Molecular Phylogenetics and Evolution 130 (2019) 169–180

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phenotypic and life-history diversification in Amazonian frogs despite past introgressions

Antoine Fouquet[a,b,*], Berengère Ferrier[b,c], Jordi Salmona[a], Sourakhata Tirera[d], Jean-Pierre Vacher[a], Elodie A. Courtois[b], Philippe Gaucher[b], Jucivaldo Dias Lima[e], Pedro M. Sales Nunes[f], Sergio Marques de Souza[g], Miguel T. Rodrigues[g], Brice Noonan[h], Benoit de Thoisy[c,d]

[a] Laboratoire Evolution et Diversité Biologique (EDB), UMR5174, Bâtiment 4R1, 118 Route de Narbonne, 31062 Toulouse cedex 9, France
[b] Laboratoire Ecologie, Evolution, Interactions des Systèmes Amazoniens (LEEISA), USR3456 Cayenne, French Guiana
[c] Association Kwata, 16 avenue Pasteur, Cayenne, French Guiana
[d] Institut Pasteur de la Guyane, 23 avenue Pasteur, Cayenne, French Guiana
[e] Instituto de Pesquisas Cientificas e Tecnológicas do Amapá (IEPA) – Nucleo de Biodiversidade (NUBIO), Laboratório de Herpetologia, Rodovia Juscelino Kubitschek, s/n, Distrito da Fazendinha, Macapá, AP, Brazil
[f] Universidade Federal de Pernambuco, Centro de Biocências, Departamento de Zoologia, Av. Professor Moraes Rego, s/n, 50670-901 Recife, PE, Brazil
[g] Universidade de São Paulo, Instituto de Biociências, Departamento de Zoologia, Caixa Postal 11.461, CEP 05508-090 São Paulo, SP, Brazil
[h] University of Mississippi, Department of Biology, Oxford, 507 Shoemaker Hall, University, MS 38677, USA

ARTICLE INFO

Keywords:
Admixture
Anomaloglossus
Amphibian
Guiana Shield
Hybridization
Speciation

ABSTRACT

The advent of genomics in phylogenetics and population genetics strengthened the perception that conflicts among gene trees are frequent and often due to introgression. However, hybridization occurs mostly among species that exhibit little phenotypic differentiation. A recent study delineating species in *Anomaloglossus*, a frog genus endemic to the Guiana Shield, identified an intriguing pattern in the *A. baeobatrachus* species complex. This complex occurs in French Guiana and Amapá (Brazil) and comprises two sympatric phenotypes contrasting not only in body size, habitat, and advertisement call, but also in larval development mode (endotrophic vs exotrophic tadpoles). However, molecular and phenotypic divergences are, in some cases, incongruent, i.e specimens sharing mtDNA haplotypes are phenotypically distinct, suggesting a complex evolutionary history. Therefore, we genotyped 106 *Anomaloglossus* individuals using ddRADseq to test whether this phenotype/genotype incongruence was a product of phenotypic plasticity, incomplete lineage sorting, multiple speciation events, or admixture. Based on more than 16,000 SNPs, phylogenetic and population genetic approaches demonstrated that exotrophic populations are paraphyletic. Species tree and admixture analyses revealed a strikingly reticulate pattern, suggesting multiple historical introgression events. The evolutionary history of one exotrophic population in northern French Guiana is particularly compelling given that it received genetic material from exotrophic ancestors but shows very strong genetic affinity with the nearby endotrophic populations. This suggests strong selection on larval development and mating call after secondary contact and hybridization. The case of *A. baeobatrachus* represents a striking example of introgression among lineages that are phenotypically distinct, even in their larval development mode, and highlights how high-resolution genomic data can unravel unexpectedly complex evolutionary scenarios.

1. Introduction

The plethoric number of studies identifying conflicts among gene trees, thanks to the recent transition from genetic to genomic approaches, strengthened the perception that incomplete lineage sorting and interspecific admixture are more common than initially thought (Smith et al., 2015). Hybridization has long been recognized as an important feature in the evolution of plants, but a growing number of studies suggest that it may be important in animal evolution as well, occasionally leading to the formation of new species (Lamichhaney

* Corresponding author at: Laboratoire Evolution et Diversité Biologique (EDB), UMR5174, Bâtiment 4R1, 118 Route de Narbonne, 31062 Toulouse cedex 9, France.
E-mail address: fouquet.antoine@gmail.com (A. Fouquet).

https://doi.org/10.1016/j.ympev.2018.09.010
Received 7 June 2018; Received in revised form 15 September 2018; Accepted 15 September 2018
Available online 05 October 2018
1055-7903/

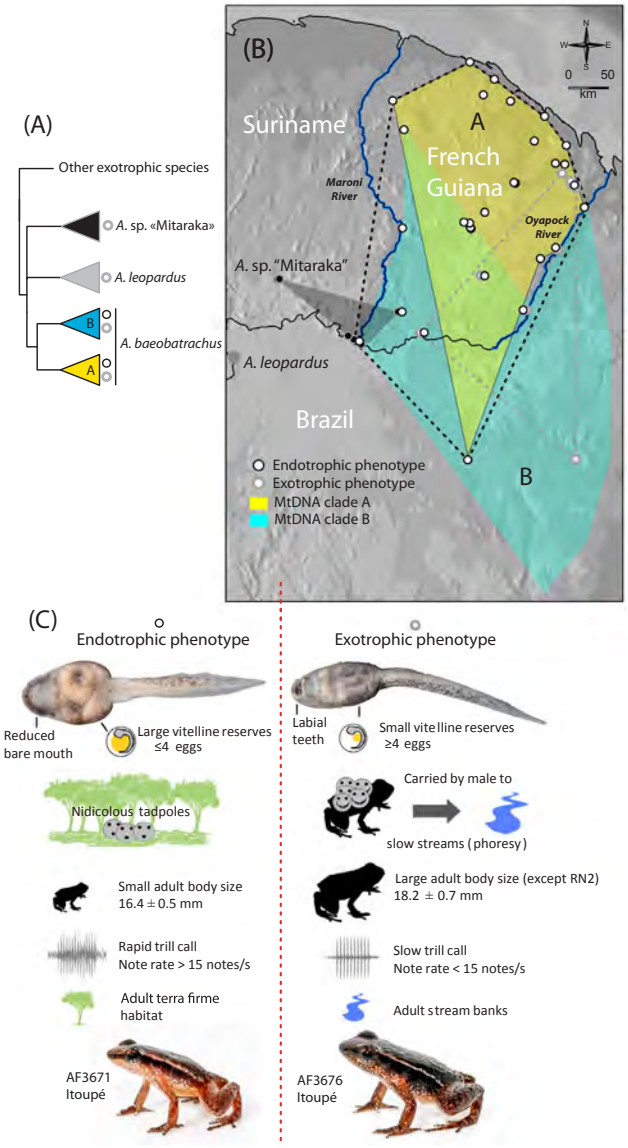

Fig. 1. Phylogenetic relationships and distribution among populations and phenotypes within the *A. baeobatrachus* complex according to Vacher et al. (2017). (A) the two mtDNA lineages and their relationship with the related species: *A.* sp. "Mitaraka" and *A. leopardus*. (B) Map of the distribution of the two mtDNA (filled polygons; note that the two mtDNA lineages face each other along most of the Oyapock River) and of the two phenotypes (endotrophic: black outlined circles; exotrophic: grey outlined circles and dotted lined polygons for their respective ranges). (C) The array of traits varying between the two defined phenotypes.

et al., 2018; Schumer et al., 2014). The emergence of reproductive isolation is often a prolonged process (Avise et al., 1998; Coyne and Orr, 2004), and if incipient species are not completely isolated geographically, hybridization and the introgression of genetic information may occur (Mallet, 2005). Reproductive isolation may then follow if hybrids have access to niches unavailable to parental species (Gross and Rieseberg, 2004) and/or reject parental species as potential mates (Mavarez et al., 2006; Melo et al., 2009). The potential incidence of introgression led to the development of criteria and tools to characterize and to detect these evolutionary events (Excoffier and Foll 2011; Hey 2009; Pickrell and Pritchard 2012; Schumer et al., 2014). Understanding admixture patterns can shed light on the history and dynamics of genetic interactions between incipient species (Duvaux et al., 2011; Neafsey et al., 2010; Teeter et al., 2010) and has potentially important implications in evolutionary biology, biodiversity, and conservation.

Vacher et al. (2017) recently identified an intriguing pattern in frogs of the *Anomaloglossus baeobatrachus* complex (Amphibia: Anura: Aromobatidae) from French Guiana (FG) and Amapá (Brazil; Fig. 1). This species complex harbors two contrasting phenotypes (body size, call, habitat, and larval development mode) with broadly overlapping ranges: one is small bodied (16.4 mm ± 0.5), emits rapid trills (> 15 notes/s), occurs away from water bodies, and produces tadpoles that are nidicolous (no free aquatic larval stage: i.e. tadpoles develop until metamorphosis in a nest in the leaf litter) and endotrophic (developing only upon vitelline reserves); the other is larger in body size (18.2 mm ± 0.7), emits slower trills (< 15 notes/s), and occurs along slow-flowing streams nearby. Males attend nests deposited in the leaf litter and, after hatching, carry the exotrophic tadpoles into the water where they complete their development feeding on aquatic resources (Fig. 1). These species can be found within meters of each other. However, they harbor in most cases distinct mtDNA lineages at a given site (except two populations, cf. results section). Phylogenetic relationships inferred from mtDNA and nuclear data (three exons) showed topological incongruences, suggesting a complex diversification pattern. The existence of contrasted developmental pathways between closely related species is extremely rare, and is seen in amphibians as a fixed trait within species and even a criterion to delineate supraspecific taxonomic groups (Dubois, 2004).

