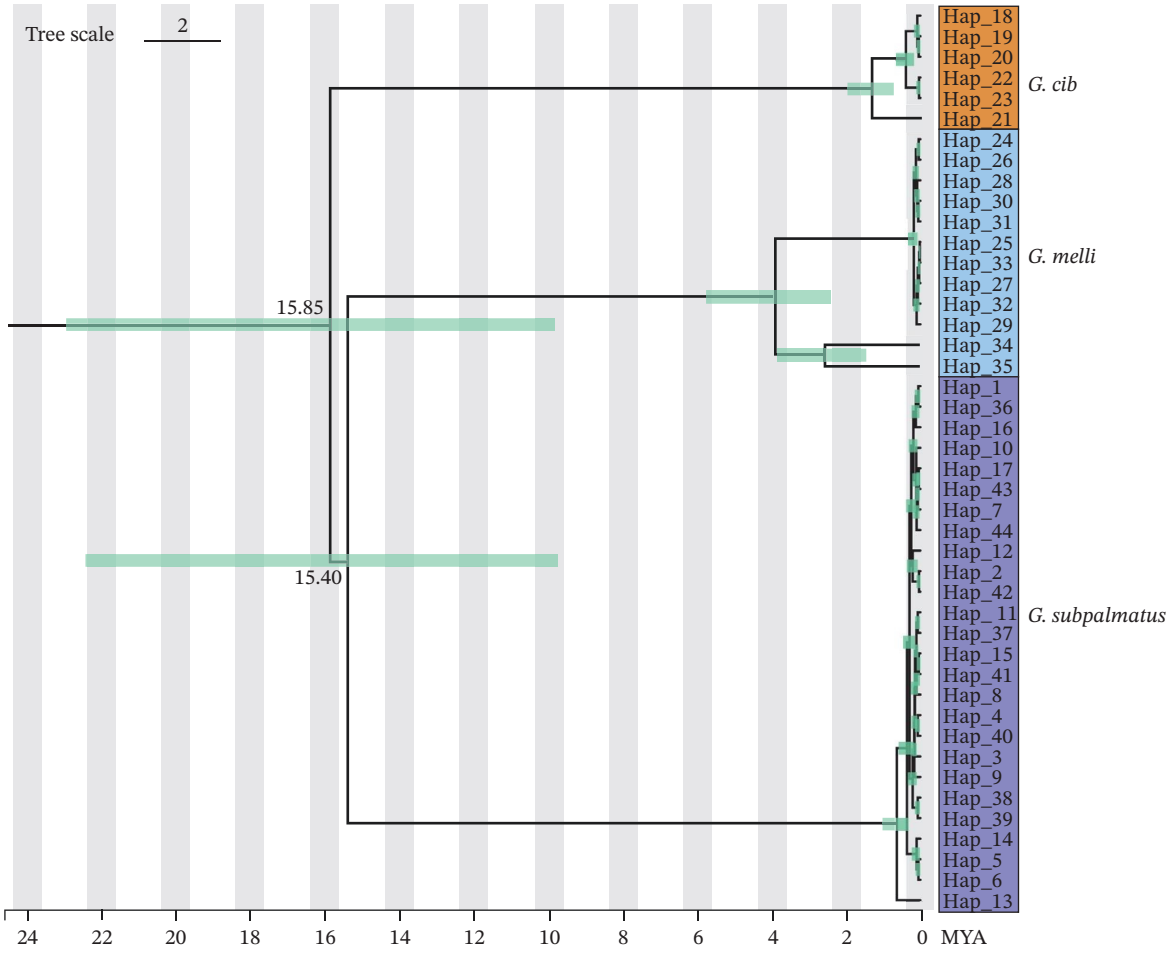

FIGURE 4 | Divergence times from mtDNA: The BEAST phylogenetic trees based on the haplotypes of mitochondrial *COXI* and *ND2* sequences of the *Gekko* species complex. The estimated divergence times (Myr) are shown near the nodes.

mtDNA tree (Figure 5). The ML and Bayesian consensus trees showed that *G. melli* and *G. subpalmatus* are monophyletic lineages (bootstrap support = 99% and posterior probability = 0.92) that together form a sister group to *G. cib*.

