Zoological Journal of the Linnean Society, 2025, **204**(**4**), zlaf096
https://doi.org/10.1093/zoolinnean/zlaf096

OXFORD

Original Article

Unravelling the cryptic diversity of the low-dispersal land planarian *Cephaloflexa bergi* to elucidate diversification processes in the Brazilian Atlantic Forest

Marina Lorenzo[1], Ricard Sabaté[1], Marta Álvarez-Presas[1,2], Fernando Carbayo[3], Marta Riutort[1,*]

[1]Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, Barcelona, Catalonia, 08028, Spain
[2]Institut de Biologia Evolutiva (CSIC-UPF), Pg. Marítim de la Barceloneta, 37-49, Barcelona, Catalonia, 08003, Spain
[3]Laboratório de Ecologia e Evolução, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo (USP), Av. Arlindo Bettio 1000, São Paulo, SP, 03828-000, Brazil

*Corresponding author. Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av Diagonal 643, Catalonia, Barcelona, 08028, Spain. E-mail: mriutort@ub.edu

ABSTRACT

The terrestrial flatworm *Cephaloflexa bergi* (Tricladida: Platyhelminthes) is an excellent model for elucidating the evolutionary processes shaping biodiversity in the Brazilian Atlantic Forest (BAF). Its limited dispersal and sensitivity to environmental changes suggest that its genetic structure may reflect the evolutionary history of the forest biota. Its high interpopulation diversity questions whether it may represent a species complex. Our objective is to conduct a genetic analysis to assess the existence of multiple species, to evaluate population structure at a finer scale than previously reported, and to explore the drivers of its diversification. Specimens were collected from protected areas within the southern BAF, and the *COI* gene and three newly developed nuclear markers were sequenced. Our phylogenetic and population genetic analyses revealed two clades with high intra- and interpopulation diversity, supporting an ancient origin for the species linked to Neogene geological processes. The contact zone previously observed was restricted to a coastal region, pointing to the Quaternary sea-level fluctuations as the potential driver. Intra-individual diversity analyses provide the first evidence of the mosaic Meselson effect in land planarians. Overall, our species' delimitation suggests that *C. bergi* potentially comprises multiple species and highlights the importance of integrating diverse molecular approaches to understand the BAF's biodiversity.

Keywords: genetic mosaicism; land planarians; phylogeography; population structure; species' delimitation

INTRODUCTION

The Brazilian Atlantic Forest (BAF) is considered the planet's fourth biodiversity hotspot (Myers *et al.* 2000), but it is also one of the most threatened ecosystems due to habitat fragmentation (Ribeiro *et al.* 2009). Some studies have reported that up to 92% of the original vegetation cover has been lost since European colonization over 500 years ago (Myers *et al.* 2000, Rezende *et al.* 2018). Despite this concerning trend, most of the remaining vegetation persists in mountain fragments larger than 50 hectares, which play a crucial role in the conservation of forest-dependent species (Bicudo Da Silva *et al.* 2020). In response to the challenges arising from the scenario of forest fragmentation, several mountain ranges have been designated as protected areas (PAs). However, recent reports indicate that only 18% of these PAs overlap with climatically stable regions, highlighting the low efficiency of these areas in preserving the endemic fauna (Sobral-Souza *et al.* 2018). To establish optimal strategies for maintaining and protecting the remaining biome in the BAF, we need to uncover the highly heterogeneous biodiversity and to understand the evolutionary history of extant species.

Received 05 February 2025; revised 27 May 2025; accepted 12 June 2025

In the past decade, studies focusing on land flatworms (Geoplaninae, Geoplanidae, Tricladida) in the BAF have been essential for identifying areas of endemism (Lago-Barcia *et al.* 2020, 2024), understanding diversification patterns (Álvarez-Presas *et al.* 2011, 2014), and assessing the conservation status of forests (Sluys 1999, Negrete *et al.* 2014). The high habitat dependence and low dispersal abilities of these invertebrates between isolated forest fragments facilitates population structure and presumably contributes to the presence of numerous endemic species, particularly in humid landscapes (Carbayo *et al.* 2002, Lago-Barcia *et al.* 2024).

On the other hand, terrestrial flatworms, although being hermaphroditic, reproduce sexually through copulation and cross-fertilization (Winsor *et al.* 2004). Nonetheless, their limited mobility may difficult finding mates, prompting the need to adopt an alternative reproductive strategy. Asexual reproduction through fission and regeneration has been documented in some non-geoplaninid species (Hyman 1954, Jones *et al.* 2020). However, the regenerative abilities of Geoplaninae are not yet fully understood, raising the question of whether these individuals may occasionally reproduce through fission to overcome their mobility constraints. A recent study on freshwater planarians has discovered that fissiparous individuals exhibit multiple distantly related haplotypes at similar frequencies, a molecular signature known as the mosaic Meselson effect (Leria *et al.* 2019). The presence of this intra-individual genetic pattern may suggest the occurrence of asexual reproduction, but it has yet to be verified in land planarians.

Previous phylogeographic and population genetic studies have identified the terrestrial flatworm *Cephaloflexa bergi* (Graff, 1899) (Fig. 1) as a valuable model organism for investigating the evolutionary dynamics of the biodiversity within the BAF (Álvarez-Presas *et al.* 2011, 2014). In the southern region of the Atlantic Forest (SAF), *C. bergi* exhibits high levels of nucleotide diversity, strong population structure, little or no significant evidence of recent expansion, and isolation-by-distance. These characteristics were particularly unexpected for a low-mobility organism in a recently colonized area, suggesting that its diversity probably has an ancient origin rather than resulting from a recent colonization event. Specifically, the genetic divergence among three main clades within the species–designated as N-SAF, C-SAF, and S-SAF for the northern, central, and southern SAF populations, respectively, was inferred to predate the early Miocene epoch, dating back approximately 8.5 million years ago (Mya). In contrast, the coexistence of highly divergent populations within a single locality suggested a recent secondary contact between the genetically isolated C-SAF and N-SAF clades, probably influenced by the Pleistocene glacial cycles (2.58 to 0.0117 Mya). The low sampling density in these areas, where secondary contacts and high diversity have been observed, underscores the necessity for more detailed analyses at a finer geographical scale. Moreover, deep coalescent times within these three lineages and strong population structure raise the question of whether *C. bergi* may indeed represent a species complex. However, no molecular species' delimitation has been conducted to determine whether the observed diversity corresponds to distinct species or divergent lineages within *C. bergi*.

Therefore, we aim to conduct a finer scale analysis of the variability observed in *Cephaloflexa bergi* building upon previous research (Álvarez-Presas *et al.* 2014). We specifically focus on a smaller region covering part of C-SAF and N-SAF, where a secondary contact has occurred, situated within the states of São Paulo and southern Rio de Janeiro (Fig. 2). Our goal is to elucidate the role of geographic barriers in shaping species' diversification and to evaluate the extension of the secondary contact region. Our study includes new sampling localities, and newly developed nuclear markers (*DOM4*, *DOM5*, and *DOM6*) to improve the resolution of previous phylogenetic analyses that only relied on the mitochondrial cytochrome *c* oxidase I (*COI*) and the nuclear ribosomal internal transcribed spacer-1 (ITS1) markers (Álvarez-Presas *et al.* 2014). Second, we aim to explore whether *C. bergi* represents multiple species, thereby providing deeper insight into the hidden diversity and evolutionary processes shaping the biota of the Atlantic Forest.

MATERIALS AND METHODS

Sampling localities and specimen's identification

Specimens of *Cephaloflexa bergi* were collected from 15 PAs across the states of São Paulo and southern Rio de Janeiro in the SAF (Fig. 2) between 2009 and 2016 (Supporting Information, Table S1). The localities are situated within the Serra da Mantiqueira and Serra do Mar Mountain ranges. Each specimen was assigned a unique code and dissected into two portions preserved in 10% formalin for morphological analyses and in absolute ethanol at −80°C for molecular analyses. Specimens were preliminary identified by their external morphology. Histological sections of two specimens were prepared following Carbayo and Almeida (2015) and subsequently examined by light microscopy to confirm their identity.

DNA extraction, amplification, and sequencing

DNA was extracted employing the Wizard® Genomic Purification kit (Promega Corporation). Specific primers were used to amplify four molecular markers with polymerase chain reaction (PCR): a fragment of the mitochondrial cytochrome *c* oxidase I gene (~820 bp—*COI*), a nuclear marker called *DOM5* (~840 bp) responsible for encoding a fragment of the heat shock protein 70 (hsp70) (Lago-Barcia *et al.* 2023), and two novel anonymous nuclear markers, *DOM4* (~430 bp) and *DOM6* (~720 bp), tested here for the first time. The *DOM4* locus encodes for another fragment of hsp70, while the coding region of *DOM6* codifies for the lactate dehydrogenase enzyme. *DOM4* and *DOM6* markers were developed employing the DOMINO program (Development of Molecular Markers in Non-model Organisms) (Frías-López *et al.* 2016) by using scaffolds from a low-coverage genome (see Supporting Information, File S1 for details). The amplification of *COI* and *DOM* markers was performed under conditions and primers outlined in Supporting Information, Tables S2 and S3, respectively. PCR products were purified using the MultiScreen® PCR96 Vacuum Manifold (Millipore Corporation) equipped with an AcroPrep™ Advance 96 Filter Plate 30K Omega™ (Pall Corporation) or using ExoSAP-IT™ PCR Cleanup Reagent (Applied Biosystems™). Finally, purified products were sequenced in both directions by Macrogen Inc. Spain, using the PCR primers.

Sequence alignments and saturation test

Sequence quality was checked by analysing peak patterns and contigs were assembled in GENEIOUS R10 (https://www.geneious.com). Nucleotide alignments for each locus were generated in MAFFT v.7 (Katoh and Standley 2013), and translated into amino acids to verify the reading frame using BioEdit v.7.0.9.0

Figure 1. *Cephaloflexa bergi* specimens photographed from three localities. A, Sertão do Cambury (CAM) from the eastern clade. B, Caraguatatuba (SMC) from the eastern clade. C, CAM from the western clade. D, SMC from the western clade. E, Campos do Jordão (CJ) from the western clade. Codes for each individual are represented in each image.