Vacher et al. (2017) proposed four scenarios to explain the origin of this intriguing pattern: (1) phenotypic plasticity; (2) incomplete lineage sorting (ILS) of the mt and nuDNA previously analyzed; (3) phenotypic convergence, or (4) secondary contact with introgression. To distinguish among these scenarios, we generated genomic (mitogenomic and ddRADseq) data from a comprehensive set of well-sampled populations. Using this high-resolution genomic data, we expected to be able to distinguish among these four hypotheses. In the presence of (1) phenotypic plasticity, we expected to recover close genetic relationships among co-occurring phenotypes; in the case of (2) ILS in the previously used mtDNA data, we expected with new genomic data to recover reciprocal monophyly or if one phenotype emerged from a subset of populations, a nested position of one species within the other; with (3) phenotypic convergence we expected two pairs of phenotypically distinct species to form distinct clades, and finally with (4) secondary contact with introgression we expected non-monophyletic species and to detect spatially explicit gene flow across phenotypically distinct populations.

2. Materials and methods

2.1. Sampling

Specimens were collected between 2001 and 2016 in French Guiana, Suriname, and the State of Amapá in Brazil. Specimens were collected by hand during active search of appropriate habitat during the day, their call and habitat recorded when possible. Frogs were euthanized by injection of a solution of lidocaine 2% immediately after being photographed. Liver tissue for genetic work was preserved in 95% ethanol; whole specimens were fixed in 10% formalin for 24 h and then transferred in 70% ethanol for permanent storage. In a few instances, clutches and tadpoles were collected and raised. Some of these individuals were fixed in 10% formalin in order to examine morphology, and others were preserved in ethanol as tissue samples to confirm species identity.

2.2. Phenotype assignment

The direct observation of larval development mode is difficult in the field, and clutches/tadpoles were directly observed or collected only in Mana, Kaw, Saint Georges, and Mitaraka for the endotrophic

populations, and Mitaraka, RN2, and Serra do Navio (Fig. 1) for the exotrophic populations. However, Vacher et al. (2017) established a strong link among various traits suggesting coevolving morphology, habitat, call, and larval development. The first three traits are much easier to record in the field and were used to infer and to segregate phenotypes.

We reexamined the acoustic variation from Vacher et al. (2017), this time focusing on the *Anomaloglossus baeobatrachus* complex and adding 10 recorded specimens (total 51 specimens). The material used for call recording includes Olympus LS11 and Zoom H4N digital recorders, attached to a Sennheiser ME-66 supercardioid microphone powered with a K6P module. We measured six call variables using Audacity v.2.1.1 from the train of pulsed notes emitted by these species. Variables follow those standardized in Köhler et al. (2017): call rate (number of calls divided by their window duration), call duration, note duration, inter-note interval, note repetition rate (note rate: call duration divided by the number of notes in the call), and the dominant frequency. Four calls per record were measured, and the mean values were examined through principal component analysis (PCA). Details of the records examined and data are in Appendix B. PCA were conducted with the package FactoMineR in R v.3.2.4 (R Development Core Team, 2016; Lê et al., 2008). Variables were not corrected by the temperature at the time of the recording because the variation was low (21–25 °C) and because most of the variables analyzed herein are generally little influenced by temperature (Köhler et al., 2017). They were not corrected by body-size either, despite that it is known to influence call dominant frequency (Köhler et al., 2017). Nevertheless, variables other than dominant frequency (not tightly linked with body size) contributed importantly to the first axis, and the RN2 population is clearly distinct from endotrophic populations despite complete overlap in body-size (see results). Therefore, we believe that correcting the acoustic variable was unnecessary to provide the information needed for this work.

We also reexamined the morphological variation from Vacher et al. (2017), focusing on the *Anomaloglossus baeobatrachus* complex and adding 45 more specimens (total 93 specimens). We measured 17 variables: snout-vent length (SVL); head length from corner of mouth to tip of snout (HL); head width at level of angle of jaws (HW); snout length from anterior edge of eye to tip of snout (SL); eye to naris distance from anterior edge of eye to centre of naris (EN); internarial distance (IN); horizontal eye diameter (ED); interorbital distance (IO); diameter of tympanum (TYM); forearm length from proximal edge of palmar tubercle to outer edge of flexed elbow (FAL); hand length from proximal edge of palmar tubercle to tip of finger (HAND); width of disc on Finger III (WFD); tibia length from outer edge of flexed knee to heel (TL); foot length from proximal edge of inner metatarsal tubercle to tip of toe IV (FL); width of disc on Toe IV (WTD); thigh length from vent opening to flexed knee (ThL); length of Finger I from inner edge of thenar tubercle to tip of disc (1FiL) following Fouquet et al. (2015). Specimens examined and data are listed in Appendix C. All measurements were taken by the same person (AF) using a digital calliper to the 0.1 mm. Specimens were categorized according to the main genetic groups (for endotrophic phenotypes: South, North and Amapá, for exotrophic phenotypes: *Anomaloglossus* sp. "Mitaraka", *A. leopardus*, RN2 and southern populations independently; see results). Only 10 of these measured specimens were not genotyped and were left unassigned. We examined morphometric data through principal component analysis (PCA) via FactoMineR in R v.3.2.4 (R Development Core Team 2016; Lê et al., 2008). To control for variation in body-size among individuals, we additionally performed analyses on a size-corrected dataset obtained by using the residuals of the linear regression between the original morphometric measures of each variable with SVL (Strauss, 1985).

2.3. MtDNA

The phylogenetic relationships among delineated species of the *Anomaloglossus stepheni* species group (which includes the *A. baeobatrachus* complex) remained poorly supported in Vacher et al. (2017) due to the small size of the DNA locus used (396–398 bp; 413 bp once aligned). Therefore, we selected one representative of each major mtDNA lineage of the *A. stepheni* species group (except the early diverging *A. stepheni* and *A. apiau*) to generate additional mitogenomic data and to improve our understanding of their relationships. Mitogenomes were sequenced and assembled following the method described in Vacher et al. (2016). We used 200 ng of DNA to build a DNA sequencing library that was hybridized and sequenced on a 1/24th of lane of an Illumina HiSeq 2500 flow cell (Illumina Inc., San Diego, CA). The mitochondrial genomes were assembled using the organelle assembler Org.Asm (Coissac, 2016). They were then annotated with MITOS (Bernt et al., 2013) and completed when necessary with an iterative mapping strategy implemented in Geneious v.9 (Kearse et al., 2012). All 14 new mitogenomes were deposited in GenBank (Appendix A). We then used MAFFT v.7 (Katoh et al., 2013) with default parameters to align the 14 mitogenomes together with 258 additional 16S DNA sequences from Vacher et al. (2017) (Appendix A) of the *A. stepheni* species group, excluding the control region, resulting in an alignment of 15,380 bp. We conducted a Maximum Likelihood phylogenetic analysis in RAXML (Stamatakis, 2014) using a single partition, the GTR + I + G model and 1,000 bootstrap replicates. The tree was rooted with the clade formed by *A.* sp. "Acari" + *A.* sp. "Brownsberg" + *A.* sp. "Paru" + *A.* sp. "Bakhuis" according to Vacher et al. (2017).

2.4. RADseq

2.4.1. Data acquisition

We selected a total of 125 individuals from 22 *Anomaloglossus* populations (Appendix A) harboring distinct phenotypes and one or two mtDNA lineages (N > 4 ind/pop/species with few exceptions); only the RN2 and Mémora Right/Lourenço localities harbored populations with distinct phenotypes sharing the same haplogroup. The closely related species *A.* sp. "Mitaraka", *A. leopardus,* and *A.* sp. "Brownsberg" were included to serve as outgroups (Vacher et al., 2017). Genomic DNA was extracted using the Wizard Genomic extraction kit (Promega©; Madison, WI, USA) and quantified using Qubit fluorometry (Thermo Fisher©; Waltham, MA, USA). Reduced representation RADseq libraries were constructed from 50 ng of DNA at 20 ng/μL following a 3RAD approach (Graham et al., 2015). Genomic DNA was digested with three enzymes. While EcoR1 and Xba1 create the 'sticky' ends to which adapters are ligated, NheI was included to cut adapter dimmers, improving ligation efficiency. After this triple digestion, double-stranded adapters were ligated to add Illumina priming sites and unique combinations of barcodes for each sample. Ligation products were amplified by PCR (14 cycles), and resulting amplicon pools were size selected (350–400 bp) using a PippinPrep (Sage Science; Beverly, MA, USA). Libraries were then quantified via qPCR using a KAPA Illumina Library Quantification Kit (KAPA Biosystems, USA). Individual libraries were then pooled into a single library containing equal amounts of each and denatured and diluted to 1.8 pM as recommended in the Illumina NextSeq documentation. The prepared library was spiked with 15% PhiX, to introduce necessary nucleotide diversity in early cycles, and sequenced on a 75-cycle Illumina NextSeq 550 High-Output flowcell.