For the SNP species tree, 10% of posterior samples were discarded as burn-in. The overall species tree topology was unequivocally supported and provided the same relationships among lineages as those obtained for the *ND2*/*COXI* mtDNA tree (Figure 6). Only two topologies were detected in the post-burn-in samples: ((*G. melli* and *G. subpalmatus*) and *G. cib*) with a posterior probability of 0.99 (Figure 6), and (*G. melli* and (*G. subpalmatus* and *G. cib*)) with a posterior probability of 0.01.

The joint posterior distribution of species assignments supported all three species with a posterior probability of 1.0, with each of the three clusters having a marginal posterior probability of 1.0, strongly supporting species delimitation.

3.4 | Morphological Divergence

The ANCOVAs demonstrated significant linear relationships between each of the six log-transformed body dimensions and the $\log_{10}$-SVL covariate ($p < 0.001$ in all cases) and significant differences among the adjusted means of the three focal groups for the following body dimensions: Log_{10}-HL ($F_{2,111} = 59.42$, $p < 0.001$), $\log_{10}$-HW ($F_{2,111} = 16.14$, $p < 0.001$), $\log_{10}$-AL ($F_{2,109} = 6.00$, $p = 0.003$), $\log_{10}$-DBA ($F_{2,111} = 4.96$, $p = 0.009$), and $\log_{10}$-DBI ($F_{2,111} = 26.99$, $p < 0.001$). Log_{10}-LL was the only non-significant body dimension ($F_{2,111} = 2.41$, $p = 0.095$) but was still used in subsequent multivariate analyses as the significance was <0.1. The multivariate permutation test across the three focal groups was highly significant (Wilks Lambda = 18.323, $p = 0.0001$) indicating significant differences among the three groups. The npmv post hoc subsets test also rejected the hypothesis of equality between all factor levels using a closed multiple testing procedure at a Type I error rate of $\alpha = 0.001$ [42].

In the DFA, the first discriminant function (DF; 87.2% of total variation) was most heavily weighted by adjusted $\log_{10}$HL and adjusted $\log_{10}$DBI, with the latter also showing the heaviest loading on the second DF (see Supporting Information 6: Table S3 for details). Absolute values for the other coefficients were low. DF1–DF2 scatterplots (Figure 7) show clear morphological separation among the three groups, with the exception of G. cib, which occupies an intermediate position and overlaps partially with *G. melli* and *G. subpalmatus*. The greatest morphological similarity, as indicated by the Mahalanobis D^2 distance, was between *G. subpalmatus* and *G. cib* (0.0072), while the largest Mahalanobis D^2 distance was between *G. subpalmatus* and *G. melli* (0.0171; Table 1).