(Hall 1999). The echinoderm mitochondrial genetic code (table 9) was used for *COI* and the universal code for *DOM* markers. Exonic and intronic regions of the *DOM6* gene were identified to enable future analyses using them separately (Supporting Information, File S2). The proportion of invariant sites was estimated, and a saturation test was applied (Xia *et al.* 2003) with 60 replicates for each gene using DAMBE v.7.3.32 (Xia 2018), considering the three codon positions collectively and individually.

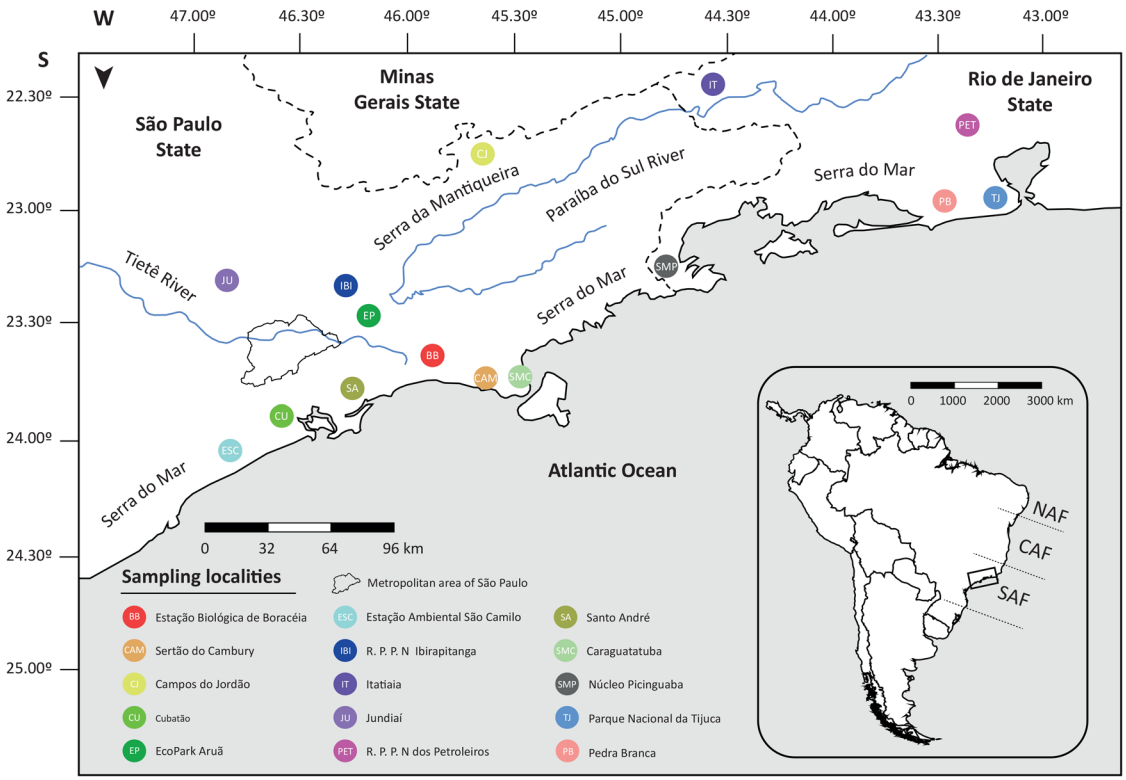

Figure 2. Map of South America showing three stability areas of the Brazilian Atlantic Forest as determined by Carnaval and Moritz (2008): Northern (NAF), Central (CAF), and Southern Atlantic Forest (SAF). A small region of the SAF is highlighted, showcasing the 15 sampled localities situated in two prominent mountain ranges (Serra da Mantiqueira and Serra do Mar), separated by the metropolitan area of São Paulo along with two main rivers (Paraíba do Sul and Tietê). Abbreviations correspond to the localities listed in Supporting Information, Table S1. Map was created using SimpleMappr (https://www.simplemappr.net) and edited in Illustrator CC® v.2015.2.1.

Table 1. Summary of the number of individuals and alignment base pairs length (bp) for each dataset used in the phylogenetic trees.

Dataset	Total individuals	Length (pb)	SNPs used
COI* + *DOMs	132	2823	380
EASTERN CLADE	43	2823	214
WESTERN CLADE	89	2823	286
DOMs	130	1999	221
EASTERN CLADE	43	1999	105
WESTERN CLADE	87	1999	156
COI	129	824	160
EASTERN CLADE	38	824	113
WESTERN CLADE	91	824	127
DOM4	132	438	
DOM5	110	841	
DOM6	132	720	

The last column indicates the number of single nucleotide polymorphisms (SNPs) used in the principal component analysis (PCA), with eastern and western individuals considered separately for each different dataset.

Phylogenetic inference

Phylogenetic relationships were inferred using six datasets (Table 1): one alignment for each of the four loci alone (*COI*, *DOM4*, *DOM5*, and *DOM6*), one concatenated alignment of three nuclear markers (*DOMs: DOM4* + *DOM5* + *DOM6*), and one concatenated alignment of four loci (*COI* + *DOMs: COI* + *DOM4* + *DOM5* + *DOM6*). Only those individuals with at least two sequenced loci were included in the concatenated datasets (Supporting Information, Table S1). Phylogenetic trees were inferred using different outgroup species depending on the data availability (Supporting Information, Table S4).

Phylogenetic reconstructions were performed using maximum likelihood (ML) and Bayesian inference (BI). For the ML approach, IQ-TREE 2 v.2.2.0 (Minh *et al.* 2020) was used with 10000 ultrafast bootstrap replicates. The best evolutionary models were estimated by ModelFinderPlus (MFP), applying the Bayesian information criterion (Supporting Information, Table S5). BI analyses were performed using MrBayes v.3.2.1 (Ronquist *et al.* 2012) with 5000000 generations, saving sampling parameters every 1000 generations. The convergence of both runs was confirmed by the average standard deviation of split frequencies being less than 0.01. The GTR+I+G model was chosen for all datasets as it was the closest available option to the model proposed by jModelTest v.0.1.1 (Posada 2008) based on the Akaike information criterion. For both methods, the inference of phylogenetic trees using the single-locus datasets entailed the assumption that (i) all positions follow the same evolutionary model or (ii) partitioning the data based on the three codon positions. Concatenated datasets were partitioned by locus or by locus and codon positions. The visualization of the trees was performed with the FigTree v.1.4.4 program (http://tree.bio.ed.ac.uk/software/figtree/) and edited in Illustrator CC® v.2015.2.1.

Cloning and haplotype networks

To assess potential mosaicism (Leria *et al.* 2019), PCR-purified *DOM4* products were cloned using the TOPO TA Cloning® Kit (Invitrogen, California, USA). One or two individuals per locality were cloned, prioritizing phylogenetically differentiated individuals within each locality when two were analysed. One to ten colonies per individual were PCR amplified and Sanger-sequenced using T3 and T7 universal primers. PCR amplification was performed as follows: initial DNA denaturation of 95°C 5 min, 40 cycles of 95°C 30 s, 45°C 45 s, 72°C 90 s, and a final extension of 72°C 7 min. Several alignments were generated: one for each locality and one for the nine localities together (*DOM4*-cloned dataset). DnaSP v.6.12.03 (Rozas *et al.* 2017) generated haplotype files from these alignments, and NETWORK v.10.2 (https://www.fluxus-engineering.com) computed haplotype networks. Singleton sequences that differ by one mutation from other non-singleton sequences were re-coded as the latter to minimize errors introduced by DNA polymerases during the initial PCR, the cloning PCR, or the sequencing reaction. Subsequently, haplotype networks were reconstructed and edited with Adobe Illustrator CC® v.2015.2.1.

Population genetic analyses

DnaSP was used to quantify the number of haplotypes (h), haplotype diversity (H_D), number of polymorphic positions (S), and nucleotide diversity (π) for the single-locus datasets considering each geographical locality, or each monophyletic group detected in the phylogenetic trees. S and π were also calculated for the *DOMs* dataset. Mean H_D and π were computed per locality or monophyletic group for single nuclear locus (*DOM4*, *DOM5*, and *DOM6*) and four loci (*COI*, *DOM4*, *DOM5* and *DOM6*). Additionally, mean H_D and π were computed for single-locus datasets considering all localities or monophyletic groups as a single entity. Genetic parameters were also calculated for *DOM4*-cloned dataset, at population and intra-individual level. To assess whether the populations were evolving neutrally or under selective pressure, Tajima's D (Tajima 1989) neutrality test was estimated for single-locus datasets and for *DOMs* using DnaSP software. The overall linkage disequilibrium was assessed for single-locus datasets using the Z_{nS} statistic (Kelly 1997), except for *COI* as it does not recombine.

F_{ST} (fixation index) matrices for *COI + DOMs*, *DOMs*, and *COI* datasets were calculated using ARLEQUIN v.3.5.2.2 (Excoffier *et al.* 2005), defining populations by geographical localities or monophyletic groups. F_{ST} values > 0.25 (Wright 1978) were considered high and statistically significant when p-value < 0.05. A heatmap of genetic distances (F_{ST}) for the monophyletic groups using *COI + DOMs* dataset was generated with the *ape* R package (Paradis and Schliep 2019). Isolation-by-distance signal was assessed with a Mantel test to evaluate the correlation between pairwise geographic and genetic distances (F_{ST}) using 999 permutations in GenAlex v.6.503 (Peakall and Smouse 2006). Population genetic structure was evaluated by principal component analysis (PCA) using single nucleotide polymorphisms of *COI + DOMs*, *DOMs* and *COI* datasets with the *fasta2genlight* function of the *adegenet* R package (Jombart and Collins 2015), based on original localities and monophyletic groups (Table 1). PCAs were also computed for the main clusters independently. STRUCTURE 2.3.4 (Pritchard *et al.* 2000) software was employed to assess admixture between specimens using *COI + DOMs* dataset. Individuals were grouped into two subsets, corresponding to the East–West division observed in the phylogenetic trees, excluding those with admixture. We performed 100000 Markov chain Monte Carlo (MCMC) replicates after a burn-in of 20000 iterations, each simulation was run 10 times for K = 1 to K = 16 without information about predefined clusters. Results were submitted to the Structure Harvester online program (http://taylor0.biology.ucla.edu/structureHarvester/) (Earl and vonHoldt 2012) to determine the optimal value of K using the ΔK method (Evanno *et al.* 2005).