2.4.2. Quality filtering, de novo locus assembly, and SNP calling

The resulting raw data consisted of 163 million 76 bp reads, which were demultiplexed using bcl2fastq v. 1.8.4 (Illumina) allowing for two errors in the barcode sequence (barcodes of Graham et al., 2015 have 3 bp degeneracy), effectively removing reads that lacked identifiable barcode pairs. An average of 1.3 M reads per individual (N = 125) were

then trimmed to 62 bp in length using the FASTX-Toolkit (Gordon and Hannon, 2010) to remove adapter sequences. Read quality, assessed with FastQC, was > 32 in average for all reads and per sequence quality averaged > 30. Clustering reads of the same locus within individuals was performed with Ustacks using a maximum distance (M) of 2 and a minimum coverage (m) of 3, followed by alignment of loci among individuals with Cstacks to create population consensus loci catalog, allowing 2 mismatches (n) and a maximum of 2 bp gaps (Catchen et al., 2011, 2013). To limit the impact of missing data, we retained 2427 loci shared by > 100 individuals (80%) with only 2 possible loci per individual. The *sstacks* module was used to map back individual loci to the catalog with default settings. We used the STACKS *populations* module (Catchen et al., 2011, 2013) to call SNPs considering: (1) a minimum number of populations a locus must be present in (p) = 3; (2) a minimum percentage of individuals in a population required to retain a locus for that population (r) = 50%; (3) minimum stack depth required for individuals at a locus (m) = 3. Under these parameters, we retained 2300 loci with 16,197 variant sites (149,978 bp).

The primary dataset retained 106 individuals (dataset#1). All individuals of the most distant outgroup, *Anomaloglossus* sp. "Brownsberg" were filtered out. Therefore, we also included the individuals of *A*. sp. "Brownsberg" in an additional dataset (#1-B). A third dataset was produced (#2) filtering out SNPs with minor allele frequency < 2%, and selecting a single SNP per locus (1950 SNPs, 106 individuals, dataset#2). Like the previous dataset, we also generated dataset#2-B, including individuals of *A*. sp. "Brownsberg".

We then applied the following analytical framework to these data: (1) identify population groups and admixed individuals (ADMIXTURE, RAXML) using the largest set of variants (dataset#1 and 1-B); (2) investigate relationships among these groups (SVDquartets, SNAPP) using only unlinked SNPs (dataset#2 and 2-B); (3) analyze the phenotypic (acoustic and morphometric) data in the light of the genetic structure; (4) infer admixture events among these groups (TREEMIX) (dataset#2 and 2-B).

2.5. Species complex structure

We investigated the species complex structure using dataset#1 in ADMIXTURE (Alexander et al., 2009) with ten iterations per K value (K = 2–20). The log likelihood was used to evaluate convergence among iterations, retaining the iteration with the lowest value. Finally, we chose the number of genetic clusters (K) that showed low cross-validation error as the one best describing the structure of the populations (Alexander et al., 2009).

2.6. Phylogenetic analysis

2.6.1. Concatenated loci

We investigated the phylogenetic relationships among individuals with dataset#1 and dataset#1-B concatenated loci in RAXML v8.1.18 (Stamatakis, 2014) with a GTR + G + I model (selected using a Smart Model and Bayesian Information Criterion in PhyML; Lefort et al., 2017) of sequence evolution and 1,000 rapid bootstrap replicates (BP). The tree was rooted with *Anomaloglossus* sp. "Mitaraka" + *A. leopardus*. A preliminary analysis, rooted with *A*. sp. "Brownsberg" (dataset#1-B) led to the same topology.

2.6.2. SVDquartets

The relationships among clusters recovered by ADMIXTURE and RAXML analyses were further investigated with analyses that account for differences in the genealogical histories of individual loci. Specifically, we used the 1950 unlinked SNPs (dataset#2 and #2-B) in SVDquartets (Chifman and Kubatko, 2014), which samples quartets of individuals' variants and infers an unrooted phylogeny, followed by species-tree inference using all sampled quartets. Three independent runs of SVDquartets were conducted in PAUP* 4.0 (Swofford, 2003) to assess topological convergence, each of which included 500 bootstrap replicates and exhaustive quartet sampling.

2.6.3. SNAPp

We additionally investigated the relationships among populations using the SNAPP coalescent-based approach (Bryant et al., 2012) implemented in BEAST v.2 (Bouckaert et al., 2014). This likelihood species-tree method uses independent SNPs to identify potentially reticulate gene histories. Due to SNAPP's computational intensity we used a single random subset of 500 unlinked SNPs from dataset#2. We used empirical estimates of major and minor allele base frequencies to inform mutation-rate priors, gamma rate priors for the alpha and beta parameters, with priors for theta set to default values. Phenotypically distinct populations were used as terminals. The Serra do Navio population was removed because it showed strong admixture between distinct clusters, the investigation of which was beyond the scope of this paper. In order to assess the stability of the results, we also performed an analysis without the exotrophic Borne 4 and Mémora Right/Lourenço populations given that we suspected these populations to have undergone introgression with others (see results). Two replicates of SNAPP with a pre-burnin of 100,000 generations were run for five million MCMC generations, sampling tree and parameter estimates every 500 generations, with a 10% burnin determined *a posteriori* when convergence stationarity was reached in AWTY (Nylander et al., 2007).

2.6.4. Treemix

Finally, TREEMIX v1.1 (Pickrell and Pritchard, 2012) was used to evaluate the occurrence of historical admixture events among populations. This program performs a Gaussian approximation to genetic drift and constructs a maximum likelihood tree of population ancestry for a given set of populations. In addition, it identifies potential unidirectional admixture episodes from the residual covariance matrix. We used 1950 unlinked SNPs (dataset#2), with the 21 populations defined in previous analyses, and rooted the trees with *A. leopardus* + *A*. sp. "Mitaraka". We assessed the addition of one to four migration events, with 10 replicates starting with different seeds and random order of populations and yielding a total of 50 trees. We disabled sample size correction, to avoid overcorrection for very small populations (Pickrell and Pritchard, 2012). As the likelihood function in TREEMIX is a not a maximum likelihood but a composite one, it cannot be used for formal significance testing (Shriner et al., 2016). Hence, out of the ten output trees per number of migration events, the most frequent tree was retained, yielding a reduced number of five trees.

3. Results

3.1. Bioacoustics

The bioacoustic data of the *A. baeobatrachus* complex provides a clear delineation of the two phenotypes. The first two components, with eigenvalues > 1.0 account for 84.5% of the total variation. Coefficients of the first component, which explain 67.1% of the variation, have significant positive loadings for inter-note interval, note length, and call duration, and significant negative loadings for call rate, dominant frequency, and note rate. The second component explains 19.4% of the variation and has significant positive loadings for call rate and note duration, and significant negative loadings for dominant frequency and call duration. Along the first axis, there is no overlap between individuals calling from *terra firme* habitat, emitting rapid trill calls (note rate > 15 notes/s) and with endotrophic tadpoles vs. those calling along streams, emitting slow trill calls (note rate < 15 notes/s) and with exotrophic tadpoles (Vacher et al., 2017). On the two first axes, the representation of the three groups of populations identified by SNP analyses harboring the endotrophic phenotype completely overlap. In contrast, among the exotrophic populations, a clear gap segregates RN2 from the other exotrophic populations, which in turn, overlap with *A.*

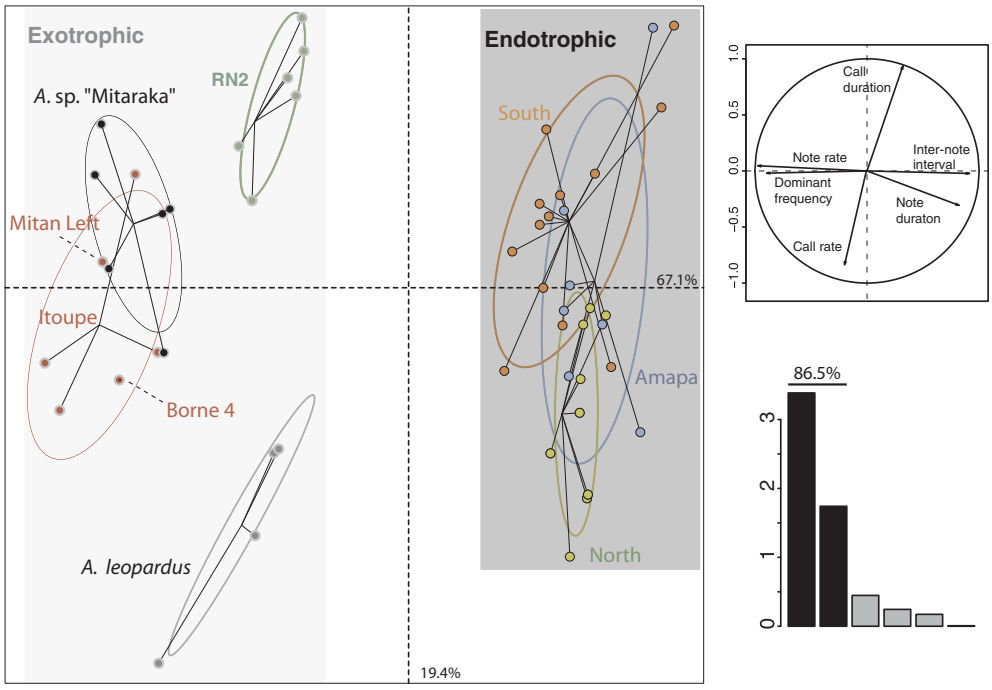

Fig. 2. Principal component analysis of six acoustic variables showing the distinctiveness of the exotrophic (light grey zone) and endotrophic (dark grey zone) populations. Diffusion ellipses and dots are colored according to the major groups of populations recovered from SNPs analyses. Percentage of variance explained by components 1 and 2 are indicated on the right and at the bottom of the axes respectively. On the top right, circle of correlations shows the loadings of individual acoustic characters on the first two principal components, and on the bottom right, eigenvalues show the variance explained by each axis.

sp. "Mitaraka" but not with the more distant *A. leopardus* (Fig. 2).