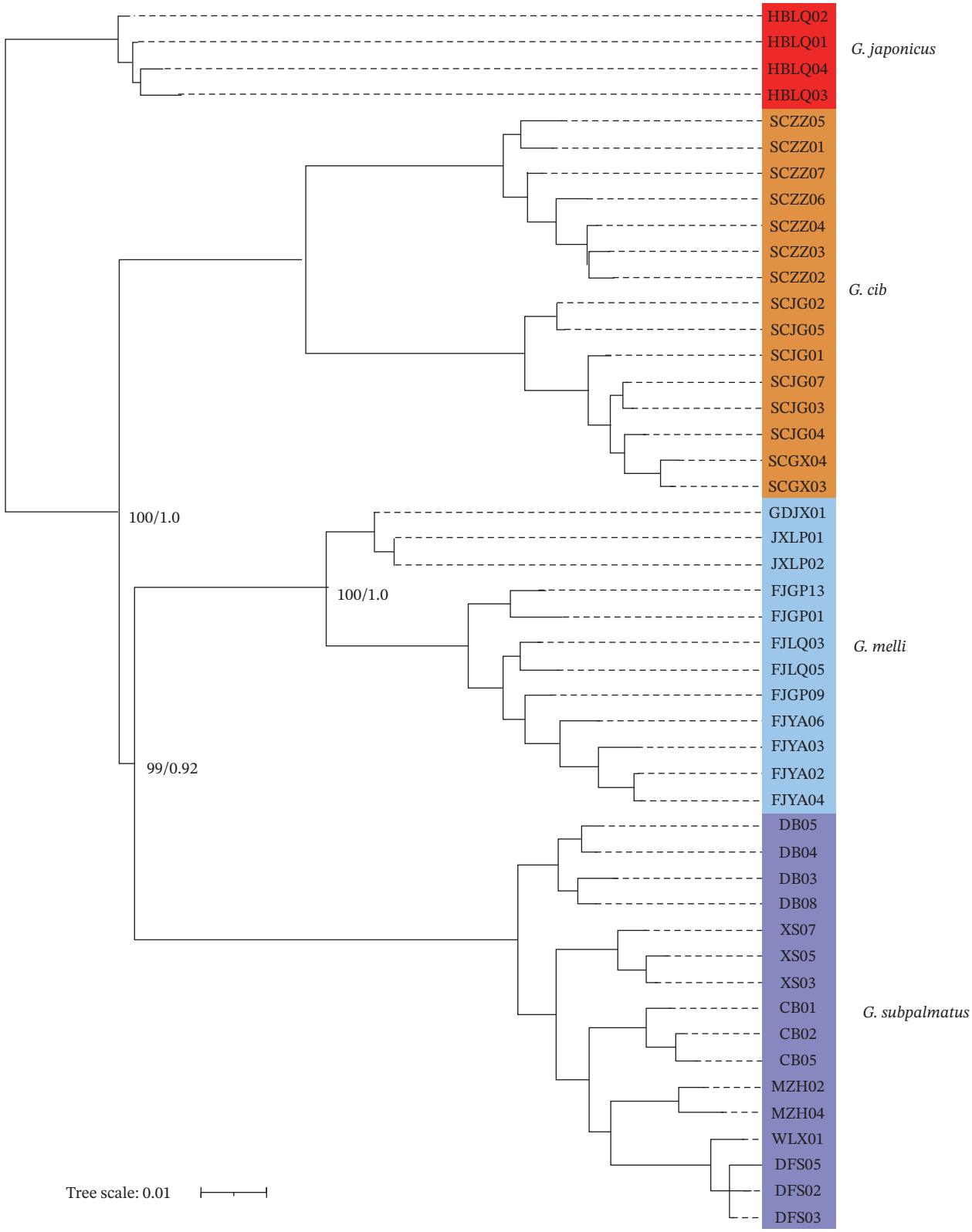

FIGURE 5 | SNP topologies obtained for RAxML and MrBayes phylogenetic trees (estimated from the concatenated SNPs). Node labels represent RAxML bootstrap values/MrBayes posterior probabilities.

4 | Discussion

A mtDNA analysis of three species of Chinese gecko provided a different mtDNA tree to that published previously by Lyu et al. [12]. Here, the mtDNA tree had the same topology as (i) the species tree we inferred from nuclear SNPs and (ii) a mtDNA tree inferred from a single representative mitochondrial genome from each species. However, we found that our favored tree was not consistently inferred when sequences from individual mtDNA

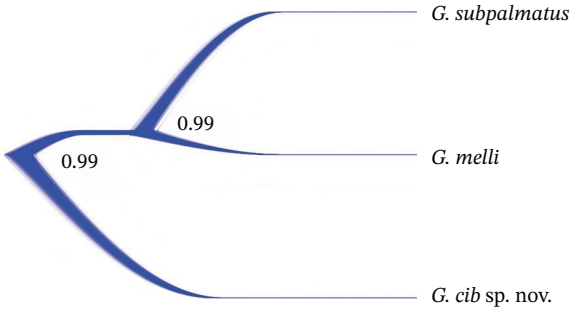

FIGURE 6 | The SNAPPER species tree estimated from SNPs (treated as independent loci), shown as a DensiTree plot of the posterior distribution.