Species' delimitation analyses

Discovery methods

The species' delimitation methods multi-rate Poisson tree processes (mPTP) (Kapli *et al.* 2017) and generalized mixed Yule-coalescent (GMYC) (Pons *et al.* 2006) were applied to the *COI* dataset. A *COI*-based ML tree was used as an input for the mPTP method conducted in the online server (https://mptp.h-its.org) after removing the outgroup species. For the GMYC approach, optimal priors were determined in BEAST v.1.10.4 (Suchard *et al.* 2018) by comparing marginal likelihood estimates (MLE) across different combinations of molecular clocks and tree models, using path sampling (PS) and stepping-stone methods (SS) (Baele *et al.* 2012), following the framework established by Kass and Raftery 1995. To assess the phylogenetic signal of the data, we ran the best model with and 'without' the data (sampling from the priors). We then evaluated the distribution of posterior and prior density probabilities for both runs using TRACER v.1.7.1 (Rambaut *et al.* 2018) to determine whether the priors influenced the data. After performing 10000000 iterations of the MCMC chains and sampling every 1000 generations, we used TRACER to verify that the effective sample size (ESS) for each parameter exceeded 200. The burn-in process was then performed using LogCombiner v.1.10.4 (Suchard *et al.* 2018), discarding the initial 10% of the trees. Finally, we generated a consensus ultrametric tree using TreeAnnotator v.1.10.4 (Suchard *et al.* 2018), which served as input for the single-threshold GMYC analysis (Fujisawa and Barraclough 2013) using the SPLITS R package (SPecies Limits by Threshold Statistics) (Ezard *et al.* 2009).

Validation approaches

Two validation methods were explored using multi-locus datasets: Bayesian phylogenetics and phylogeography (BPP) software (Yang 2015) and Bayes factor delimitation (BFD) approach (Grummer *et al.* 2014). For both, the starting species' hypothesis derived from GMYC results. BPP was performed excluding the outgroup species and using *COI + DOMs* and *DOMs* datasets with the A10 model (Yang and Rannala 2010, Rannala and Yang 2013) implemented in BPP v.4.7 (Flouri *et al.* 2018). Different prior combinations were tested to build either a more conservative or permissive model to estimate the posterior probabilities of the splitting events for our candidate species' hypotheses. Two parameters were included, which follow an inverse gamma prior distribution (IG (α, β)): the ancestral population size (θ) and

the relative coalescent times (τ). We tested four prior combinations of θ and τ for each dataset following the guidelines of Yang (2015): (i) a relatively small ancestral population size and recent divergences [M1: IG(3, 0.002) for θ and τ], (ii) a relatively large ancestral population size and deep divergences [M2: IG(3, 0.1) for θ and τ], (iii) large ancestral population sizes and shallow divergences among species [M3: IG(3, 0.1) for θ and IG(3, 0.002) for τ], and (iv) vice versa [M4: IG(3, 0.002) for θ and IG(3, 0.1) for τ]. Each combination was run under rjMCMC algorithm 0 (with a fine-tunning prior ε = 2) and algorithm 1 (with fine-tuning priors α = 2 and m = 1). The rjMCMC chains were run for 500000 generations with a burn-in of 10000 and a sampling interval of five generations. Groups were classified as putative species when the corresponding PP values of the nodes exceeded a threshold of 0.95. For BFD analyses, nine species' delimitation hypotheses were tested based on different groupings using *COI + DOMs*, and *COI* datasets. Individuals were initially assigned to distinct species for each scenario. Models 1 and 2 were based on the hypothetical groups determined by GMYC and mPTP, respectively (Fig. 3). Model 3 considered *Cephaloflexa bergi* as a single species, and model 4 tested the East–West division observed in the phylogenetic trees as two candidate species. The other combination models were derived from successively splitting the groups into more clusters, as observed in the F_{ST}, PCAs, and STRUCTURE analyses. Additionally, *COI + DOMs* dataset was split into two subsets (East and West) that were independently tested for different models. For all species' hypotheses, we used the GTR substitution model, an uncorrelated log-normal relaxed clock, and a birth–death speciation tree prior. MLEs were evaluated to assess whether the best-fitting model is significantly better than the rest (Kass and Raftery 1995).

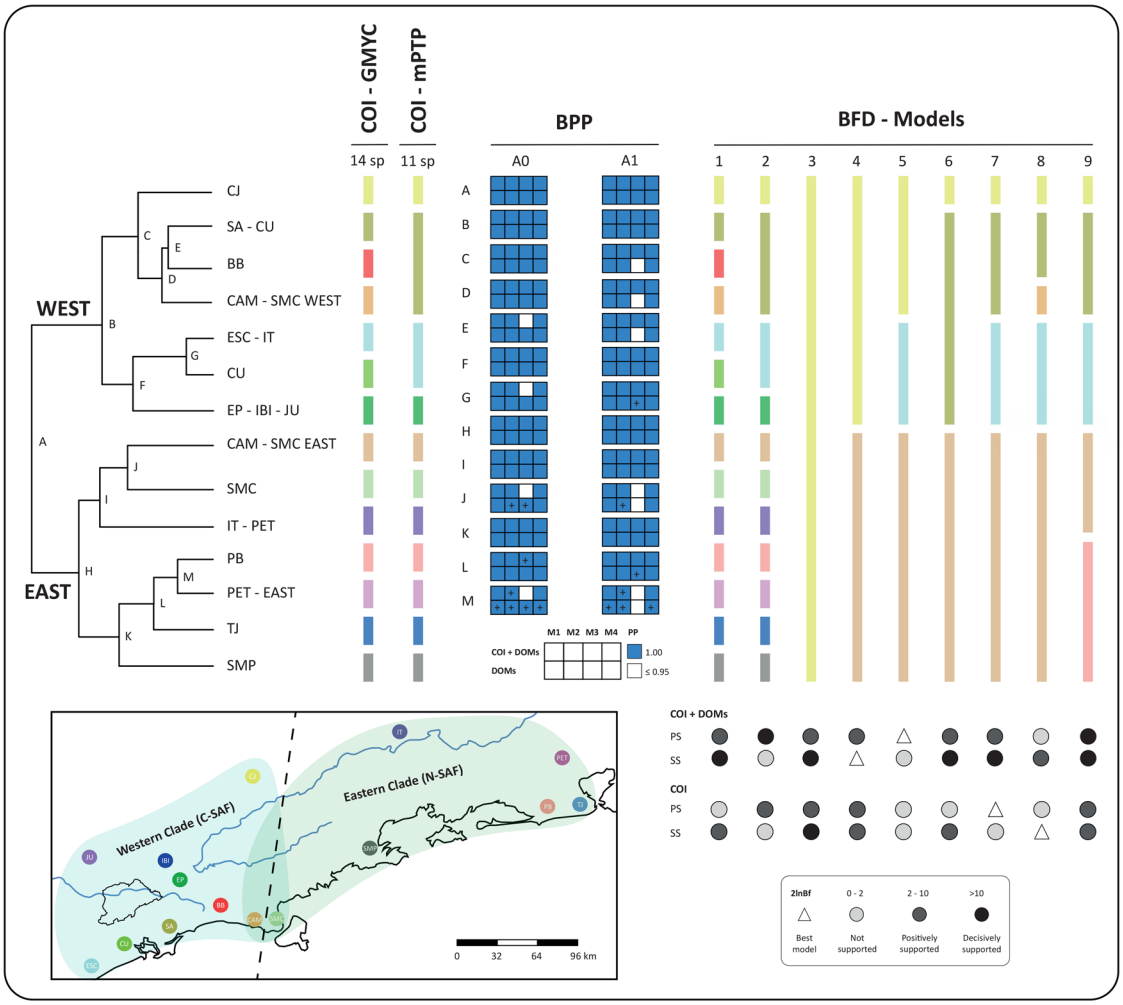

Figure 3. Dendrogram based on *COI + DOMs* dataset, followed by species' delimitation results from discovery methods (GMYC and mPTP), and validation approaches (BPP and BFD). For BPP, dendrogram nodes are labelled by a letter and colour-coded based on posterior probabilities (PP): strongly supported (PP = 1, blue) and non-supported (white). A plus symbol (+) indicates nodes with PP > 0.99. Four different prior models (M1–M4) were tested using two algorithms (A0, A1) and two concatenated datasets (*COI + DOMs*, and *DOMs*). Next to BPP results, nine species' delimitation models (columns) tested by BFD are shown, each including a unique combination of lineages (rows) using *COI + DOMs* and *COI* datasets. Circles in a grey-scale represent Bayes factor (2lnBf) results for path-sampling (PS) and stepping-stone (SS) methods: non-supported (0–2) indicates no significant difference in support values for two models, positively supported (2–10) shows moderate evidence, and decisively supported (>10) indicates strong evidence favouring the best-fitting model over the alternative. The inset map shows the study region with a dashed black line marking the eastern and western candidate species boundaries.

RESULTS

Loci alignments and saturation tests

Alignments' length and number of individuals included in each one are shown in Table 1. All sequences were deposited in GenBank (Supporting Information, Table S1). The mitochondrial gene exhibited less invariable sites compared to the nuclear genes (Table S6). A higher proportion of invariant sites was observed in *DOM6* exons compared to introns (Supporting Information, Table S6). For all molecular markers considering three codon positions collectively and independently, *p*-values in the saturation test were equal to zero. Therefore, no significant saturation was detected, and all codon positions were considered in further analyses.

Phylogenetic inference

The phylogenetic trees inferred for each nuclear marker independently exhibited distinct topologies with low support values (Supporting Information, Figs S1–S3). However, concatenating the three nuclear markers (*DOMs*) yielded higher support values (Supporting Information, Fig. S4). Nonetheless, except for a few monophyletic groups, the ML and BI trees lacked resolution, resulting in a great number of nodes with low bootstrap (<80%) and PP supports (<0.95). Two main clades were identified, divided geographically into eastern and western groups. The *COI*-based phylogeny (Supporting Information, Fig. S5) yielded these same clades with greater support and longer branches. Notably, the phylogenetic relationships within each clade were also better resolved than in the *DOMs* tree. Integrating nuclear information with the mitochondrial locus (*COI* + *DOMs*; Fig. 4) resulted in the eastern and western divisions with higher support values when employing partitions by gene and codon position, rather than using gene partitioning (see Supporting Information, Figs S6–S9 for uncollapsed trees). The resolution within each clade was also better supported when combining mitochondrial and nuclear data. Given the improved phylogenetic resolution obtained with *COI* + *DOMs*, *DOMs*, and *COI* datasets compared to single nuclear locus, we used these three datasets whenever possible in the subsequent analyses.