3.2. Morphometrics

The morphometric data also discriminate the endotrophic and the exotrophic populations, but the difference is less pronounced than with the acoustic data, notably because the RN2 exotrophic population overlaps partially with the endotrophic populations (Fig. 3). The first two components with eigenvalues > 1.0 account for 68.5% of the total variation. Coefficients of the first component, which explains 61.3% of the variation, have significant positive loadings for most variables and are primarily driven by body size. The second component explains 7.2% of the variation. Overall, the endotrophic and exotrophic populations are well differentiated, the first harboring individuals smaller than 17.3 mm while the second (with the exception of the RN2 exotrophic population) harbors individuals larger than 17.3 mm. The use of residuals of the linear regression of the measurement over the body size does not segregate any groups, suggesting that the body proportions are

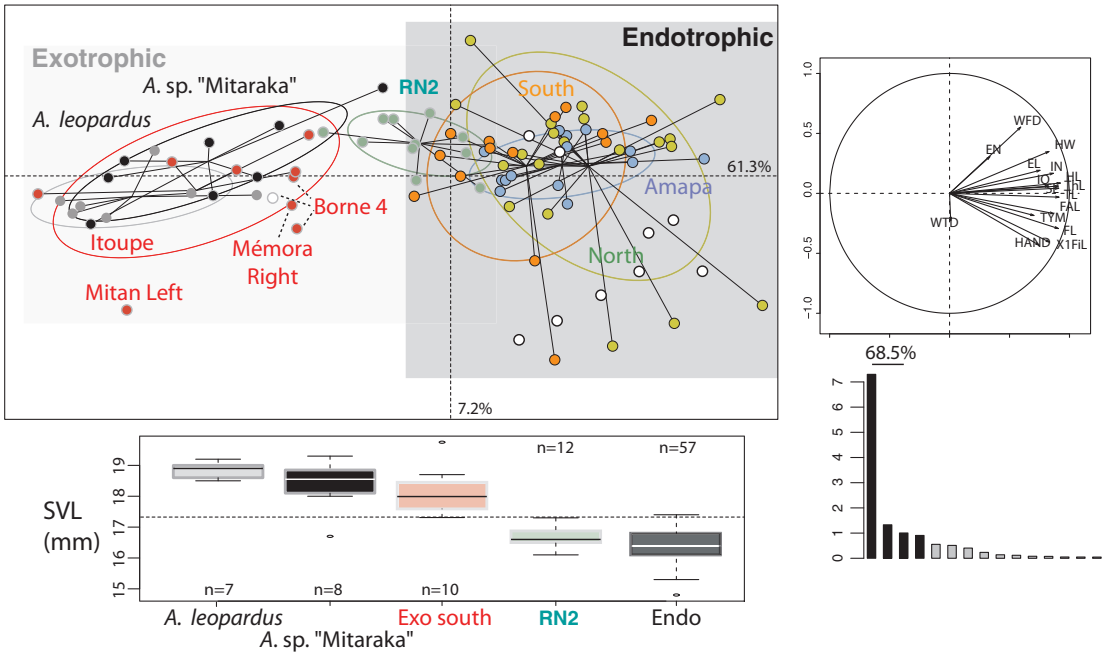

Fig. 3. Principal component analysis of 17 morphometric variables showing the distinctiveness of the exotrophic (light grey zone) and the endotrophic (dark grey zone) populations. Diffusion ellipses and dots are colored according to the major groups of populations recovered from SNPs analyses; unassigned individuals are depicted as white dots with black contour. Percentage of variance explained by components 1 and 2 are indicated on the right and at the bottom of the axes respectively. On the top right, the circle of correlations shows the loadings of individual morphometric variables (see text for abbreviations) on the first two principal components and on the bottom right, eigenvalues show the variance explained by each axes. The body length (SVL) of each major group are plotted as boxplots at the bottom.

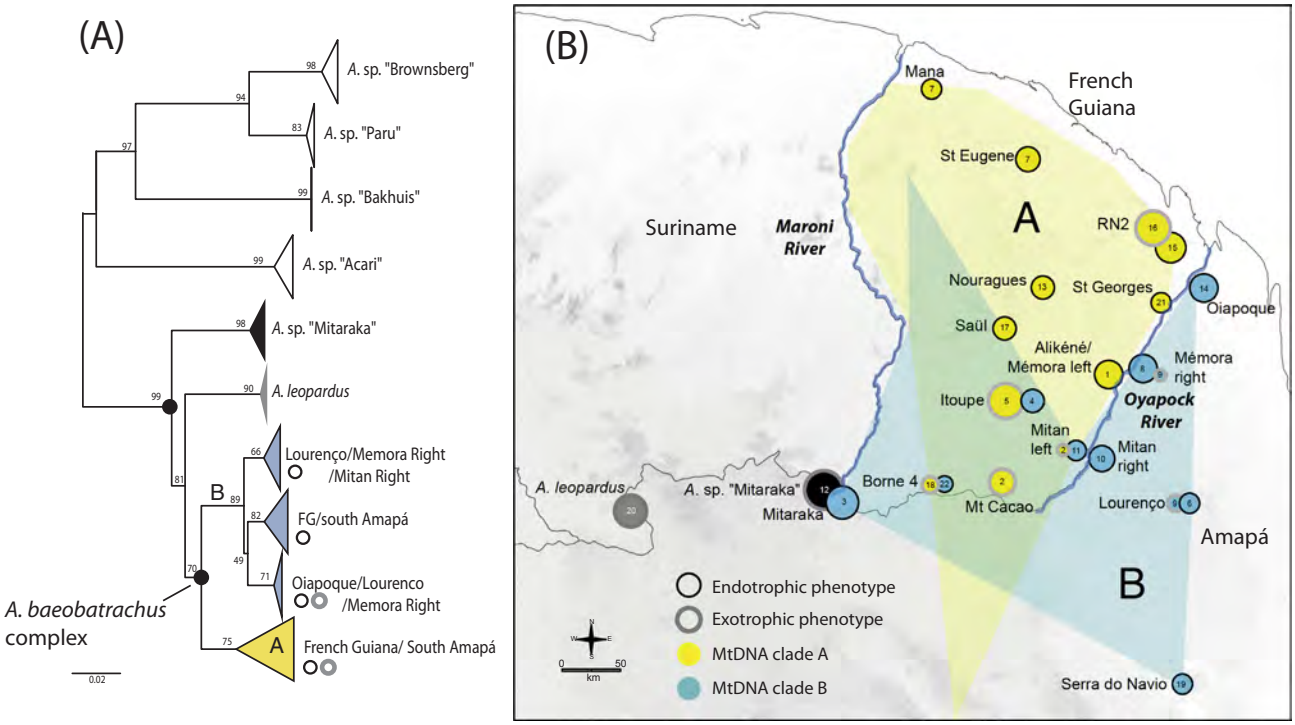

Fig. 4. Phylogenetic relationships based on mtDNA and the respective distribution of the major lineages and phenotypes showing various discordances and spatial overlaps i.e. northern FG endotrophic populations display lineages A (black outlined yellow filled circles) but southern FG endotrophic populations display lineage B (black outlined blue filled circles). (A) Maximum Likelihood tree of 272 terminals represented by 413 bp of 16S rDNA and 14 mitogenomes (at least one for each mtDNA major clade). Bootstrap % are indicated above each node. Relationships among collapsed branches are not strongly supported; (B) Map illustrating the distribution of the two major mtDNA clades (A and B) as filled colored polygons, indented along most of the Oyapock River along which the two clades face each other. Only the populations sampled for analyses of SNP data are depicted by circles whose size is proportional to the number of individuals sampled; fill color follows (A) and illustrates mtDNA lineages, and outlines indicate individuals with exotrophic (grey) and endotrophic (black) phenotypes.

very similar across populations. All unassigned individuals unambiguously cluster with endotrophic populations, with the exception of one individual from Borne 4, which clusters with the exotrophic specimens.

3.3. MtDNA

The combination of mitogenomes and short 16S sequences yields relatively robust support for the monophyly of the *Anomaloglossus baeobatrachus* complex (BP = 70%), the sister group to *A. leopardus* (Fig. 4A). These are, in turn, the sister group to *A.* sp. "Mitaraka" (BP = 99%). The earliest divergence within the *A. baeobatrachus* complex segregates two clades (A and B, respectively BP = 75 and 89%). Clade A is found in most FG localities except the southwest corner and extends to southern Amapá, while clade B occurs in Amapá and southern FG (extending to the Lucifer massif in northern FG, Fig. 4B). The genetic structure within A is complex and poorly supported (not shown), while clear genetic structure exists within clade B, with three distinct clades displaying poorly supported relationships among them, a pattern also reported by Vacher et al. (2017). When examining the correspondence between mtDNA clades and phenotypes, clade A is exclusively harbored by the endotrophic phenotype in the northern part of FG (except population RN2, where both endotrophic and exotrophic individuals harbor clade A) and by the exotrophic populations in southern FG. Conversely, mitochondrial clade B is harbored by the endotrophic phenotype in southern FG and Amapá, and the exotrophic individual from Mémora Right and Lourenço. It is noteworthy that along much of the Oyapock River the two clades A and B face each other, i.e. each occurring only on one bank (Fig. 4).