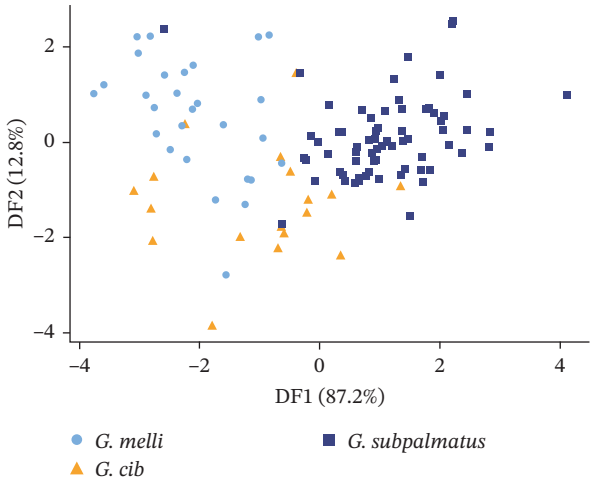

FIGURE 7 | Plot of individual scores on the first two discriminant functions (DF1 and DF2), from the discriminant function analysis of morphology.

regions were used, which appears to explain why the inferred phylogeny differed between studies.

Unlike unlinked nuclear DNA markers, differences among genes within the mtDNA cannot be due to differences in gene history. The result might appear unexpected given that the phylogenetic signal should increase over time across the genome and divergence times were shown to be quite ancient (i.e., possibly early or mid-Miocene). However it is likely explained by the finding of similar divergence times among the three main lineages, which makes correct inference of the topology more difficult. Relatively high divergence could lead to saturation of synonymous substitutions, which could cause differences in topology between protein-coding and other regions; however, our analyses rule out this possibility. Instead, it seems likely that similar divergence times combined with minor stochastic differences in phylogenetic signal between short mtDNA regions, combined with their lower individual phylogenetic information content, means that short sequences are susceptible to either failing to recover a well-supported topology or recovering the incorrect one. These findings raise significant questions about the reliability of short mtDNA sequences for reconstruction of phylogenetic histories.

Is the SNP phylogeny correct? It is known that large numbers of concatenated nuclear markers can provide incorrect but well-supported phylogenies [48, 49]. However, this can be overcome with multilocus multispecies coalescent methods [50]. Here, both the individual (concatenated SNPs), the multilocus multispecies coalescent phylogenies and also the whole mitochondrial genomes of three individuals all provided strong support for the same topology, indicating robust phylogenetic inference. We therefore demonstrate the power of the SNP approach (as well as the value of whole mitochondrial genomes).

Morphological differences reported here were largely concordant with those identified by Lyu et al. [12]. These authors used their findings to diagnose the three species and support their taxonomic proposals. It is interesting that the pattern of morphological divergence is not concordant with the SNP tree. *G. melli* and *G. subpalmatus* were the most divergent groups while *G. cib* was intermediate. We found that HL and DBIs had the greatest impact on among-group discrimination. *Gekko melli* has been described as having a larger body size and distinct color stripe [51], but here we find that it has a relatively shorter but wider head, relatively longer anterior limb length and is intermediate in terms of DBIs, relative to the two other groups. A natural assumption is that morphological divergence will generally increase over time, but the discordance with species history indicates that rates of divergence must differ between the different lineages. If rates of morphological divergence were equal, then our observations would support a species relationship between *G. cib* and *G. melli*. As this runs counter to our molecular analyses, we suggest that accelerated evolution might have led greater morphological divergence in *G. subpalmatus*. The finding of discordance between morphological divergence and molecular-based phylogenetic analyses is not unusual, and it has become clear how this has sometimes led to inappropriate morphological taxonomies at the species/subspecies levels [52–54].