Monophyletic groups within eastern and western clades

The phylogenetic tree inferred from *COI* + *DOMs* dataset (Fig. 4) revealed that the eastern clade was subdivided into five distinct groups and a singleton (PET-EAST), all showing high support for both ML and BI analyses or at least for one of them. We assigned abbreviations to these monophyletic groups based on the localities from which individuals were collected. Most groups clustered individuals from single localities, except for CAM-SMC-EAST, which included specimens from Sertão do Cambury (CAM) and Caraguatatuba (SMC), and IT-PET, which comprised individuals from Itatiaia (IT) and Petroleiros (PET). Sister to these two, we identified four highly supported monophyletic groups constituted by individuals from four localities: Pedra Branca (PB), Petroleiros (PET), Tijuca (TJ), and Núcleo Picinguaba (SMP). Notably, the PET locality is a particular case, as individuals were distributed across distant phylogenetic groups: one individual (F5234) formed the singleton PET-EAST, another (F5233) clustered in the IT-PET group within the eastern clade, and a third (F5238)

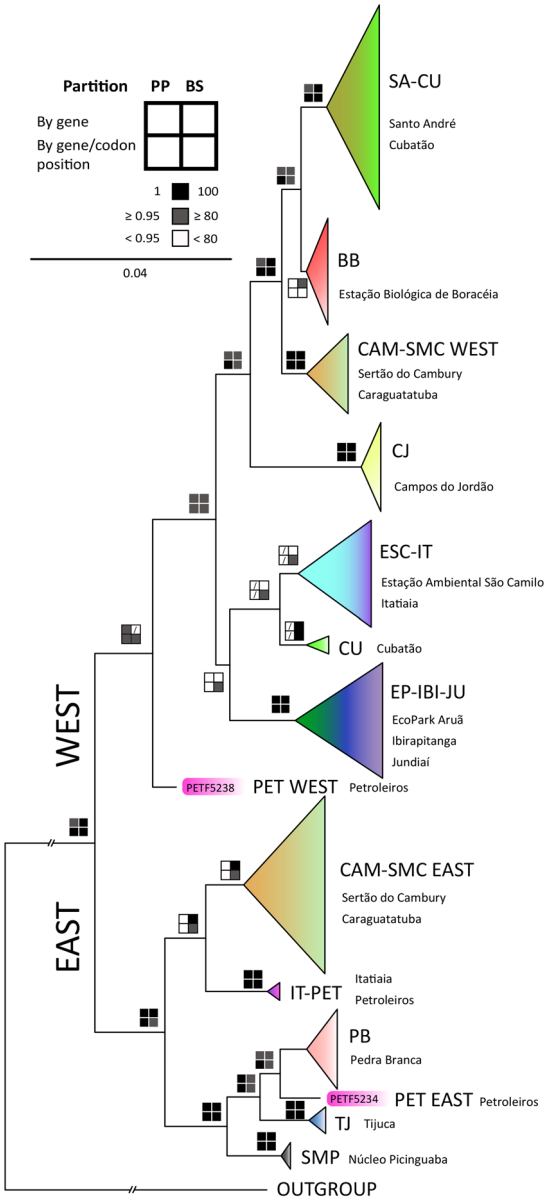

Figure 4. Collapsed phylogenetic tree based on *COI* + *DOMs* dataset, where triangles are proportional to the number of individuals for each monophyletic group. Legend shows support values for each partition scheme (by gene or by gene and the three codon positions) using bootstrap values (BS) and posterior probabilities (PP). Missing groups for certain partitions are indicated with a slash (/). The scale bar represents the number of changes per position. Group names and localities are labelled next to the tree, with abbreviations and colors representing each locality. Strongly supported node values are considered when BS ≥ 80 and PP ≥ 0.95.

constituted the singleton PET-WEST, which is closely related to western clade. Additionally, a specimen from PET (F5239) exhibited a different phylogenetic position in the *DOMs* tree compared to the *COI* tree, and was excluded from analyses when employing *COI* + *DOMs* dataset.

The western clade was subdivided into seven distinct groups (Fig. 4), and a singleton (PET WEST). The poorly supported BB group, which exclusively included individuals from Boraceia, was identified as the sister-group to SA-CU, which comprised

specimens from Santo André and Cubatão. Sister to these two, the CAM-SMC-WEST group included individuals from two localities (Sertão do Cambury and Caraguatatuba), which also constituted a cluster within the eastern clade (CAM-SMC-EAST). Serving as the sister-group to all the previous localities, individuals from Campos do Jordão (CJ) were clearly differentiated in a monophyletic group with robust support. Individuals from Cubatão (CU) clustered into two different groups: SA-CU and CU. However, the monophyly of CU, identified as the sister-group to ESC-IT, was not supported in the BI trees. The ESC-IT group included individuals from Estação Ambiental São Camilo (ESC) and one individual (F5208) from the eastern locality of Itatiaia (IT). Sister to this cluster, EP-IBI-JU comprised individuals from three localities: EcoPark Aruã (EP), Ibirapitanga (IBI), and Jundiaí (JU). Overall, only two individuals from the eastern localities, F5208 from Itatiaia and F5238 from Petroleiros, clustered in the western clade rather than in their respective geographic groups.

Genetic diversity analyses

Intrapopulation diversity, monophyletic groups' diversity, and neutrality tests

The mean haplotype diversity across localities was higher for *COI* (0.8239) compared to the nuclear markers, with the lowest diversity observed for *DOM6* (0.4281) (Supporting Information, Table S7). The CJ locality exhibited the lowest haplotype diversity values across all markers (mean: 0.1808). The highest haplotype diversity values were found in IT and PET localities, with values close or equal to 1 for all markers (mean: 0.9168 for both localities) (Supporting Information, Table S8). Similarly, the mean nucleotide diversity was higher for *COI* (0.0167) than for the nuclear markers, with *DOM6* again exhibiting the lowest nucleotide variability (0.0038) (Supporting Information, Table S7). CJ also had the lowest nucleotide diversity (mean: 0.0006), and the highest was observed in IT and PET localities for all markers (means: 0.0321 and 0.0318, respectively) (Supporting Information, Table S8).

Most populations appeared to evolve under the neutral theory, as most of the observed values for Tajima's *D* neutrality test were not statistically significant. Only six tests exhibited values outside the confidence interval, corresponding to CAM and SA for *COI*, CAM, CJ, and SMC for *DOM5*, as well as ESC for *DOM6*. Tajima's *D* values for these significant cases were negative, except for SMC in *DOM5*. However, it is crucial to interpret these results cautiously, as they were not consistent across all loci. When using *DOMs* dataset, Tajima's *D* remained non-significant for all localities (Supporting Information, Table S8). In agreement with the neutrality test results, the estimated linkage disequilibrium using the Z_{nS} statistic was consistent with neutral evolution.

Population genetic parameters and neutrality tests were also calculated for each monophyletic group observed in the phylogenetic trees, given the clustering of some individuals from specific localities in genetically distant groups. Haplotype and nucleotide diversity values for these groups were generally lower than those estimated for geographical localities. Significant Tajima's *D* values were observed in the SA-CU group for *COI* and *DOM5*, in CAM-SMC-EAST for *DOM4*, and in CJ for *DOM5*. When using *DOMs* dataset, Tajima's *D* was only significant and negative for the CAM-SMC-EAST group (Supporting Information, Table S9). Furthermore, haplotype and nucleotide diversities were also higher for *COI* compared to the nuclear markers (Supporting Information, Table S10).

Interpopulation divergences and isolation by distance test

F_{ST} values for the 15 localities were mostly significant and high for *COI*, *DOMs* (Supporting Information, Fig. S10), and *COI* + *DOMs* datasets (Supporting Information, Fig. S11). Pairwise F_{ST} values for *COI* were generally higher, averaging around 0.8, compared to the approximately 0.7 for the nuclear markers. SMC and CAM localities, which are known to exhibit individuals in the eastern and western clades, also showed relatively high F_{ST} values (around 0.4 for *COI*). When calculating distance matrices based on the monophyletic clusters identified in the trees using the *COI* and *DOMs* datasets, the genetic distances were quite large (Supporting Information, Fig. S12), which is a clear indicator of substantial genetic differentiation among the monophyletic groups. The east–west separation observed in the phylogenetic trees was also evident in the F_{ST} heatmap generated from the *COI* + *DOMs* dataset (Fig. 5A). Within the eastern clade, the identification of distinct groups was comparatively less defined, whereas the western clade exhibited a more structured pattern, divided into three main groups: (i) SA-CU + BB + CAM-SMC-WEST, (ii) CJ, and (iii) ESC-IT + CU + EP-IBI-JU.

Mantel's tests (Supporting Information, Fig. S13) revealed a linear correlation between geographic (Supporting Information, Fig. S14) and genetic distances (Fig. 5A) across all localities. However, the R^2 values remained consistently low yet significant for the two concatenated datasets: $R^2 = 0.0102$, Rxy = 0.101, $P = 0.269$ for *COI*, $R^2 = 0.1789$, Rxy = 0.423, $P = 0.001$ for *DOMs*, and $R^2 = 0.2169$, Rxy = 0.466, $P = 0.0003$ for *COI* + *DOMs*. The geographic distances explained the 17% and 21% of the F_{ST} variation, using the *DOMs* and *COI* + *DOMs* datasets, respectively. However, a significant isolation-by-distance pattern was not detected for *COI*.