3.4. Genetic structure of the Anomaloglossus baeobatrachus complex

Most retained K values displayed well-converged log likelihood values. ADMIXTURE cross-validation errors (CV) suggest that the number of genetic clusters best explaining the observed population structure stands between K = 6 and K = 10. Admixture pattern interpretation was difficult beyond K = 9 (Appendix E). We focus here on the structure recovered for K = 6, the smallest value, which is also most suitable for comparison with the following genetic analyses (Fig. 5B). The endotrophic populations are grouped into three spatially explicit clusters: North FG, Amapá, and South FG. The exotrophic populations cluster in two distinct groups (southern populations and RN2).

Where endotrophic and exotrophic phenotypes co-occur (Mitaraka, Itoupé, Borne 4, Cacao, Mitan Left, Lourenço, Mémora Right, and RN2), they are recovered in distinct clusters, suggesting an absence of contemporary gene flow between phenotypically distinct populations. One exception is the pattern observed in Mémora Right and Lourenço exotrophic populations which display signs of admixture between southern exotrophic and the co-occurring endotrophic populations. Admixture is also detected in the Saül and Serra do Navio endotrophic populations that are located in contact zones among the major groups of endotrophic populations (Saül individuals cluster either with North FG or South FG, and one individual from Serra do Navio displays ambiguous assignation between South FG and Amapá – Fig. 5B).

3.5. Phylogenetic analyses

Phylogenetic reconstructions from analysis of concatenated loci (RAXML) and independent SNPs (SVDquartets) using datasets#1-B and #2-B (i.e. including *A.* sp. "Brownsberg") support *Anomaloglossus* sp. "Mitaraka" + *A. leopardus* as the sister clade to the *A. baeobatrachus* complex (BP = 100%) (not shown). The concatenated loci (RAXML,

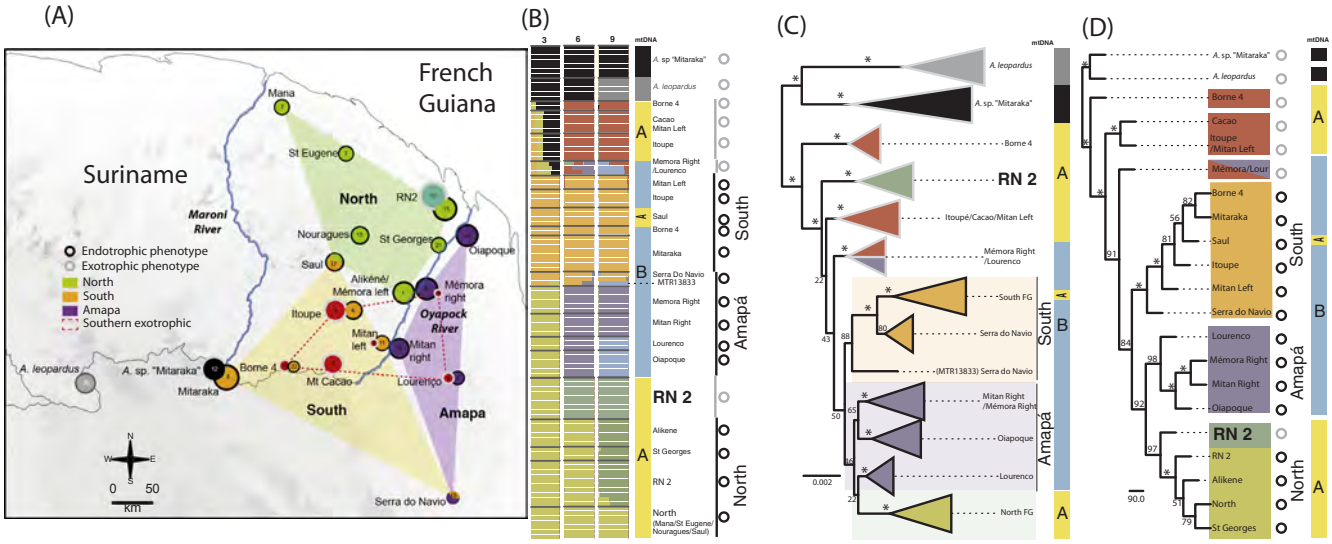

Fig. 5. Genetic structure inferred from SNPs data analyses (ADMIXTURE, RAXML, SVDquartets) showing the nested position of three major groups of endotrophic populations (black outlined circles) within southern FG exotrophic populations (grey outlined circles) except the RN2 population (bold). (A) Map showing the geographic distribution of the major clusters recovered at K = 6 with ADMIXTURE. Filled polygons represent the distributions of the three clusters grouping endotrophic populations (black outlined circles) and the red dotted line polygon corresponds to the exotrophic south FG and Amapá populations (grey outlined circles). (B) Individuals assignments for K = 3–9 from ADMIXTURE based on 16,197 SNPs (2419 loci; dataset#1); corresponding mtDNA lineages are indicated on the right following the yellow/blue color scheme of Figs. 1 and 2. (C) RAXML tree obtained using dataset#1; bootstrap supports are indicated above branches. Branches within clades are collapsed and colors correspond to the results from ADMIXTURE at K = 6. Outline of the clades indicates exotrophic (grey) or endotrophic (black) populations; corresponding mtDNA lineages are indicated on the right. (D) Species tree obtained via SVDquartets using 1,950 SNP (dataset#2) with bootstrap support indicated above branches. Population colors correspond to the results from ADMIXTURE at K = 6 and corresponding mtDNA lineages are indicated on the right.

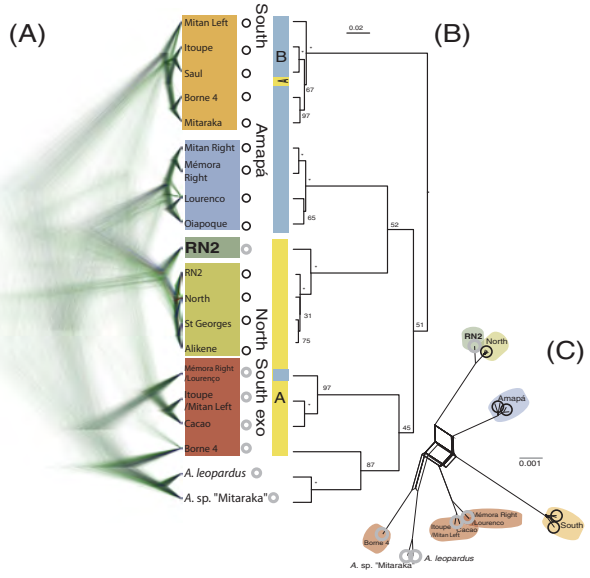

Fig. 6. Species tree obtained from SNAPP analysis of unlinked SNPs showing the multiple topologies supported by the data: (A) All topologies recovered visualized using Densitree, corresponding phenotypes (black outlined circle = endotrophic, grey outlined circles = exotrophic) and mtDNA lineages (yellow = A; blue = B) are indicated on the right; (B) maximum clade credibility tree with node posterior probability (*100) indicated above branches (* = 100); (C) tree network from SPLITSTREE.

Fig. 5C) analysis using dataset#1 further showed that, with few exceptions, *A. baeobatrachus* individuals clustered according to their locality and phenotype, confirming the ADMIXTURE results (Fig. 5B). Across the four methodologies (concatenated loci with RAXML, independent SNPs with SVDquartets, Fig. 3D and SNAPP, Fig. 6, and independent SNP allele frequencies in TREEMIX Fig. 7), we recovered consistent patterns of relationships among the three groups of endotrophic populations, confirming results of ADMIXTURE analysis (North FG, Amapá, and South FG; Figs. 5–7).

The RN2 exotrophic population is recovered as an early diverging lineage among other exotrophic clades in the analysis of concatenated loci with low support. However, it is strongly supported as the sister group to endotrophic northern FG populations in all other methods (Figs. 5–7). Our resulting topologies also differ in the placement of the southern exotrophic clade supported by all methods but SNAPP (Figs. 5–7). This incongruence is echoed by low bootstrap values in the concatenated loci tree (Fig. 5C) and by the reticulation and low posterior probabilities in the SNAPP species tree (Fig. 6B). SNAPP can only produce midpoint rooted trees, and resolution of basal nodes remained ambiguous. However, an additional analysis with SNAPP conducted without the Borne 4 and the Mémora Right/Lourenço populations yielded *A.* sp. "Mitaraka" and *A. leopardus* as the sister group of the *A. baeobatrachus* complex (similar to the other analyses) and a nested position of the southern exotrophic population (Itoupé/Mitan Left, Cacao) within the endotrophic populations with strong support (Appendix F). In summary, apart from the SVDquartets tree (Fig. 5D), the interrelationships remain ambiguous among major groups of the *A. baeobatrachus* complex.