To date, relatively few studies have performed species delimitation analyses using the SNAPPER/SPEEDEMON implementation that was applied here. Gene concordance, which generally forms the basis for statistical delimitation, should be treated as just one piece of evidence for delimiting species [55]. Nonetheless, it did provide very strong statistical support for three gecko species as first proposed by Lyu et al. [12]. However, we have now definitively established the historical relationships that underpin this taxonomy, which differ from those inferred by their study.

TABLE 1 | Mahalanobis distances between group centroids with associated 95% bootstrap confidence intervals, indicating that morphological characters showed greatest similarity between *Gekko cib* and *G. subpalmatus*.

Species	*G. subpalmatus*	*G. cib*
G. subpalmatus	—	—
G. cib	0.0072 (0.0047.0.0115)	—
G. melli	0.01701 (0.0120,0.0253)	0.01064 (0.0053, 0.0197)

Note: The groups are: *G. subpalmatus* (*G. subpalmatus*), *G. cib*, and *G. melli*.

5 | Conclusion

Following a long history of contentious systematics, we obtained sequence and species-trees using genome-wide SNPs and mtDNA analyses and unequivocally inferred that *G. melli* and *G. subpalmatus* from Zhejiang and the Zhoushan Islands are sister taxa that are outgrouped by *G. cib*. All lineages appear to have diverged in the early-mid Miocene. The described evolutionary history has not been proposed previously. Hence, we demonstrate the benefits of characterizing cross-genomic SNPs to enable robust phylogenomic inference and resolve uncertainties that may have arisen from studies of short mtDNA fragments and morphology. We highlight how discordant but well-supported phylogenies can be observed for different loci within the mtDNA. Although previous studies have proposed two or three species, taxonomic changes are now required because we now show that three species are supported. While a full taxonomic description of the three species is beyond the scope of our phylogenetic study, we do provide the evolutionary evidence to support this description.

Acknowledgments

Thanks to Jianchao Gao, Kun Yang, and Gang Shao for their assistance during sampling. This work was supported by the National Natural Science Foundation of China (Grant 32370441) and the Zhejiang Natural Science Foundation (Grant LY21C040002) and partially supported by the Guangxi Key Laboratory of Rare and Endangered Animal Ecology, Guangxi Normal University. We thank Ben Reiser for statistical discussions. AI was not used by any of the authors during the preparation of this work.

Funding

This work was supported by the National Natural Science Foundation of China (Grant 32370441) and the Zhejiang Natural Science Foundation (Grant LY21C040002) and partially supported by the Guangxi Key Laboratory of Rare and Endangered Animal Ecology, Guangxi Normal University.

Ethics Statement

All field sampling protocols and measurements were performed in accordance with guidelines from the China Council on Animal Care and approved by the Ethics Committee of Animal Experiments at China Jiliang University.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The complete set of genetic/genomic data accessions is provided in Supporting Information 1: Table S1. The SNP dataset is also provided in the compressed datafile (Supporting Information 3).

References

1. J. C. Avise, *Molecular Markers, Natural History and Evolution* (Springer Science and Business Media, 2012).

2. B. DeSalle, B. Schierwater, and H. Hadrys, "MtDNA: The Small Workhorse of Evolutionary Studies," *Frontiers in Bioscience* 22, no. 5 (2017): 873–887.

3. N. Galtier, B. Nabholz, S. Glemin, and G. D. D. Hurst, "Mitochondrial DNA as a Marker of Molecular Diversity: A Reappraisal," *Molecular Ecology* 18, no. 22 (2009): 4541–4550.

4. K. de Queiroz, "Systematics and the Darwinian Revolution," *Philosophy of Science* 55, no. 2 (1988): 238–259.

5. J. W. Wägele, *Foundations of Phylogenetic Systematics*, 2nd ed. (Verlag Dr. Friedrich Pfeil, 2005).

6. J. W. O. Ballard and D. M. Rand, "The Population Biology of Mitochondrial DNA and Its Phylogenetic Implications," *Annual Review of Ecology, Evolution, and Systematics* 36, no. 1 (2005): 621–642.