Principal component and structure analyses

Principal component analysis (PCA) revealed a clear separation between the eastern and western clades (Supporting Information, Fig. S15). Remarkably, CJ locality was differentiated from other western groups, and CAM and SMC localities were clustered within both main clades, consistent with the phylogenetic trees. When performing a PCA for each clade, higher population structure was observed in *COI* (Supporting Information, Fig. S16) compared to *DOMs* (Supporting Information, Fig. S17). The PCA clusters generated using *COI* + *DOMs* dataset (Fig. 5B) closely aligned with those derived from the F_{ST} heatmap. Specifically, the western clade exhibited the same division [(i) SA-CU + BB + CAM-SMC-WEST, (ii) CJ, and (iii) ESC-IT + CU + EP-IBI-JU] along PC1 (30.17% of the variance) and PC2 (21.15% of the variance). In contrast to the F_{ST} heatmap, which lacked clarity in defining groups, the eastern clade was clearly divided into four clusters [(i) CAM-SMC-EAST, (ii) IT-PET, (iii) PB + PET-EAST + TJ, and (iv) SMP] along PC1 (44.28% of the variance) and PC2 (11.79% of the variance).

The results of PCA analyses were congruent with those obtained from structure analyses. The optimal number of clusters

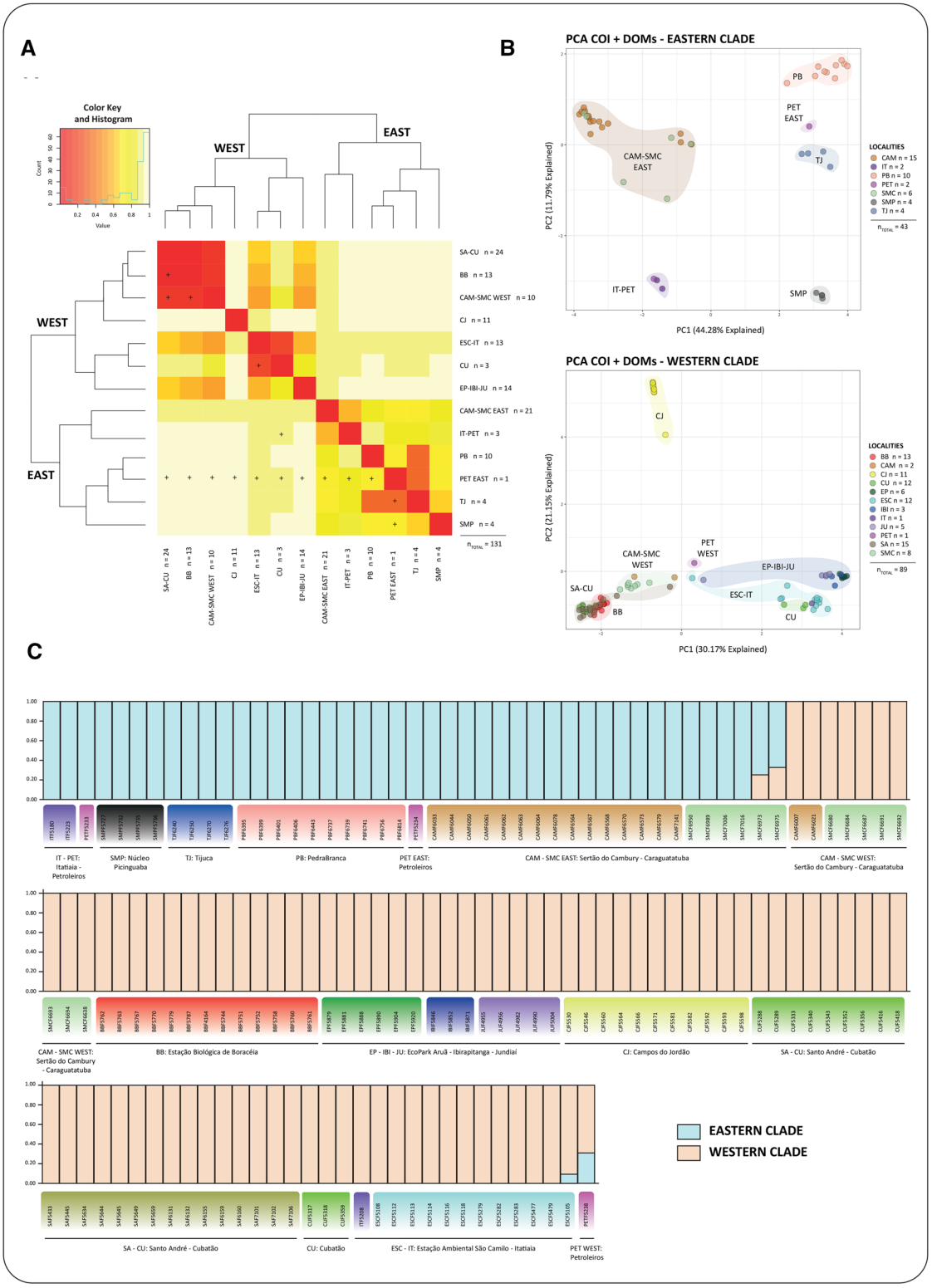

Figure 5. A, F_{ST} heatmap for *COI + DOMs* dataset. The singleton PET-WEST was not included as it represents only one individual, whose phylogenetic position is uncertain. The colour gradient indicates genetic distance, where 1 denotes high differentiation and 0 no differentiation. No statistical significance (p-value > 0.05) is represented by a plus symbol (+). Next to the heatmap, the number of individuals (N) of each monophyletic group is indicated. B, principal component analysis (PCA) plot based on *COI + DOMs* dataset considering PC1 and PC2 for the eastern and western clades separately. The monophyletic groups within each clade are highlighted. Locality colours, abbreviations, and number of individuals (N) are specified in the legend. C, population structure analysis based on *COI + DOMs* dataset when K = 2, corresponding to an east–west separation (coloured in light blue and light orange, respectively).

reached its maximum at K = 2, and had a secondary peak at K = 6 (Supporting Information, Fig. S18). At K = 2, the eastern and western clusters were identified (Fig. 5C). Notably, four individuals exhibited admixed ancestry: two SMC individuals from the eastern clade that shared ancestry with the western group, and two individuals from the western clade, one from ESC and one from PET, that shared little ancestry with the eastern clade. When only considering the eastern individuals, STRUCTURE identified K = 2 as the optimum number of clusters (Supporting Information, Fig. S19), with a minor peak at K = 5. The K = 2 scenario consisted of two monophyletic groups observed in the phylogenetic trees: (i) CAM-SMC-EAST + IT-PET and (ii) PB + PET EAST + TJ + SMP. In contrast, only two optimal groups were delimited within the western clade [(i) SA-CU + BB + CAM-SMC-WEST and (ii) CJ + ESC-IT + CU + EP-IBI-JU), which did not consider CJ as an independent group, as observed in other analyses (Supporting Information, Fig. S20).

Cloning, haplotype networks and intra-individual diversity

A total of 14 individuals of *Cephaloflexa bergi* were successfully cloned for the *DOM4* locus, resulting in 97 sequences (Supporting Information, Table S11). The haplotype networks generated from the *DOM4*-cloned dataset provided evidence of high heterozygosity within individuals (Fig. 6), with most of them showing a high number of differentiated haplotypes. Notably, when analysing two specimens from the same locality, individuals did not share haplotypes, except for the BB and CJ localities, indicating strong intrapopulation variability. When considering all individuals across populations, the haplotype networks exhibited a topology consistent with the phylogenetic trees (Supporting Information, Fig. S21). At the population level, CJ locality exhibited the lowest nucleotide and haplotype diversity, whereas CAM and SMC exhibited the highest values. The same diversity pattern was observed at intra-individual level (Supporting Information, Table S12).

Molecular delimitation of candidate species

The application of two discovery delimitation methods recovered two different species' delimitation scenarios (Fig. 3). The mPTP approach identified 11 candidate species. Within the eastern clade, seven species were recognized: (i) CAM-SMC-EAST, (ii) SMC, (iii) IT-PET, (iv) PB, (v) PET-EAST, (vi) TJ, and (vii) SMP. Six of these candidate species also formed monophyletic groups in the phylogenetic analysis (Fig. 4), except for 'SMC', which constituted a monophyletic group with only two individuals in the *COI*-based tree (Supporting Information, Fig. S5). The western clade exhibited a division into four candidate species: (i) CJ, (ii) SA-CU + BB + CAM-SMC-WEST, (iii) ESC-IT + CU, and (iv) EP-IBI-JU. Before running the GMYC method, we assessed the most suitable priors, which were the uncorrelated log-normal relaxed clock (Drummond *et al.* 2006), and a birth–death speciation tree (Gernhard 2008) (Supporting Information, Table S13). GMYC subsequently delimited 14 candidate species (confidence interval: 12–20). The maximum log-likelihood (1091.628) was better than the null scenario (1084.605), and the likelihood ratio test was significant (p-value = 0.0008). Within the eastern clade, seven candidate species were identified, which was consistent with the hypothesis proposed by mPTP. In contrast, the western clade was divided into the same seven monophyletic groups observed in the phylogenetic analysis (Fig. 4): (i) CJ, (ii) SA-CU, (iii) BB, (iv) CAM-SMC-WEST, (v) ESC-IT, (vi) CU, and (vii) EP-IBI-JU.

