Genomic phylogeography illuminates deep cyto-nuclear discordances in midwife toads (Alytes)

Ambu, J.; Martínez-Solano, Í.; Suchan, T.; Hernandez, A.; Wielstra, B.M.; Crochet P.-A.; Dufresnes, C.

Genomic phylogeography illuminates deep cyto-nuclear discordances in midwife toads (*Alytes*)

Johanna Ambu [a], Íñigo Martínez-Solano [b], Tomasz Suchan [c], Axel Hernandez [a], Ben Wielstra [d,e], Pierre-André Crochet [f], Christophe Dufresnes [a,*]

[a] *Laboratory for Amphibian Systematics and Evolutionary Research (LASER), College of Biology and the Environment, Nanjing Forestry University, Nanjing, People's Republic of China*
[b] *Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain*
[c] *W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków, Poland*
[d] *Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, the Netherlands*
[e] *Institute of Biology Leiden, Leiden University, P.O. Box 9505, 2300 RA Leiden, the Netherlands*
[f] *CEFE, CNRS, Univ Montpellier, EPHE, IRD, Montpellier, France*

ARTICLE INFO

ABSTRACT

The advent of genomic methods allows us to revisit the evolutionary history of organismal groups for which robust phylogenies are still lacking, particularly in species complexes that frequently hybridize. In this study, we conduct RAD-sequencing (RAD-seq) analyses of midwife toads (genus *Alytes*), an iconic group of western Mediterranean amphibians famous for their parental care behavior, but equally infamous for the difficulties to reconstruct their evolutionary history. Through admixture and phylogenetic analyses of thousands of loci, we provide a comprehensive phylogeographic framework for the *A. obstetricans* complex, as well as a fully resolved phylogeny for the entire genus. As part of this effort, we carefully explore the influence of different sampling schemes and data filtering thresholds on tree reconstruction, showing that several, slightly different, yet robust topologies may be retrieved with small datasets obtained by stringent SNP calling parameters, especially when admixed individuals are included. In contrast, analyses of incomplete but larger datasets converged on the same phylogeny, irrespective of the reconstruction method used or the proportion of missing data. The *Alytes* tree features three Miocene-diverged clades corresponding to the proposed subgenera *Ammoryctis* (*A. cisternasii*), *Baleaphryne* (*A. maurus*, *A. dickhilleni* and *A. muletensis*), and *Alytes* (*A. obstetricans* complex). The latter consists of six evolutionary lineages, grouped into three clades of Pliocene origin, and currently delimited as two species: (1) *A. almogavarii almogavarii* and *A. a. inigoi*; (2) *A. obstetricans obstetricans* and *A. o. pertinax*; (3) *A. o. boscai* and an undescribed taxon (*A. o.* cf. *boscai*). These results contradict the mitochondrial tree, due to past mitochondrial captures in *A. a. almogavarii* (central Pyrenees) and *A. o. boscai* (central Iberia) by *A. obstetricans* ancestors during the Pleistocene. Patterns of admixture between subspecies appear far more extensive than previously assumed from microsatellites, causing nomenclatural uncertainties, and even underlying the reticulate evolution of one taxon (*A. o. pertinax*). All *Ammoryctis* and *Baleaphryne* species form shallow clades, so their taxonomy should remain stable. Amid the prevalence of cyto-nuclear discordance among terrestrial vertebrates and the usual lack of resolution of conventional nuclear markers, our study advocates for phylogeography based on next-generation sequencing, but also encourages properly exploring parameter space and sampling schemes when building and analyzing genomic datasets.

1. Introduction

Obtaining reliable insights on species evolution and diversity is the core basis of various life sciences disciplines. Yet, despite the increasing availability of molecular genetic data in a wide variety of animal species (e.g., Mongiardino Koch et al., 2018; Kuhl et al., 2021), it remains notoriously challenging to obtain robust phylogenies in many groups (Degnan and Rosenberg, 2009; Andrews et al., 2016).