3.6. Patterns of admixture

The topology recovered using TREEMIX without migration (Fig. 7; Appendix G) is identical to the one inferred with SVDquartets (Fig. 5D). This tree ($m = 0$) provided a relatively good fit to the allele frequency data and explained 96.2% (s.e. 0.001) of the observed covariance. However, the residuals of the covariance matrix strongly suggest a poor fit of the model for the relationship between the exotrophic RN2 population and the southern exotrophic populations. Four admixture events (m = 4) improved the fit of the tree up to 99.1% (s.e. 0.001), and the highest scaled residual covariance was low (< 3.3 s.e.). Therefore, we did not consider additional migration events. Trees recovered with

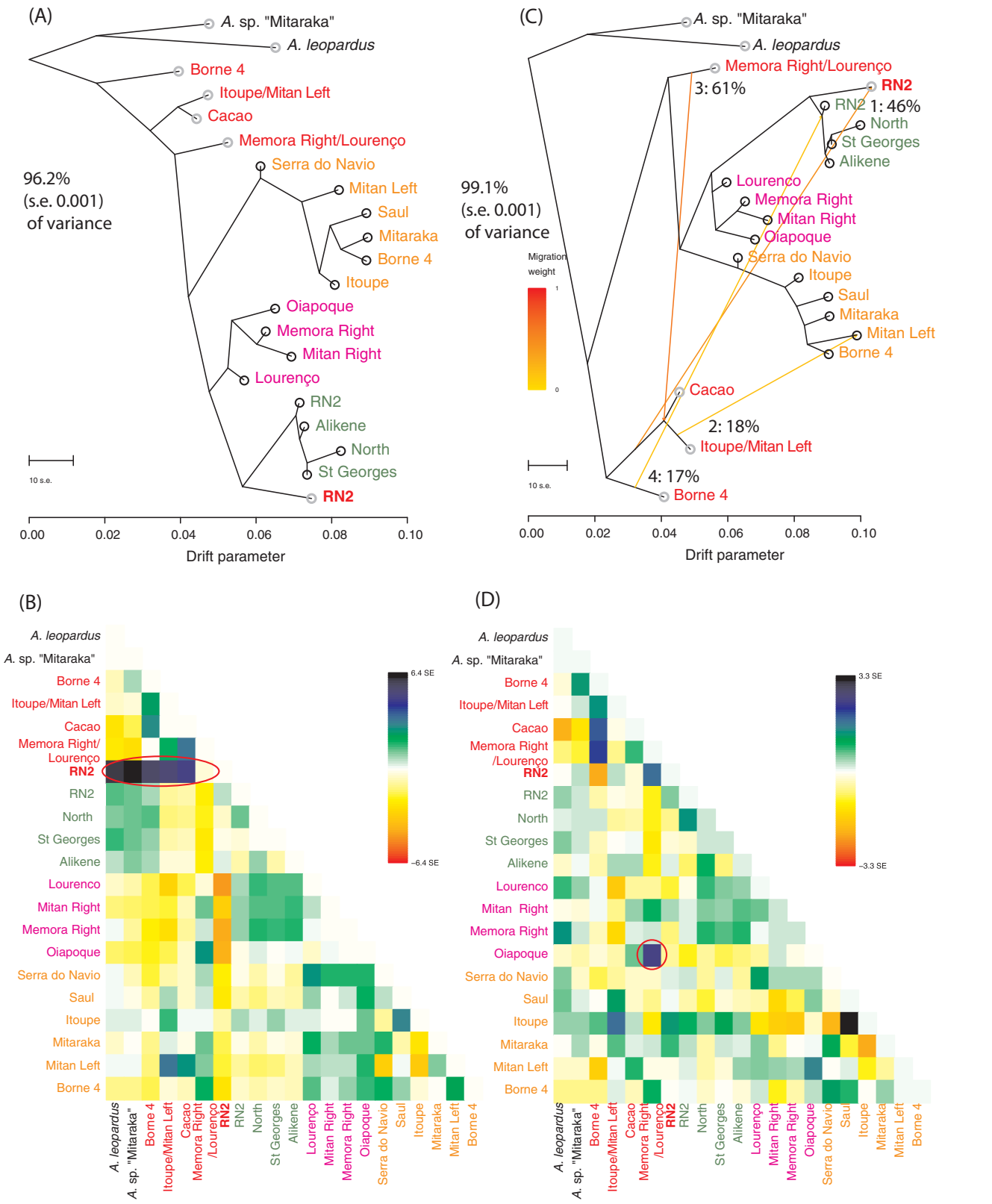

Fig. 7. Maximum Likelihood analysis of the SNPs showing probable admixture events; (A) Maximum likelihood tree resulting from analysis using TREEMIX with no migration. Phenotypes of the populations are indicated according to Fig. 5, black outlined circles for endotrophic; grey outlined circles for exotrophic; (B) Residual fit from the maximum likelihood tree in A. Residuals above zero represent populations that are more closely related to each other in the data than in the best-fit tree and thus are candidates for admixture events. (C) Structure of the phylogeny inferred by TREEMIX allowing for four migration events. (D) Residual fit considering four migration events.

$m = 0$ and $m = 4$ only differ in the position of the exotrophic populations. All inferred admixture events at $m = 4$ involve exotrophic populations.

The exotrophic populations are distributed on a southwest – northeast axis (Fig. 5A) with a growing degree of admixture along this axis (A-Borne 4: 17% of ancestry shared with another population; B-Mitan 18%; C-Mémora Right/Lourenço: 61%; D-RN2: 49%). (A) The southernmost population – Borne 4 – is recovered as the sister lineage to the remainder of the complex (RAXML, SVDquartets) or related to the outgroups (SNAPP, TREEMIX $m = 0$) but shifts to a nested position within the exotrophic population when allowing migration with TREEMIX. We interpret this pattern as the possible consequence of shared relatedness with the other southern exotrophic populations and *A*. sp. "Mitaraka". (B) Eastward, the first three splits separate the Itoupé, Mont Cacao, and then the Mitan Left populations from the remaining ingroup. This group of exotrophic populations is, however, recovered as nested within the endotrophic populations by SNAPP analyses when Borne 4 and Mémora Right/Lourenço are removed (Appendix F), and an admixture event is inferred with the co-occurring endotrophic populations with TREEMIX. This could reflect the fact that these southern exotrophic populations underwent past admixture events when the two newly formed endotrophic and exotrophic ancestors entered into secondary contact. However, this remains relatively speculative with the data at hand, and because subsequent admixture events have most likely complicated the pattern (see below). (C) Further eastward, the Mémora Right and Lourenço populations represent the next group being related to the endotrophic populations in most analyses (RAXML, SVDquartets, TREEMIX). However, these cluster with the southern exotrophic clades in SNAPP analyses and display ambiguous assignment in ADMIXTURE below K = 13 (Appendix E). TREEMIX indicates a high proportion of shared ancestry (61%) with the other exotrophic populations, and residuals indicate that admixture with the co-occurring endotrophic populations in Amapá is likely (Fig. 7). Therefore, we hypothesize that these populations underwent recent admixture with endotrophic populations in Amapá. Unfortunately, we are limited in our inference given the low sampling of Amapá. (D) The most compelling case for admixture is presented by the exotrophic RN2 population, which shares 46.0% of its ancestry with southern exotrophic populations, but is nested as the sister group of the northern FG endotrophic populations in all analyses. The call of this population is more similar to those of the southern exotrophic populations but the phenotype, based on morphometrics, is intermediate between exo- and endotrophic populations. We interpret this pattern as resulting from recent introgression between exotrophic and endotrophic ancestors.

4. Discussion

The first decade of large-scale phylogenetic analyses spawned by next-generation sequencing unraveled signals of incomplete lineage sorting and introgression with an unexpectedly high frequency (Bravo et al., 2018). Notwithstanding, distinguishing between the processes potentially causing conflicting gene trees can be challenging (Choleva et al., 2014). Our exploration of the evolutionary history of French Guianan *Anomaloglossus* reveals them to be a compelling case of multiple introgressions amongst species strikingly distinct across many discrete phenotypic traits. The interrelationships among spatially structured endotrophic populations are stable across analyses. On the contrary, the exotrophic populations display conflicting signals in topologies across analyses, echoing mtDNA-phenotypic incongruences and suggesting historical admixture events. This is particularly striking in the case of the exotrophic RN2 population, which is deeply nested within the endotrophic lineages. We suggest that this population represents one of the most recent instances in a history of admixture events that successively occurred along a southwest-northeast axis.

4.1. A history of introgression

Phylogenetic results suggest that the endotrophic phenotype originated from exotrophic ancestors, confirming the conclusion of Vacher et al. (2017). In addition, our inferences demonstrate that the relationships among the three major groups of endotrophic populations are stable across all analyses. Our analyses provided no support for contemporary interbreeding among the phenotypically distinct, sympatric groups within *A. baeobatrachus*. However, there are clear conflicts across inferences with respect to phylogenetic relationships among the southern exotrophic populations. These conflicting signals provide strong evidence for repeated episodes of gene flow from endotrophic to exotrophic lineages.