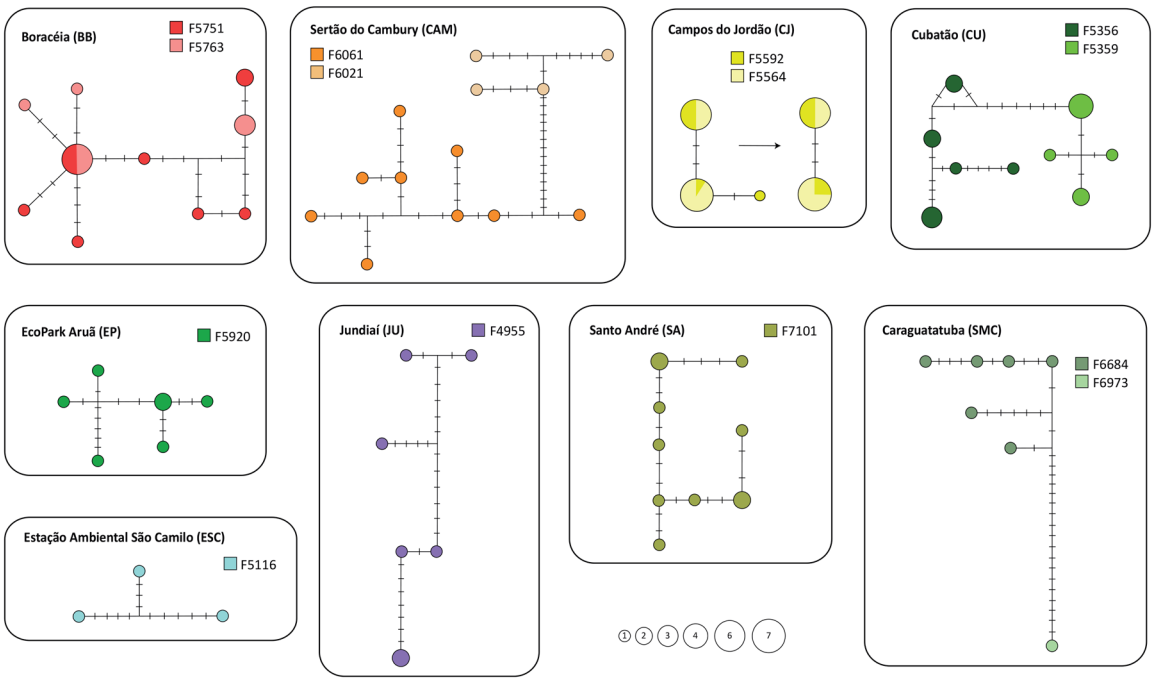

Figure 6. Haplotype networks based on *DOM4*-cloned dataset for each locality. Each circle represents a different haplotype and black lines represent mutations between haplotypes. The size of the circle indicates the frequency of each haplotype within the individual, represented with a distinct colour. CJ haplotypes could not be recoded, due to the limitations of the software, which is unable to construct a network with only two haplotypes.

When testing BPP's model A10, which essentially splits or not the candidate species provided as the starting hypothesis, we evaluated the 14 candidate species derived from the GMYC results, which included the 11 species identified by mPTP. Among the models tested in BPP, the third model, which assumed large ancestral population sizes and shallow divergences among species, was the only scenario that exhibited non-robust posterior probabilities for several candidate species. However, most models supported the presence of 14 candidate species as proposed by GMYC (Fig. 3).

BFD analyses revealed discrepancies depending on the method (PS: path-sampling or SS: stepping-stone) and dataset (*COI* + *DOMs* or *COI*) employed (Fig. 3). For instance, delimitation analysis using *COI* + *DOMs* dataset with the PS method provided the strongest support for model 5, which identified two candidate species within the western group as observed in STRUCTURE analyses [(i) CJ + SA-CU + BB + CAM-SMC-WEST and (ii) ESC-IT + CU + EP-IBI-JU], and supported the eastern group as a single candidate species (Fig. 3; Supporting Information, Table S14). However, pairwise comparisons proposed by Kass and Raftery (1995) revealed that this 'best-fitting scenario' was not significantly better than model 8, which considered four candidate species within the western clade [(i) CJ, 2: SA-CU + BB, (iii) CAM-SMC-WEST, and (iv) ESC-IT + CU + EP-IBI-JU], and identified the eastern group as a single candidate species. When considering the two clades separately, both methods consistently divided the western group into two candidate species: (i) SA-CU + BB + CAM-SMC WEST and (ii) ESC-IT + CU + EP-IBI-JU + CJ. However, no models including the eastern cluster received significantly better support than the alternative models (Supporting Information, Table S15, Fig. S22). Among all datasets and methods, *COI* + *DOMs* with PS appeared to be the most resolutive. In this case, only one alternative model showed no significant difference compared to the best model, whereas in all other cases, two or even four models compete with the optimal one.

DISCUSSION

The mosaic Meselson effect in *Cephaloflexa bergi*

We detected for the first time in land planarians the presence of the mosaic Meselson effect, as described in Leria *et al.* (2019) for freshwater planarians, where individuals carry multiple highly divergent haplotypes at similar frequencies. This genetic pattern arises from repeated fission events followed by regeneration, allowing the accumulation of distinct cell lineages within an individual. Accordingly, the presence of the mosaic pattern in most of our studied populations (Fig. 6), except for CJ specimens, may indicate the presence of fissiparous or facultative reproduction, although we currently lack direct observations of self-fission in *C. bergi*. Interestingly, *C. bergi* has been observed several times predating on harvestmen (Opiliones: Arachnida) in natural environments (Cseh *et al.* 2017, Cuevas-Caballé, Riutort and Álvarez-Presas 2019), and laboratory experiments have also shown that harvestmen can fragment flatworms with its legs IV into two pieces (Silva *et al.* 2018), potentially followed by body regeneration from each piece. However, the frequency of such predation events in the wild are unclear, and it remains uncertain whether they significantly contribute to the genetic pattern observed. Therefore, the underlying biological processes responsible for the high intra-individual haplotype divergence remain to be clarified. Nevertheless, this high genetic variability, both among populations and within individuals, is particularly relevant for species inhabiting highly fragmented ecosystems like the Atlantic Forest, where small, isolated populations are more vulnerable to genetic erosion.

East meets west in multi-locus analyses

Our phylogenetic analyses showed a marked east–west division, grouping localities from the eastern and western PAs in two distinct phylogenetic clades. East and west groups were also observed in the haplotype networks, PCA plots, F_{ST} matrices, and STRUCTURE analyses. This clear division is defined by an imaginary barrier (Fig. 3) that also fits the separation observed in Álvarez-Presas *et al.* (2014) between two sister-clades: the central SAF (C-SAF) and the northern SAF (N-SAF), which correspond, respectively, to our western and eastern clades. The barrier is situated in a region that is currently mostly inhospitable for terrestrial planarians due to human impact. However, the high genetic differentiation found between both groups in our study suggests that their divergence is unlikely to be a result of recent anthropogenic activity. In Álvarez-Presas *et al.* (2014), the clades coalesced back to 7 Mya (HDP: 5.8–8.5 Mya). This dating agrees with Simpson's hypothesis (Simpson 1979), which suggests that initial speciation events in the BAF may have been promoted by geological processes during the Neogene (23–2.6 Mya). Specifically, the uplift of the Brazilian east coast during this period would have produced geographic, hydrographic, and climatic modifications, which might have caused forest fragmentation with divergence in the associated fauna (Grazziotin *et al.* 2006). The Tietê and Paraíba do Sul rivers (Fig. 2) once formed a single drainage system, flowing towards the Atlantic Ocean until the Early Miocene. Tectonic activity that gave rise to the formation of the Serra do Mar resulted in the division of these rivers (Ab'Saber 1957). During this epoch, the uplift of this mountain range may have triggered cooler, rainier, and more humid climate conditions, which could have favoured speciation by promoting the colonization of new niches.

However, a few localities present individuals belonging to the western and eastern clades mixed. This is the case of IT and PET situated in the east and of CAM and SMC geographically located near each other in a continuous, forested coastal area (Fig. 2). As may be expected, IT and PET constitute a monophyletic group in the eastern clade (only three individuals), but each harbour one individual that is grouped within or close to western groups. The low number of specimens collected in those localities does not allow further analyses to be performed that would make it possible to draw solid conclusions on the origin of the western individuals. In contrast, the situation found in CAM and SMC localities, was already observed in a previous sampling in Cambury analysed in Álvarez-Presas *et al.* (2014), where ABC modelling supported the hypothesis of a secondary contact to explain the observed differences among individuals within the locality. Our present analyses of genetic diversity within these two localities revealed a high number of haplotypes (h) and nucleotide diversity (π), resulting

from analysing individuals from both clades together. As expected, analyses conducted separately for each clade revealed significantly lower nucleotide diversity within the CAM-SMC-EAST and CAM-SMC-WEST groups. Although in our present sampling CAM and SMC were considered separated localities, the high genetic similarity among individuals of the same clade, along with the forest continuity and their close proximity, points to them as representing a single locality where the eastern and western clades coexist. Interestingly, STRUCTURE analyses identified two SMC individuals exhibiting co-ancestry from both clades (Fig. 5), suggesting potential introgression between eastern and western lineages in this locality. This east–west coexistence appears to be unique to CAM and SMC, as it is not observed in close inland localities (BB, EP, and IBI), along the southern coast (SA and CU) or northern coast (SMP).

All of these results support the hypothesis of an ecological and geographic barrier that resulted in the formation of the two main clades, as discussed above. However, subsequent gene flow has occurred in a delimited region in the Serra do Mar region, where CAM and SMC localities are situated. Eustatic processes associated with the Pleistocene (2.58–0.0117 Mya), ice mass expansions and retractions (Suguio and Martin 1978), may have led to repeated cycles of isolation and merging of these two pre-existing lineages distributed along the coast. To better understand the history of the two subpopulations, we performed neutrality tests for each monophyletic group (CAM-SMC-EAST and CAM-SMC-WEST). The neutrality test was only significant and negative for the eastern group using the *DOMs* dataset and the *DOM4* locus. This deviation from the neutral evolutionary model could be attributed to a recent selective sweep or population expansion after a recent bottleneck (Tajima 1989). Although Tajima's *D* estimates based on small samples should be approached with caution (Marroni *et al.* 2011), these results could indicate that while the western clade shows signs of stability in this locality, the eastern may have endured some evolutionary process, either a recent expansion or selective sweep.

Genetic structure within *Cephaloflexa bergi*

Within the eastern and western clades, we observed strong genetic structure among populations, particularly in the west, where we identified four distinct groups (Figs 4, 5A). First, localities situated in the Serra do Mar region, such as SA, CU, BB, CAM, and SMC, were clearly differentiated from EP, IBI, and JU populations, which are geographically located above the Tietê River and the metropolitan area of São Paulo (Fig. 2). This subdivision was observed in previous research (Álvarez-Presas *et al.* 2014), in which they found high differentiation between a locality situated in the Serra do Mar and another above the river, dating back to around 4 Mya (HDP: 3–5.1 Mya) during the Pliocene. The third group corresponding to CJ was not included in the previous study. The delimitation of these three western groups can be explained by the existence of putative barriers. The Tietê River and Serra do Mar probably acted as natural barriers to dispersal, limiting gene flow between populations on opposite sides of the river and mountain. In contrast, the recent formation of the megapolis, dating back to the arrival of colonizers less than 500 years ago, is unlikely to be a significant factor responsible for the observed diversification. However, the human-fragmented landscape may accentuate the lack of connectivity between populations nowadays, leading to isolation, decline of mixed populations, and reduced gene flow in the future.