One of the main issues affecting many phylogeographic studies is that they are still essentially based on mitochondrial DNA (mtDNA), both to reconstruct phylogenetic relationships (e.g., Chan et al., 2022) and map lineage distributions via DNA barcodes (e.g., Hebert et al., 2003). While mtDNA is considered a time- and cost-effective tool for species identification (Hebert et al., 2003), it faces inherent limitations that greatly compromise its reliability (Rubinoff and Holland, 2005). Mitochondrial sequences are clonal, haploid, exclusively maternally inherited, and represent a single, small portion of the entire genome, hence they sometimes reflect a different evolutionary history than the species/populations' true ancestry (Ballard and Whitlock, 2004; Zink and Barrowclough, 2008; Cairns et al., 2021; Kimball et al., 2021).

Complementing mtDNA phylogeographies by a small number of nuclear loci typically does not alleviate these problems (especially when they are not analyzed independently of mtDNA), because such datasets often lack informativeness to elucidate recent diversifications. First, the widely used intron gene sequences are usually characterized by slow rates of evolution, hence most of the phylogenetically informative signal is carried by the more rapidly evolving mtDNA (Avise, 1994). Second, because they have a four-times larger effective population size, nuclear markers retain more ancestral polymorphism and exhibit incomplete lineage sorting (ILS) more often (e.g., Pabijan et al., 2012; Cruaud et al., 2014; Dufresnes et al., 2018). Microsatellites are more variable, making them suitable to recover intraspecific patterns of genetic structure (e.g., Zardi et al., 2015; Garcia-Cisneros et al., 2018), but their low cross-amplification performance and elevated rate of homoplasy make them difficult to use in multi-species studies (Chenuil, 2006).

Another common yet still underappreciated problem with mtDNA is the widespread occurrence of discordant patterns between mitochondrial and nuclear markers, so called "cyto-nuclear discordance" (reviewed in Toews and Brelsford, 2012). Most cases of cyto-nuclear discordance originate from introgressive hybridization (e.g., Currat et al., 2008; Lee-Yaw et al., 2019) and can be exacerbated by selective processes (e.g., Bonnet et al., 2017), sex-biased dispersal (e.g., Brandt et al., 2012; Dai et al., 2013; Seixas et al., 2018) or hybrid zone movement (Wielstra, 2019). In theory, ILS can also generate cyto-nuclear discordances (Ballard and Whitlock, 2004) although this is difficult to demonstrate empirically with traditional nuclear loci. The phenomenon appears common especially in recent diversifications involving lineages that have recurrently hybridized during past range shifts, sometimes leading to extreme patterns such as mitochondrial captures, i.e., processes by which populations or even entire species carry foreign mtDNA (e.g., Zieliński et al., 2013; Bonnet et al., 2017; Seixas et al., 2018; Dufresnes et al., 2019b), or ghost mtDNA lineages, i.e., deeply diverged matrilines that do not correspond to any extant nuclear lineage (Recuero et al., 2014; Dufresnes et al., 2020a).

Genomic phylogeographies, i.e., phylogeographic analyses based on genomic data obtained by high-throughput sequencing, such as RAD-sequencing (RAD-seq), offer the opportunity to resolve ambiguous phylogenies and thus shed light on cyto-nuclear discordance (e.g., Cruaud et al., 2014; Edwards et al., 2016; Firneno et al., 2020). In turn, by empirically demonstrating the limits of conventional markers, especially mtDNA, such inferences become instrumental to inform on the potential inaccuracy of many currently accepted phylogenies, which are often taken for granted to support biological and taxonomic conclusions (Rubinoff and Holland, 2005; Chan et al., 2022). Yet, the high resolutive power of RAD-seq and other genomic methods also calls for precautions. Phylogenomic reconstructions may be sensitive to the completeness of alignments (Huang and Knowles, 2016), the method of analysis (Rancilhac et al., 2019; Widhelm et al., 2019), the number of loci considered (Edwards et al., 2015; Hodel et al., 2017), and the admixed ancestry of samples (Morales-Briones et al., 2018; Firneno et al., 2020). More generally, multilocus trees are highly sensitive to gene flow (Leaché et al., 2014a) and can thus be distorted by introgression (e.g., Firneno et al., 2020), including in populations located outside current contact areas, as neutral alleles can escape selection against hybrids and freely diffuse across species boundaries, even between strongly isolated species (e.g., Baldassarre et al., 2014; Dufresnes et al., 2020b, 2021a).