Fouquet et al. (2012) estimated the diversification of the *A. baeobatrachus* complex to have begun in the early Pleistocene, a period of rapid fluctuation of temperatures and possibly rainfall in Amazonia (Rull, 2008). The dry-cold conditions of the Pleistocene were likely unfavorable for endotrophic populations, which may require stable wet conditions to breed in the leaf litter on the hilly landscape of the Guiana Shield. Interestingly, the distribution of the endotrophic species suggests that it originated in the easternmost part of the Guiana Shield, where the conditions are today the wettest and were possibly more stable during dry-cold conditions of the Pleistocene. Moreover, the current warm and wet climatic cycle may favor the endotrophic species, which is currently much more abundant than the exotrophic species. Conversely, the interior of the Guiana Shield, from where the exotrophic species seems to originate, is significantly drier. Today, the exotrophic species is rare and found only in small patches isolated along slow-flowing streams. Such conditions may have favored the observed unbalanced directionality of the introgression from endotrophic into exotrophic genomes (see below). The exotrophic populations are distributed on a southwest – northeast axis (Fig. 5A) with an expanding degree of admixture along this axis (A-Borne 4; B-Mitan; C-Mémora Right/Lourenço; D-RN2). Among them, the exotrophic northern population of RN2 is the clearest case of admixture because it is nested as the sister group of the northern FG endotrophic populations in all analyses. We interpret this pattern as resulting from successive introgression during the Pleistocene after secondary contact between exotrophic and endotrophic ancestors (Fig. 8).

Introgressive hybridization arises when species that are not reproductively isolated come into contact. In these cases, asymmetric introgression may occur because the permeability of interspecific barriers varies among genomic regions. In particular, mtDNA easily introgresses and/or replaces pre-existing mtDNA in allospecific gene pools, thanks to an interplay of selection (Bachtrog et al., 2006; Gompert et al., 2008), drift (smaller *N*e), hereditary (maternal) and species behavioral characteristics (philopatry) (Currat et al., 2008; Nevado et al., 2009; Petit and Excoffier, 2009; Seehausen, 2004; Wilson and Bernatchez, 2008). In *Anomaloglossus* our results suggest that the two mtDNA clades A and B were ancestrally harbored by the two phenotypically distinct lineages. MtDNA clade A then introgressed from exotrophic to endotrophic populations (Fig. 8). Subsequently, northern endotrophic populations carrying the mtDNA clade A dispersed southward where they formed a contact zone with the clade B endotrophic populations (Saül, French Guiana). The recent expansion of the clade A, may explain its lack of strong genetic structure, in contrast with clade B.

4.2. Hybridization between reproductively distinct species

Reproductive isolation in animals depends upon the evolution and maintenance of associations between multiple traits contributing to reproductive barriers (Coyne and Orr, 2004; Hoskin et al., 2005). However, in most cases, divergence of phenotypes among sister species remains relatively subtle, except in these traits. Conversely, examples of introgression involving species with qualitatively distinct life history traits such as larval development are rare, probably in part because

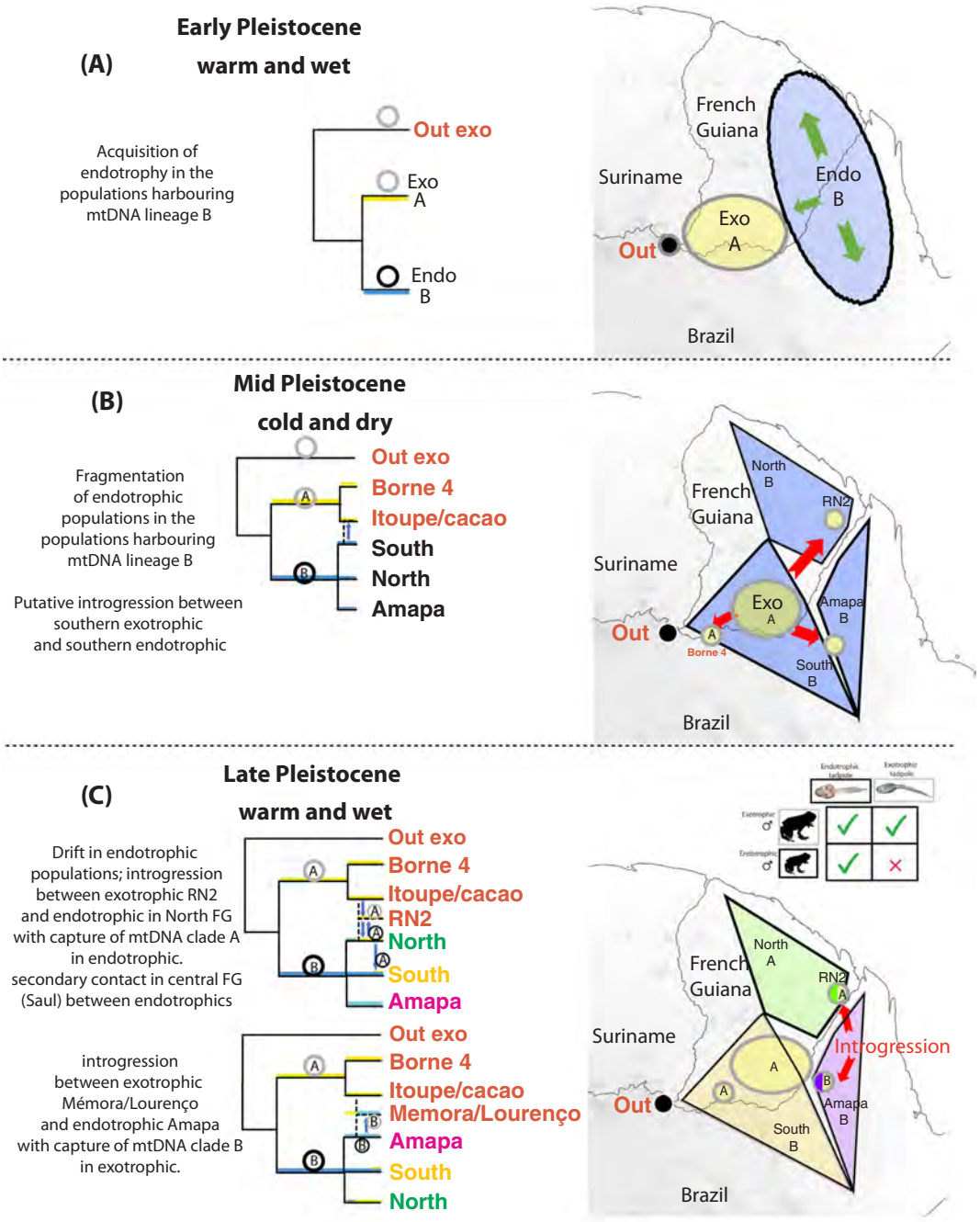

Fig. 8. Hypothesized scenario of successive (A-C) population divergence and introgression during Pleistocene. The yellow/blue colors accompanying the branches depict the trajectory of the mtDNA lineages A & B and the outlines depict the phenotypes (black = endotrophic; grey = exotrophic). The inset depicts the four possible combinations of nesting males and their descents with one combination being necessarily unsuccessful (endotrophic male with exotrophic tadpoles).

there is little intermediate phenotype possible for the offspring. The case of the spring salamanders *Gyrinophilus* (Plethodontidae) is a noteworthy exception given epigean and cave dwelling (paedomorphic) species originated despite ongoing gene-flow during speciation (Niemiller et al., 2008). Similarly, introgression occurs among species of snails of the genera *Alopia* and *Euhadra* despite their distinct chirality that is expected to prevent interbreeding (Koch et al., 2017; Richards et al., 2017). The case of *Anomaloglossus* is no less striking. The nested phylogenetic position of the exotrophic RN2 population and the inferred history of admixture leave little doubt that these admixture events occurred even though the lineages were already characterized by a number of distinct phenotypic traits (larval development, habitat preference, and tadpole transportation behavior, Fig. 1). Such hybridization seems unlikely given (1) there is currently no known intermediate phenotype for some of the traits (larval development), and (2) the large diversity of traits contributing to these alternative life-history strategies. Moreover, hybridization may be successful only in certain conditions. Indeed, it is unlikely that an endotrophic male could successfully raise exotrophic tadpoles that require transportation to a water body because the behavior of transporting may be absent, the weight of the larvae may be too high, and because the available water bodies may be simply too far from the nest (Fig. 8 inset). Conversely, endotrophic tadpoles may successfully be raised by an exotrophic male given they can remain in the nest without specific parental care. Such

asymmetry may be consistent with the inferred directionality of introgression(s) from endotrophic to exotrophic genomes for nuclear SNPs (Fig. 8C). However, since the genic system triggering the larval development is unknown, the understanding of the dominance/inheritance mechanisms remains speculative and out of the scope of the present work.

4.3. Perspective

Our findings highlight the possibility that such cases of ancestral introgression may have been overlooked and may not be as uncommon as previously thought. Nevertheless, introgression between species that differ strikingly in so many traits, especially reproductive and developmental, is particularly intriguing. It deserves further scrutiny and many of the hypotheses we propose, such as a simple genic system driving shifts between exo- and endotrophy and population expansion during favorable climatic cycles, remain highly speculative. Testing these hypotheses will be challenging and requires an understanding of the genic system determining phenotypes, necessitating a search for patterns of expression and selection throughout the genome. Nevertheless, the present study highlights the high resolving power of genomic data for unraveling unexpectedly complex evolutionary scenarios and brought to light a remarkable instance of interspecific hybridization despite extreme differentiation.