The fourth group included two populations from the Serra do Mar region: ESC and some individuals from CU. Although these populations were more closely related to the EP-IBI-JU group, located above the Tietê River, than to other Serra do Mar localities, the support for this grouping is weak, leaving their relationships unresolved. Additionally, it is also possible that they are more closely related to southern localities in Paraná, Santa Catarina, and Rio Grande do Sul, which previous research (Álvarez-Presas *et al.* 2014) identified as a distinct monophyletic group. Interestingly, individuals from CU clustered within or near two phylogenetic groups that included localities situated geographically at each side of CU: ESC to the west (CU group) and SA to the east (SA-CU group) (Supporting Information, Figs S1–S9). These findings suggest that CU, situated in the coastal region, represents a mixed zone where secondary contacts between diverged clades may have also occurred. In this area, the Serra do Mar is narrower but remains connected by suitable habitat, facilitating potential gene flow between adjacent populations.

The genetic structuring and the relatively high diversity observed within the three western groups suggest again that the western region has remained relatively stable over a long period. In contrast, Campos do Jordão (CJ) exhibited the lowest nucleotide and haplotype diversity for all markers, with no sign of recent expansion or selective pressures that might have influenced its current diversity. Low intrapopulation variability could, then, be attributed to genetic drift or a historically small effective population size. Significant genetic differentiation from the rest of the PAs could be associated with the topographic and physiognomic heterogeneity of the forest. BAF contains humid areas that concentrate on slopes or medium altitudes of mountains, as well as valleys and plains dominated by marine-influenced vegetation (Lago-Barcia *et al.* 2020). In particular, CJ is situated in Serra da Mantiqueira, which receives a significant amount of rainfall due to its geographical position (Fig. 2). This mountain range, which originated in parallel to the formation of the Paraíba do Sul River Valley lying directly beneath it, may have acted as a geographical and ecological barrier to gene flow between populations during the Upper Cretaceous (de Almeida and Carneiro 1998), contributing to the isolation of CJ.

To split or not to split: is *Cephaloflexa bergi* a species complex?

Recent taxonomic revisions have demonstrated that some terrestrial flatworms exhibit more limited distributions than previously assumed, due to the discovery of cryptic species under the same name (Álvarez-Presas *et al.* 2015, Carbayo, Francoy and Giribet 2016, do Amaral *et al.* 2018, Carbayo *et al.* 2018, Negrete *et al.* 2020, Lago-Barcia *et al.* 2023). Classically, the delimitation of land flatworm species has relied on anatomical features, including the morphology of the pharynx and cephalic region, the organization of the parenchymal and cutaneous musculature, and, especially, the copulatory apparatus. In certain genera, it is difficult to ascertain whether the observed morphological polymorphisms respond to intraspecies

variation or species-specific traits. In these cases, an integrative approach combining molecular and morphological evidence has been key to delimit new species within the Geoplanidae family (Almeida *et al.* 2019, Lemos *et al.* 2014, Álvarez-Presas *et al.* 2022).

In the case of *C. bergi*, the low variability in body colour among individuals (Fig. 1) from the eastern and western clades was previously interpreted as intraspecific variation (Álvarez-Presas *et al.* 2014). In the present study, the general shape of the copulatory apparatus of the specimens studied conforms to that of *C. bergi*. Nevertheless, our study revealed a marked genetic structure pointing to *C. bergi* as a complex of several species. The F_{ST} values were high in nearly all comparisons, suggesting that speciation events may have already isolated most populations or are currently in progress. Nonetheless, we are not describing any species in the present work. Since speciation is a continuous process (de Queiroz 1998), the use of a singular source of evidence, such as genetic data in our case, can entail a degree of incongruence. It is crucial now to combine diverse lines of evidence to test the validity of the candidate species delimited genetically. Complementary information such as other molecular markers, morphology, karyology, and ecological aspects are needed to shed light on the actual number of species and proceed to their description. The most conservative species' hypothesis to be tested involves the presence of two candidate species within *C. bergi*, corresponding to the eastern and western clades. However, considering the results from BFD, a three-species' or even a five-species' scheme should be considered. In any case, our results clearly showed that *C. bergi* comprises multiple independent evolutionary lineages credited to be upgraded to the species level, something that would have to be considered when examining the conservation of the protected areas in the Brazilian Atlantic Forest.

ACKNOWLEDGEMENTS

We thank Laia Leria for her intellectual contribution and guidance through the laboratory experiments and preliminary analyses, Daniel Dols for helpful advice with population genetic analyses and species' delimitation methods, and Gema Blasco for technical support.

CREDIT STATEMENT

Marina Lorenzo sequenced the data used in the present work, computed the analyses, created the figures and tables, wrote the original draft, and edited the final manuscript. Ricard Sabaté sequenced some preliminary data used in the present work. Marta Álvarez-Presas revised and edited the final manuscript, and contributed to the original design of the project. Fernando Carbayo sampled the fifteen localities, contributed to the original design of the project, and revised and edited the final manuscript. Marta Riutort designed the investigation framework, supervised all the analyses, and provided intellectual contribution to the investigation, grant funding acquisition, and reviewed and edited the final manuscript.

SUPPLEMENTARY DATA

Supplementary data are available at *Zoological Journal of the Linnean Society* online.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

FUNDING

This research was supported by Fundación BBVA grant BIOCON 06—112 to MR and Ministerio de Economía y Competitividad (PGC2018-093924-B-100) of Spain, and São Paulo Research Foundation (#2014/13661-8, #2016/18295-5, # 2017/05760-4, #2019/12357-7). Collecting permits were provided by INEA-RJ (#60/2014), Santo André Municipality, (dated 22 February 2013), Jundiaí municipality (dated 8 August 2011), Instituto de Botânica (#3/2014; #15/2015), COTEC-SP (#12.640/2011; #95/2015, #D203/2014 BA), MZUSP-SP (#20/2013), ICM-Bio (# 11473, # 42212, # 32779). Fernando Carbayo research group is thanked for sampling support. Marta Álvarez-Presas acknowledges the support of the PDJ-2014 grant from Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR).

DATA AVAILABILITY

Sequences for *COI*, *DOM4*, *DOM5*, *DOM6*, and the cloned individuals for the *DOM4* locus are accessible on GenBank. Details regarding sampling localities, specimen vouchers, sample codes, and GenBank accession numbers are available in Supporting Information, Table S1.

REFERENCES

Ab'Saber AN. O problema das conexões antigas e da separação da drenagem do Paraíba e do Tietê. *Boletim Paulista de Geografia* 1957;**26**:38–49.

Almeida AL, Marques FPL, Carbayo F. 'Endless forms most beautiful': taxonomic revision of the planarian *Geoplana vaginuloides* (Darwin, 1844) and discovery of numerous congeners (Platyhelminthes: Tricladida). *Zoological Journal of the Linnean Society* 2019;**185**:1–65. https://doi.org/10.1093/zoolinnean/zly022

de Almeida FFM, Carneiro CDR. Origem e evolução da Serra do Mar. *Revista Brasileira de Geociências* 1998;**28**:135–50. https://doi.org/10.25249/0375-7536.1998135150

Álvarez-Presas M, Carbayo F, Rozas J *et al.* Land planarians (Platyhelminthes) as a model organism for fine-scale phylogeographic studies: understanding patterns of biodiversity in the Brazilian Atlantic Forest hotspot. *Journal of Evolutionary Biology* 2011;**24**:887–96. https://doi.org/10.1111/j.1420-9101.2010.02220.x

Álvarez-Presas M, Sánchez-Gracia A, Carbayo F *et al.* Insights into the origin and distribution of biodiversity in the Brazilian Atlantic forest hot spot: a statistical phylogeographic study using a low-dispersal organism. *Heredity* 2014;**112**:656–65. https://doi.org/10.1038/hdy.2014.3

Álvarez-Presas M, Amaral SV, Carbayo F *et al.* Focus on the details: morphological evidence supports new cryptic land flatworm (Platyhelminthes) species revealed with molecules. *Organisms Diversity & Evolution* 2015;**15**:379–403. https://doi.org/10.1007/s13127-014-0197-z

Álvarez-Presas M, Mateos E, Sluys R *et al.* Cryptic diversity in European terrestrial flatworms of the genus Microplana (Platyhelminthes, Tricladida, Geoplanidae). In: *Cryptic Species: Morphological Stasis, Circumscription, and Hidden Diversity*. Cambridge: Cambridge University Press, 2022, 281–93.

do Amaral SV, Ribeiro GG, Valiati VH *et al.* Body doubles: an integrative taxonomic approach reveals new sibling species of land planarians. *Invertebrate Systematics* 2018;**32**:533–50. https://doi.org/10.1071/IS17046

Baele G, Lemey P, Bedford T *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 2012;**29**:2157–67. https://doi.org/10.1093/molbev/mss084

Carbayo F, Almeida AL. Anatomical deviation of male organs of land planarians from Rio de Janeiro, Brazil, with description of two new species of Cratera (Platyhelminthes, Tricladida). *Zootaxa* 2015; **3931**:27–40. http://dx.doi.org/10.11646/zootaxa.3931.1.2

Carbayo F, Leal-Zanchet AM, Vieira EM. Terrestrial flatworm (Platyhelminthes: Tricladida: Terricola) diversity versus man-induced disturbance in an ombrophilous Forest in Southern Brazil. *Biodiversity & Conservation* 2002;**11**:1091–104. https://doi.org/10.1023/A:1015865005604

Carbayo F, Francoy TM, Giribet G. Non-destructive imaging to describe a new species of Obama land planarian (Platyhelminthes, Tricladida). *Zoologica Scripta* 2016;**45**:566–78. https://doi.org/10.1111/zsc.12175

Carbayo F, Silva MS, Riutort M *et al.* Rolling into the deep of the land planarian genus *Choeradoplana* (Tricladida, Continenticola, Geoplanidae) taxonomy. *Organisms Diversity & Evolution* 2018;**18**:187–210. https://doi.org/10.1007/s13127-017-0352-4

Cseh A, Carbayo F, Froehlich EM. Observations on food preference of neotropical land planarians (Platyhelminthes), with emphasis on *Obama anthropophila*, and their phylogenetic diversification. *Zoologia (Curitiba)* 2017;**34**:e12622–8. https://doi.org/10.3897/zoologia.34.e12622

Carnaval AC, Moritz C, Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* 2008;**35**:1187–201. https://doi.org/10.1111/j.1365-2699.2007.01870.x

Cuevas-Caballé C, Riutort M, Álvarez-Presas M. Diet assessment of two land planarian species using high-throughput sequencing data. *Scientific Reports* 2019;**9**:8679. https://doi.org/10.1038/s41598-019-44952-3

Da Silva RFB, Millington JDA, Moran EF *et al.* Three decades of land-use and land-cover change in mountain regions of the Brazilian Atlantic Forest. *Landscape and Urban Planning* 2020;**204**:103948. https://doi.org/10.1016/j.landurbplan.2020.103948

Drummond AJ, Ho SYW, Phillips MJ *et al.* Relaxed phylogenetics and dating with confidence. *PLoS Biology* 2006;**4**:e88. https://doi.org/10.1371/journal.pbio.0040088

Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 2012;**4**:359–61. https://doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 2005;**14**:2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 2005;**1**:117693430500100003. https://doi.org/10.1177/117693430500100003

Ezard T, Fujisawa T, Barraclough TG. Splits: species' limits by threshold statistics. R package version 1.0-14/r31, 2009.