With these issues in mind, here we attempt to clarify the complex phylogeography of a Western Mediterranean group of anuran amphibians, the midwife toads (genus *Alytes*). Midwife toads are famous for their parental care behavior, in which the male carries the eggs on its back until they are ready to hatch (Lange et al., 2022). They are also a group of high conservation concern (Griffiths et al., 2008; Barrios et al., 2012). Some taxa inhabit extremely fragmented ranges, vulnerable to global warming and invasive species (e.g., *A. muletensis* in Mallorca, Moore et al., 2004; *A. dickhilleni* in the Betic Mountains of Spain, García-París and Arntzen, 2002; *A. maurus* in northern Morocco, Donaire-Barroso et al., 2022), while others are sensitive to the devastating chytrid fungus *Batrachochytrium dendrobatidis* (e.g., *A. obstetricans* in Spain, Bosch et al., 2001, 2021). Taxonomic authorities presently list six species (Speybroeck et al., 2020) that branch into three major clades, sometimes attributed to distinct subgenera (Dubois, 1987) or genera (Dubois et al., 2021): (1) the basal species *A. cisternasii* (subgenus *Ammoryctis*); (2) a clade featuring *A. muletensis*, *A. dickhilleni* and *A. maurus* (subgenus *Baleaphryne*); and (3) the *A. obstetricans* species complex (subgenus *Alytes*), which includes no less than six described and undescribed lineages of uncertain relationships and recently divided into two species (*A. obstetricans* and *A. almogavarii*, reviewed by Dufresnes and Hernandez, 2021).

Although many phylogeographic studies on *Alytes* were conducted over the last decades, based on allozymes (García-París, 1995; Arntzen and García-París, 1995), mitochondrial sequences (Gonçalves et al., 2007, 2015), nuclear intron sequences (Maia-Carvalho et al., 2014; Gonçalves et al., 2015), or microsatellites (Maia-Carvalho et al., 2018; Lucati et al., 2022), a fully resolved species tree for the genus has so far eluded herpetologists. Genetic relationships in the *A. obstetricans* complex are particularly puzzling. Mitochondrial and nuclear topologies are radically different, with only a few supported nodes in each (Maia-Carvalho et al., 2014). All delimited lineages seem to hybridize (Maia-Carvalho et al., 2018; Dufresnes and Martínez-Solano, 2020), and genetic screening revealed multiple spatial disparities between the distribution of allozyme variation, microsatellite clusters, and respective matrilines (García-París, 1995; Dufresnes and Hernandez, 2021). A phylogenomic approach thus appears timely, both to resolve pending taxonomic issues and to understand true species' diversity and evolution within the complex. Such insights would also widen perspectives to exploit the genus in evolutionary and behavioral ecology studies (e.g., Penna et al., 2015; Martínez-Gil et al., 2022) – midwife toads are a historical amphibian model to study the epigenetics of reproductive behavior (Arntzen, 2022). In turn, this group provides an opportunity to assess the reliability of conventional phylogeographic analyses (i.e., based on few markers) *vs* genomic inferences when cyto-nuclear discordance and admixture potentially blur our appreciation of species diversifications.

In this study, we generated and analyzed RAD-seq data obtained from geographically representative samples of all known *Alytes* lineages. We first used population genetic analyses to verify whether spatially-distant samples of the *A. obstetricans* complex could still show a trace of recent admixture. Second, we reconstructed *Alytes* species trees to infer evolutionary relationships and pinpoint cyto-nuclear discordance. In this respect, we explored how the phylogenetic results were affected by the number of loci shared between species, as adjusted by the stringency of filtering, and whether recently admixed samples are included in the analyses or not.