7. W. S. Moore, "Inferring Phylogenies From mtDNA Variation: Mitochondrial-Gene Trees Versus Nuclear-Gene Trees," *Evolution* 49, no. 4 (1995): 718–726.

8. D. Rubinoff and B. S. Holland, "Between Two Extremes: Mitochondrial DNA is Neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference," *Systematic Biology* 54, no. 6 (2005): 952–961.

9. W. Wüster, "Shedding the Mitochondrial Blinkers: A Long-Overdue Challenge for Species Delimitation in Herpetology," *Vertebrate Zoology* 75 (2025): 259–275.

10. D. P. Toews and A. Brelsford, "The Biogeography of Mitochondrial and Nuclear Discordance in Animals," *Molecular Ecology* 21, no. 16 (2012): 3907–3930.

11. J. Eberle, E. Bazzato, S. Fabrizi, et al., "Sex-Biased Dispersal Obscures Species Boundaries in Integrative Species Delimitation Approaches," *Systematic Biology* 68, no. 3 (2019): 441–459.

12. Z. T. Lyu, C. Y. Lin, J. L. Ren, et al., "Review of the *Gekko* (*Japonigekko*) *Subpalmatus* Complex (Squamata, Sauria, Gekkonidae), With Description of a New Species From China," *Zootaxa* 4951, no. 2 (2021): 236–258.

13. H. Rösler and F. Tiedemann, "*Gekko melli*, Vogt, 1922 and Its Types (Reptilia, Sauria, Gekkonidae)," *Zoosystematics and Evolution* 83, no. S1 (2007): 105–108.

14. J. H. Yang, Y. Y. Wang, T. D. Zhang, Y. J. Sun, and S. S. Lin, "Genetic and Morphological Evidence on the Species Validity of, *Gekko melli*, Vogt, 1922 With Notes on Its Diagnosis and Range Extension (Squamata: Gekkonidae)," *Zootaxa* 3505, no. 1 (2012): 67–74.

15. E. Zhao and K. Adler, *Herpetology of China* (Society for the Study of Amphibians and Reptiles, 1993).

16. P. L. Wood, X. G. Guo, S. L. Travers, et al., "Parachute Geckos Free Fall Into Synonymy: *Gekko* Phylogeny, and a New Subgeneric Classification, Inferred From Thousands of Ultraconserved Elements," *Molecular Phylogenetics and Evolution* 146 (2020): 106731.

17. K. Tamura, G. Stecher, and S. Kumar, "MEGA11: Molecular Evolutionary Genetics Analysis Version 11," *Molecular Biology and Evolution* 38, no. 7 (2021): 3022–3027.

18. J. Rozas, A. Ferrer-Mata, J. C. Sánchez-DelBarrio, et al., "DnaSP. 6: DNA Sequence Polymorphism Analysis of Large Data Sets," *Molecular Biology and Evolution* 34, no. 12 (2017): 3299–3302.

19. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott, "PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses," *Molecular Biology and Evolution* 34 (2016): 772-773.

20. F. Ronquist, M. Teslenko, P. Van Der Mark, et al., "MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space," *Systematic Biology* 61, no. 3 (2012): 539–542.

21. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, and A. Stamatakis, "RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference," *Bioinformatics* 35, no. 21 (2019): 4453–4455.

22. S. Hao, D. N. Yu, J. Ping, H. B. Zhou, and Y. P. Zhang, "Complete Mitochondrial Genomes of Two Gecko Species, *Gekko hokouensis* and *Gekko japonicus* (Squamata, Gekkonidae)," *Mitochondrial DNA Part B-Resources* 1, no. 1 (2016): 346-347.

23. S. Sitthivong, L. O. Van, T. Q. Nguyen, et al., "A New Species of the *Gekko japonicus* Group (Squamata: Gekkonidae) From Khammouane Province, Central Laos," *Zootaxa* 5082, no. 6 (2021): 553–571.

24. X. Xia, "DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution," *Molecular Biology and Evolution* 35, no. 6 (2018): 1550–1552.