Flouri T, Jiao X, Rannala B *et al.* Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution* 2018;**35**:2585–93. https://doi.org/10.1093/molbev/msy147

Frías-López C, Sánchez-Herrero JF, Guirao-Rico S *et al.* DOMINO: development of informative molecular markers for phylogenetic and genome-wide population genetic studies in non-model organisms. *Bioinformatics (Oxford, England)* 2016;**32**:3753–9. https://doi.org/10.1093/bioinformatics/btw534

Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* 2013;**62**:707–24. https://doi.org/10.1093/sysbio/syt033

Gernhard T. The conditioned reconstructed process. *Journal of Theoretical Biology* 2008;**253**:769–78. https://doi.org/10.1016/j.jtbi.2008.04.005

Grazziotin FG, Monzel M, Echeverrigaray S *et al.* Phylogeography of the *Bothrops jararaca* complex (Serpentes: Viperidae): past fragmentation and island colonization in the Brazilian Atlantic Forest. *Molecular Ecology* 2006;**15**:3969–82. https://doi.org/10.1111/j.1365-294X.2006.03057.x

Grummer JA, Bryson , RWJr., Reeder TW. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* 2014;**63**:119–33. https://doi.org/10.1093/sysbio/syt069

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999;**41**:95–8.

Hyman LH. Some land planarians of the United States and Europe, with remarks on nomenclature. *American Museum Novitates* 1954;**1667**:

Jombart T, Collins C. Analysing genome-wide SNP data using adegenet 2.0.0, 2015

Jones H, Mateos E, Riutort M *et al.* The identity of the invasive yellow-striped terrestrial planarian found recently in Europe: *Caenoplana variegata* (Fletcher & Hamilton, 1888) or *Caenoplana bicolor* (Graff, 1899)?*Zootaxa* 2020;**4731**: zootaxa.4731.2.2. https://doi.org/10.11646/zootaxa.4731.2.2

Kapli P, Lutteropp S, Zhang J *et al.* Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics (Oxford, England)* 2017;**33**:1630–8. https://doi.org/10.1093/bioinformatics/btx025

Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association* 1995;**90**:773–95. https://doi.org/10.1080/01621459.1995.10476572

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010

Kelly JK. A test of neutrality based on interlocus associations. *Genetics* 1997;**146**:1197–206. https://doi.org/10.1093/genetics/146.3.1197

Lago-Barcia D, DaSilva MB, Conti LA *et al.* Areas of endemism of land planarians (Platyhelminthes: Tricladida) in the Southern Atlantic Forest. *PLoS One* 2020;**15**:e0235949. https://doi.org/10.1371/journal.pone.0235949

Lago-Barcia D, Álvarez-Presas M, Riutort M *et al.* Phylogenetic relationships of the Geoplaninae land planarians (Platyhelminthes, Tricladida) assessed with a total evidence approach, with the description of a new species of *Gigantea*. *Molecular Phylogenetics and Evolution* 2023;**184**:107750. https://doi.org/10.1016/j.ympev.2023.107750

Lago-Barcia D, Conti LA, Solà E *et al.* Causal factors and time in formation of areas of endemism for land planarians in the Atlantic Forest. *Journal of Biogeography* 2024;**51**:1894–909. https://doi.org/10.1111/jbi.14857

Lemos VS, Cauduro GP, Valiati VH *et al.* Phylogenetic relationships within the flatworm genus *Choeradoplana* Graff (Platyhelminthes: Tricladida) inferred from molecular data with the description of two new sympatric species from Araucaria moist forests. *Invertebrate Systematics* 2014;**28**:605–27. https://doi.org/10.1071/IS14003

Leria L, Vila-Farré M, Solà E *et al.* Outstanding intra-individual genetic diversity in fissiparous planarians (Dugesia, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology* 2019;**19**:130. https://doi.org/10.1186/s12862-019-1440-1

Marroni F, Pinosio S, Zaina G *et al.* Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (*CAD4*) gene. *Tree Genetics & Genomes* 2011;**7**:1011–23. https://doi.org/10.1007/s11295-011-0391-5

Minh BQ, Schmidt HA, Chernomor O *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015

Myers N, Mittermeier RA, Mittermeier CG *et al.* Biodiversity hotspots for conservation priorities. *Nature* 2000;**403**:853–8. https://doi.org/10.1038/35002501

Negrete L, Colpo KD, Brusa F. Land planarian assemblages in protected areas of the interior Atlantic forest: implications for conservation. *PLoS One* 2014;**9**:e90513. https://doi.org/10.1371/journal.pone.0090513

Negrete L, Lenguas Francavilla M, Damborenea C *et al.* Trying to take over the world: potential distribution of *Obama nungara* (Platyhelminthes: Geoplanidae), the neotropical land planarian that has reached Europe. *Global Change Biology* 2020;**26**:4907–18. https://doi.org/10.1111/gcb.15208

Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England)* 2019;**35**:526–8. https://doi.org/10.1093/bioinformatics/bty633

Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 2006;**6**:288–95. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pons J, Barraclough TG, Gomez-Zurita J *et al.* Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* 2006;**55**:595–609. https://doi.org/10.1080/10635150600852011

Posada D. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 2008;**25**:1253–6. https://doi.org/10.1093/molbev/msn083

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;**155**:945–59. https://doi.org/10.1093/genetics/155.2.945

de Queiroz K, The general lineage concept of species, species criteria, and the process of speciation. *Endless Forms: Species and Speciation*, 1998.

Rambaut A, Drummond AJ, Xie D *et al.* Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 2018;**67**:901–4. https://doi.org/10.1093/sysbio/syy032

Rannala B, Yang Z. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* 2013;**194**:245–53. https://doi.org/10.1534/genetics.112.149039

Rezende CL, Scarano FR, Assad ED *et al.* From hotspot to hopespot: an opportunity for the Brazilian Atlantic Forest. *Perspectives in Ecology and Conservation* 2018;**16**:208–14. https://doi.org/10.1016/j.pecon.2018.10.002

Ribeiro MC, Metzger JP, Martensen AC *et al.* The Brazilian Atlantic Forest: How much is left, and how is the remaining Forest distributed? Implications for conservation. *Biological Conservation* 2009;**142**:1141–53. https://doi.org/10.1016/j.biocon.2009.02.021

Ronquist F, Teslenko M, van der Mark P *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 2012;**61**:539–42. https://doi.org/10.1093/sysbio/sys029

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 2017;**34**:3299–302. https://doi.org/10.1093/molbev/msx248

Silva MS, Willemart RH, Carbayo F. Sticky flatworms (Platyhelminthes) kill armored harvestmen (Arachnida, Opiliones) but are not immune to the prey's weapons. *Journal of Zoology* 2018;**306**:88–94. https://doi.org/10.1111/jzo.12570

Simpson B. Quaternary biogeography of the high montane regions of South America. *The South American Herpetofauna: its Origin, Evolution, and Dispersal*, 1979, 157–88.

Sluys R. Global diversity of land planarians (Platyhelminthes, Tricladida, Terricola): a new indicator-taxon in biodiversity and conservation studies. *Biodiversity & Conservation* 1999;**8**:1663–81.

Sobral-Souza T, Vancine MH, Ribeiro MC *et al.* Efficiency of protected areas in Amazon and Atlantic Forest conservation: a spatio-temporal view. *Acta Oecologica* 2018;**87**:1–7. https://doi.org/10.1016/j.actao.2018.01.001

Suchard MA, Lemey P, Baele G *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 2018;**4**:vey016. https://doi.org/10.1093/ve/vey016

Suguio K, Martin L. Formações quaternárias marinhas do litoral paulista e sul fluminense: quaternary marine formations of the state of São Paulo and southern Rio de Janeiro. Special Publication 1978;55.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;**123**:585–95. https://doi.org/10.1093/genetics/123.3.585

Winsor L, Johns PM, Barker GM. Terrestrial planarians (Platyhelminthes: Tricladida: Terricola) predaceous on terrestrial gastropods. *Natural Enemies of Terrestrial Molluscs* 2004;227–78. https://doi.org/10.1079/9780851993195.0227

Wright S. *Evolution and the Genetics of Populations, Volume 4: Variability within and among Natural Populations*. Chicago: University of Chicago Press, 1978.

Xia X. DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* 2018;**35**:1550–2. https://doi.org/10.1093/molbev/msy073

Xia X, Xie Z, Salemi M *et al.* An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* 2003;**26**:1–7. https://doi.org/10.1016/S1055-7903(02)00326-3

Yang Z. The BPP program for species tree estimation and species delimitation. *Current Zoology* 2015;**61**:854–65. https://doi.org/10.1093/czoolo/61.5.854

Yang Z, Rannala B. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 2010;**107**:9264–9. https://doi.org/10.1073/pnas.0913022107