2. Methods

2.1. DNA sampling

DNA samples from 52 *Alytes* representative of all species of the genus were obtained from herpetological collections, including MNCN (Museo

Nacional de Ciencias Naturales) and BEV (CNRS – EPHE collection of the Biogeography and Ecology of the Vertebrates team in CEFE, Montpellier) (Table S1). These consisted of buccal swabs stored dry and frozen (live adults), tail clips stored in 96 % ethanol (tadpoles), and skin/muscle pieces (preserved specimens). Most samples had been analyzed in previous studies (Gonçalves et al., 2015; Dufresnes and Hernandez, 2021; Vliegenthart et al., 2023), and we either reused available DNA extracts or isolated new DNA with the DNeasy Blood & Tissue Kit (Qiagen). Except for *A. muletensis*, for which only captive-bred specimens were available, samples broadly cover the ranges of each species while avoiding putative contact zones and thus potentially admixed individuals for the *A. obstetricans* complex (according to Gonçalves et al., 2015; Maia-Carvalho et al., 2018).

2.2. RAD-seq

A genomic library was prepared using a modified version of the double-digest RAD (ddRAD) protocol of Brelsford et al. (2016). Briefly, this involves enzymatic digestion by *SbfI* and *MseI* restriction endonucleases, adaptor ligation (with unique 4–8 bp barcodes on the *SbfI* end), duplicate PCR amplification of the ligated fragments, and size selection between 400 bp and 500 bp. Compared to the original protocol, two PCR replicates were run, and size selection was performed with a PippinPrep (Sage Science). The purified final library was sequenced (paired-end) on an Illumina NextSeq 550. Raw reads were demultiplexed with STACKS 2.59 (Catchen et al., 2013) and trimmed to 100 bp. We then used the denovo_map.pl pipeline of STACKS to construct RAD loci, assembly and cataloging, with default *-m*, *-n*, and *-M* values, which provided a good balance between data quality and quantity (Paris et al., 2017). The STACKS catalog contained 357,379 loci, with an average coverage of 18 reads ($SD = 5.5$). Loci were then filtered using the module *populations* for downstream analyses, as described below.

2.3. Population genetic analyses of the A. obstetricans complex

A genotype matrix was obtained for the 38 individuals of the *A. obstetricans* complex, considering only the markers present in all six lineages, and in 70 % of the samples of each (filter *-r* 0.7). This represented 3,487 RAD-tags, and to avoid physically linked loci, only a single SNP per RAD-tag was retained (filter *-write-random-snp*). The Bayesian clustering algorithm of STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to infer individual ancestries to genetic groups, based on the admixture model with uncorrelated allele frequencies. Clustering solutions for $K = 1$ to 10 were explored through at least five replicate runs per K, each of 200,000 iterations after a burnin period of 20,000. Run convergence was assessed by inspecting the STRUCTURE barplots and by monitoring the run likelihood $L(P|K)$. Converged runs were compiled and summary statistics ($L(P|K)$, ΔK) were computed for each K with STRUCTURE HARVESTER (Earl & vonHoldt, 2012). A neighbor joining tree plotting allele frequency divergences among STRUCTURE clusters for $K = 6$ was computed. In addition, a Principal Component Analysis (PCA) was performed with the R package *adegenet* (Jombart, 2008). Because PCA does not tolerate missing data – these need to be replaced by average allele frequencies – one individual (MNCN.68632) that featured an excessively large proportion of missing data (85 %) was excluded from the analysis.

To independently validate population differentiation and assignment, we constructed a co-ancestry matrix from our RAD-seq data using fineRADstructure (Malinsky et al., 2018), considering all the variation (20,449 SNPs) found at the 3,487 RAD-tags. The loci were first reordered according to linkage disequilibrium (software-provided script). We used a 100,000 burnin followed by 100,000 MCMC steps sampling every 1,000 steps and the tree was constructed with 10,000 hill-climbing iterations. The results were visualized using the R scripts provided with the software package.