25. H. D. Luo, A. Huang, B. Li, et al., "Complete Mitochondrial Genome of the Webbed-Toed Gecko *Gekko subpalmatus* (Squamata: Gekkonidae)," *Mitochondrial DNA Part B-Resources* 4, no. 1 (2019): 1725-1726.

26. D. Darriba, G. L. Taboada, R. Doallo, and D. Posada, "jModelTest 2: More Models, New Heuristics and Parallel Computing," *Nature Methods* 9, no. 8 (2012): 772.

27. S. Guindon and O. Gascuel, "A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood," *Systematic Biology* 52, no. 5 (2003): 696–704.

28. R. Bouckaert, T. G. Vaughan, J. Barido-Sottani, et al., "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis," *PLoS Computational Biology* 15, no. 4 (2019): e1006650.

29. Y.-T. Jin and N.-F. Liu, "Phylogeography of *Phrynocephalus erythrurus* From the Qiangtang Plateau of the Tibetan Plateau," *Molecular Phylogenetics and Evolution* 54, no. 3 (2010): 933–940.

30. M. Pepper, P. Doughty, and J. S. Keogh, "Molecular Phylogeny and Phylogeography of the Australian *Diplodactylus stenodactylus* (Gekkota; Reptilia) Species-Group Based on Mitochondrial and Nuclear Genes Reveals an Ancient Split Between Pilbara and Non-Pilbara *D. stenodactylus*," *Molecular Phylogenetics and Evolution* 41, no. 3 (2006): 539–555.

31. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, "Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7," *Systematic Biology* 67, no. 5 (2018): 901–904.

32. M. Schubert, S. Lindgreen, and L. Orlando, "AdapterRemoval v2: Rapid Adapter Trimming, Identification, and Read Merging," *BMC Research Notes* 9, no. 1 (2016): 88.

33. S. Andrews, "FastQC: A Quality Control Tool for High Throughput Sequence Data," 2010, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

34. A. T. O. Melo, R. Bartaula, and I. Hale, "GBS-SNP-CROP: A Reference-Optional Pipeline for SNP Discovery and Plant Germplasm Characterization Using Variable Length, Paired-End Genotyping-by-Sequencing Data," *BMC Bioinformatics* 17, no. 1 (2016): 29.

35. Y. Liu, Q. Zhou, Y. Wang, et al., "*Gekko japonicus* Genome Reveals Evolution of Adhesive Toe Pads and Tail Regeneration," *Nature Communications* 6, no. 1 (2015): 10033.

36. A. D. Leaché, M. K. Fujita, V. N. Minin, and R. R. Bouckaert, "Species Delimitation Using Genome-Wide SNP Data," *Systematic Biology* 63, no. 4 (2014): 534–542.

37. M. Stoltz, B. Baeumer, R. Bouckaert, C. Fox, G. Hiscott, and D. Bryant, "Bayesian Inference of Species Trees Using Diffusion Models," *Systematic Biology* 70, no. 1 (2021): 145–161.

38. J. Douglas and R. Bouckaert, "Quantitatively Defining Species Boundaries With More Efficiency and More Biological Realism," *Communications Biology* 5, no. 1 (2022): 755.

39. A. J. Drummond and A. Rambaut, "BEAST: Bayesian Evolutionary Analysis by Sampling Trees," *BMC Evolutionary Biology* 7, no. 1 (2007): 214.

40. R. R. Bouckaert, "DensiTree: Making Sense of Sets of Phylogenetic Trees," *Bioinformatics* 26, no. 10 (2010): 1372-1373.

41. K. O. Chan and L. L. Grismer, "A Standardized and Statistically Defensible Framework for Quantitative Morphological Analyses in Taxonomic Studies," *Zootaxa* 5023, no. 2 (2021): 293–300.

42. W. W. Burchett, A. R. Ellis, S. W. Harrar, and A. C. Bathke, "Nonparametric Inference for Multivariate Data: The *R* Package npmv," *Journal of Statistical Software* 76, no. 4 (2017): 1–18.