Acknowledgments

We thank Andy Lorenzini, Paul Ouboter, Rawien Jairam, Vanessa Kadosoe, Arthur Kocher, Michel Blanc, Sébastien Cally, Vincent Rufray, Olivier Chaline, Christian Marty, and Maël Dewynter for their notable contributions in the field. We also thank the editor Allan Larson, Christophe Thébaud and two anonymous reviewers for their input on the manuscript. Part of the fieldwork was made possible by: “Our Planet Revisited” expedition (Muséum National d'Histoire Naturelle, ProNatura international; APA n° 973-1), with support from Conseil régional de Guyane, Conseil général de Guyane, FEDER funds, Parc Amazonien de Guyane, and DEAL Guyane. The use of the genetic resources was declared to the French Ministry of Environment under the reference TSP 48868, in compliance with the Access and Benefit Sharing procedure implemented by the Loi pour la Reconquête de la Biodiversité, 2017. Part of the material was collected in the Parc Amazonien de Guyane and benefited from an Access and Benefit Sharing Agreement (APA n° 973-23) delivered by the Région Guyane and from a partnership with the Parc Amazonien de Guyane. We thank the Nature Conservation Division and the STINASU for collecting permits in Suriname and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPESP (Fundacão de Amparo à Pesquisa do Estado de São Paulo), ICMBIO (Instituto Chico Mendes de Conservação da Biodiversidade), IBAMA (Instituto Brasileiro do Meio Ambiente e dos recursos naturais renováveis) (permit numbers 07BR000355/DF and 07BR000379/DF, Parque Nacional das Montanhas do Tumucumaque permit 46951-2). In Brazil, we also thank the SETEC (Secretaria de Estado da Cîencia e Tecnologia do Amapá) and especially Erney Plessman de Camargo and José Maria da Silva, for permitting the fieldwork. We acknowledge support from an 'Investissement d'Avenir' grant managed by *Agence Nationale de la Recherche* (CEBA, ref.ANR-10-LABX-25-01) and from the INDIGEN project, attributed to Kwata NGO, through a FEDER/ERDF grant, funded by European Union, Collectivité Territoriale de Guyane, and DEAL Guyane. We thank Pierre Solbès for support with the EDB-cCacl cluster, Sophie Manzi and Amaia Pelozuelo for assistance in the lab.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2018.09.010.

References

Alexander, D., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.
Avise, J.C., Walker, D., Johns, G.C., 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. Roy. Soc. Lond. B: Biol. Sci. 265 (1407), 1707–1712.
Bachtrog, D., Thornton, K., Clark, A., Andolfatto, P., 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. Evolution 60, 292–302.
Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69 (2), 313–319.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10 (4), e1003537.
Bravo, G.A., Antonelli, A., Bacon, C.D., Bartoszek, K., Blom, M., Huynh, S., Jones, G., Knowles, L., Lamichhaney, S., Marcussen, T., Morlon, H., 2018. Embracing heterogeneity: building the tree of life and the future of phylogenomics. PeerJ Preprints 6, e26449v2.
Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29 (8), 1917–1932.
Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22 (11), 3124–3140.
Catchen, J., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J., 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes Genomes Genet. 1, 171–182.
Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30 (23), 3317–3324.
Choleva, L., Musilova, Z., Kohoutova-Sediva, A., Paces, J., Rab, P., Janko, K., 2014. Distinguishing between incomplete lineage sorting and genomic introgressions: complete fixation of allospecific mitochondrial DNA in a sexually reproducing fish (Cobitis; Teleostei), despite clonal reproduction of hybrids. PLoS One 9 (6), e80641.
Coissac, E., 2016. Org.Asm: The ORGanelle AseMbler. Available from: http://pythonhosted.org/ORG.asm/index.html.
Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland, MA, pp. 545.
Currat, M., Ruedi, M., Petit, R.J., Excoffier, L., 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62, 1908–1920.
Dubois, A., 2004. Developmental pathway, speciation and supraspecific taxonomy in amphibians. 2. Developmental pathway, hybridizability and generic taxonomy. Alytes 22, 38–52.
Duvaux, L., Belkhir, K., Boulesteix, M., Boursot, P., 2011. Isolation and gene flow: inferring the speciation history of European house mice. Mol. Ecol. 20 (24), 5248–5264.
Excoffier, L., Foll, M., 2011. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27 (9), 1332–1334.
Fouquet, A., Noonan, B.P., Rodrigues, M.T., Pech, N., Gilles, A., Gemmell, N.J., 2012. Multiple quaternary refugia in the eastern Guiana Shield revealed by comparative phylogeography of 12 frog species. Syst. Biol. 61 (3), 461–489.
Fouquet, A., Souza, S.M., Nunes, P.M.S., Kok, P.J., Curcio, F.F., de Carvalho, C.M., Grant, T., Rodrigues, M.T., 2015. Two new endangered species of *Anomaloglossus* (Anura: Aromobatidae) from Roraima State, northern Brazil. Zootaxa 3926 (2), 191–210.
Gompert, Z., Forister, M.L., Fordyce, J.A., Nice, C.C., 2008. Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. Mol. Ecol. 17, 5231–5244.
Gordon, A., Hannon, G.J., 2010. FASTX-Toolkit. FASTQ/A short-reads pre-processing tools. http://hannonlab.cshl.edu/fastx_toolkit/.
Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., Manzon, R.G., Martino, J.A., Pierson, T., Rogers, S.M., Wilson, J.Y., 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol. Ecol. Resour. 15, 1304–1315.
Gross, B.L., Rieseberg, L.H., 2004. The ecological genetics of homoploid hybrid speciation. J. Hered. 96 (3), 241–252.
Hey, J., 2009. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27 (4), 905–920.
Hoskin, C.J., Higgie, M., McDonald, K.R., Moritz, C., 2005. Reinforcement drives rapid allopatric speciation. Nature 437 (7063), 1353–1356.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Koch, E.L., Neiber, M.T., Walther, F., Hausdorf, B., 2017. High gene flow despite opposite chirality in hybrid zones between enantiomorphic door snails. Mol. Ecol. 26, 3998–4012.
Köhler, J., Jansen, M., Rodríguez, A., Kok, P.J.R., Toledo, L.F., Emmrich, M., Glaw, F., Haddad, C.F.B., Rödel, M.O., Vences, M., 2017. The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice. Zootaxa 4251, 1–124.
Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., Grant, P.R., 2018. Rapid hybrid speciation in Darwin’s fiches. Science 359 (6372), 224–228.
Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. J.

Stat. Softw. 25 (1), 1–18.
Lefort, V., Longueville, J.E., Gascuel, O., 2017. SMS: smart model selection in PhyML. Mol. Biol. Evol. 34 (9), 2422–2424.
Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20 (5), 229–237.
Mavárez, J., Salazar, C.A., Bermingham, E., Salcedo, C., Jiggins, C.D., Linares, M., 2006. Speciation by hybridization in *Heliconius* butterflies. Nature 441 (7095), 868.
Melo, M.C., Salazar, C., Jiggins, C.D., Linares, M., 2009. Assortative mating preferences among hybrids offers a route to hybrid speciation. Evolution 63 (6), 1660–1665.
Neafsey, D.E., Lawniczak, M.K.N., Park, D.J., Redmond, S.N., Coulibaly, M.B., Traore, S.F., Sagnon, N., Costantini, C., Johnson, C., Wiegand, R.C., Collins, F.H., 2010. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. Science 330 (6003), 514–517.
Nevado, B., Koblmüller, S., Sturmbauer, C., Snoeks, J., Usano-Alemany, J., Verheyen, E., 2009. Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. Mol. Ecol. 18, 4240–4255.
Niemiller, M.L., Fitzpatrick, B.M., Miller, B.T., 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. Mol. Ecol. 17 (9), 2258–2275.
Nylander, J.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2007. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24 (4), 581–583.
Petit, R.J., Excoffier, L., 2009. Gene flow and species delimitation. Trends Ecol. Evol. 24, 386–393.
Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS genetics 8 (11), e1002967.
Development Core Team, R., 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna Austria.
Richards, P.M., Morii, Y., Kimura, K., Hirano, T., Chiba, S., Davison, A., 2017. Single-gene speciation: mating and gene flow between mirror-image snails. Evol. Lett. 1 (6), 282–291.
Rull, V., 2008. Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. Mol. Ecol. 17 (11), 2722–2729.
Schumer, M., Rosenthal, G.G., Andolfatto, P., 2014. How common is homoploid hybrid speciation? Evolution 68 (6), 1553–1560.
Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207.
Shriner, D., Tekola-Ayele, F., Adeyemo, A., Rotimi, C.N., 2016. Ancient human migration after Out-of-Africa. Sci. Rep. 6, 26565.
Smith, S.A., Moore, M.J., Brown, J.W., Yang, Y., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15, 150.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.
Strauss, R.E., 1985. Evolutionary allometry and variation in body form in the South American catfish genus *Corydoras* (Callichthyidae). Syst. Zool. 34, 381–396.
Swofford, D.L., 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.
Teeter, K.C., Thibodeau, L.M., Gompert, Z., Buerkle, C.A., Nachman, M.W., Tucker, P.K., 2010. The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64 (2), 472–485.
Vacher, J.P., Fouquet, A., Holota, H., Thébaud, C., 2016. The complete mitochondrial genome of *Anomaloglossus baeobatrachus* (Amphibia: Anura: Aromobatidae). Mitochondrial DNA Part B 1 (1), 338–340.
Vacher, J.P., Kok, P.J., Rodrigues, M.T., Lima, J.D., Lorenzini, A., Martinez, Q., Fallet, M., Courtois, E.A., Blanc, M., Gaucher, P., Dewynter, M., 2017. Cryptic diversity in Amazonian frogs: integrative taxonomy of the genus *Anomaloglossus* (Amphibia: Anura: Aromobatidae) reveals a unique case of diversification within the Guiana Shield. Mol. Phylogenet. Evol. 112, 158–173.
Wilson, C.C., Bernatchez, L., 2008. The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). Mol. Ecol. 7, 127–132.