43. R Core Team, "R: A Language and Environment for Statistical Computing," 2023, R Foundation for Statistical Computing https://www.R-project.org/.

44. B. Reiser, "Confidence Intervals for the Mahalanobis Distance," *Communications in Statistics–Simulation and Computation* 30, no. 1 (2001): 37–45.

45. L. McFerrin, "_HDMD: Statistical Analysis Tools for High Dimension Molecular Data (HDMD)," 2013, R package version 1.2 https://CRAN.R-project.org/package=HDMD.

46. A. Canty and B. Ripley, "Boot: Bootstrap R (S-Plus) Functions," 2022, R Package Version 1.3-28.1 https://cran.r-project.org/web/packages/boot/boot.pdf.

47. D. M. Hillis and J. J. Bull, "An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis," *Systematic Biology* 42, no. 2 (1993): 182–192.

48. J. H. Degnan and N. A. Rosenberg, "Discordance of Species Trees With Their Most Likely Gene Trees," *PLoS Genetics* 2, no. 5 (2006): e68.

49. L. S. Kubatko and J. H. Degnan, "Inconsistency of Phylogenetic Estimates From Concatenated Data Under Coalescence," *Systematic Biology* 56, no. 1 (2007): 17–24.

50. L. Liu and S. V. Edwards, "Phylogenetic Analysis in the Anomaly Zone," *Systematic Biology* 58, no. 4 (2009): 452–460.

51. R. Herbert, Z. Thomas, N. T. Vu, H. Hans-Werner, and B. Wolfgang, "A New Lizard of the Genus, *Gekko*, Laurenti, 1768 (Squamata: Sauria: Gekkonidae) From the Phong Nha–Ke Bang National Park, Quang Binh Province, Vietnam," *Bonner zoologische Beiträge* 53 (2005): 135–148.

52. R. B. Harris, P. Alström, A. Ödeen, and A. D. Leaché, "Discordance Between Genomic Divergence and Phenotypic Variation in a Rapidly Evolving Avian Genus (*Motacilla*)," *Molecular Phylogenetics and Evolution* 120 (2018): 183–195.

53. R. A. Laroche, D. P. Duran, C. A. Lee, et al., "A Genomic Test of Subspecies in the *Eunota togata* Species Group (Coleoptera: Cicindelidae): Morphology Masks Evolutionary Relationships and Taxonomy," *Molecular Phylogenetics and Evolution* 189 (2023): 107937.

54. P. P. Sharma, R. Fernández, L. A. Esposito, E. González-Santillán, and L. Monod, "Phylogenomic Resolution of Scorpions Reveals Multilevel Discordance With Morphological Phylogenetic Signal," *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1804 (2015): 20142953.

55. K. de Queiroz, "Species Concepts and Species Delimitation," *Systematic Biology* 56, no. 6 (2007): 879–886.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information 1. Table S1. Complete sampling site information and morphological data.

Supporting Information 2. Table S2. Primer pairs for PCR amplification of mitochondrial genes ND2 and COXI.

Supporting Information 3. Data File S1. Data files containing: COX-ND2 haplotypes and corresponding GenBank accessions) (COXI_ND2_hap.fas); total SNPs and corresponding GenBank accessions (total SNPs.fasta); filtered SNPs used for RAxML-NG (filtered SNPs for RAxML running.phy); the SPEEDEMON data block (speedemon data block.xml); MtDNA genomes and analyses on previous data (MtDNA analyses previous data.docx); BEAST input for divergence time dating (BEAST2 Strict Clock Divergence Time Analysis Input.xml); R code for obtaining bootstrapped

95% confidence intervals on Mahalanobis distances (R_code_bootstrapped_mahalanobis.docx) can be downloaded as a zip file.

Supporting Information 4. Table S4. Table of the best-fit models for different mitochondrial genes.

Supporting Information 5. Table S5. Table of optimal models for partitions identified for whole mitochondrial genome analysis.

Supporting Information 6. Table S3. Pooled within-groups correlation matrix and standardized canonical discriminant function coefficients for adjusted body dimensions.