To measure genetic distances between lineages and test specific

hypotheses regarding the origin of *A. o. pertinax* (see Results), another genotype matrix was generated from a subset of 26 "pure" individuals, i. e., considered free of recent admixture (ancestry > 0.95 to their corresponding clusters, see Results and Table S1), using the same filtering parameters as above. This dataset contained 3,669 unlinked SNPs. We computed pairwise genetic distance (F_{st}) between lineages with the R package *hierfstat* (Goudet, 2005). Approximate Bayesian Computation (ABC) analyses were then performed in DYIABC 2.1.0 (Cornuet et al., 2014). Five scenarios were designed, in which the *pertinax* lineage either diverged from *obstetricans* (scenario I), or has a hybrid origin, namely between *obstetricans* (its mitochondrial ancestor and closest relative in nuclear trees) and another lineage (scenarios II–V). Coalescent events followed the most likely topology of the complex, as retrieved by our phylogenetic analyses, with model parameters including the effective population size of each lineage (N_e), their coalescent time (T_d), and the admixture rate for scenarios II-V. Preliminary analyses were run to check and optimize prior distributions, validate the choice of summary statistics, and pre-evaluate scenario-prior combinations, hence ensuring that the simulated datasets overlapped with the observed dataset and underlie realistic scenarios. For the final models, we kept uniform prior distributions and set the bounds for the N_e parameters to 1–20,000 individuals, and for T_d parameters to 10–100,000 generations, with prior conditions on T_d to avoid incongruencies in respect to the topology. We used the one-sample (genetic diversities), two-sample (F_{st} and Nei genetic distances) and three-sample (maximum-likelihood coefficient of admixture) summary statistics that pertained to the *pertinax* lineage, for a total of 60 statistics. For final analyses, we generated five million simulations, i. e., one million per scenario (the optimal amount recommended by the software), and we evaluated which scenario was the closest to the observed data by calculating their posterior probabilities from the direct (P_{dir}) and logistic (P_{log}) methods among the best 500 and 10,000 simulations, respectively.

2.4. Phylogenetic analyses

For comparative analyses, we designed three sets of *n* samples representative of all *Alytes* species and lineages (Table S2). For each set, a series of 11 concatenated supermatrices was obtained by varying the number of individuals in which each locus must be present (filter *-p*), ranging from all of them (*-p = n*), to all but ten of them (*-p = n*-10). The first sample set included the 52 individuals of the study. The second sample set included 45 individuals identified as "pure" by the population genetic analyses of the *A. obstetricans* complex (ancestry > 0.95 to their corresponding clusters, see Results). The third sample set included 38 individuals of the second set that featured $< 10 \%$ of missing data under the intermediate filtering solution (*-p* 40). This spectrum of filters thus allowed to explore the balance between the number of retained loci and the completeness of supermatrices, while the different sample sets allowed to account for potential biases driven by substantial recent admixture and missing data on the retrieved phylogenies. All alignments were exported with (option *-phylip-var-all*), and without invariant sites.

The 33 supermatrices (3 sample sets × 11 filtering options) were analyzed under three different phylogenetic analyses. First, we performed Bayesian inferences (BI) in BEAST 2.6 (Bouckaert et al., 2019) on concatenated alignments (including invariant sites), using a birth–death tree model, a strict molecular clock and a GTR + G + I model of sequence evolution. To time-calibrate the tree, we set up a normally distributed prior of 5.33 ± 1.0 million years (My) for the diversification of the common ancestor of the North African *A. maurus,* the Balearic *A. muletensis* and the Iberian *A. dickhilleni*, which corresponds to the end of the Messinian Salinity Crisis. This period is marked by the reopening of the Gibraltar Strait, which, by refilling of the Mediterranean Sea, putatively isolated North African and European populations of *Alytes* (Arntzen and García-París, 1995), a hypothesis consistent with previous mitochondrial dating (Martínez-Solano et al., 2004; Ehl et al., 2019). Trees were sampled every 10,000 iterations and chains were run until

stationarity and large (>200) effective sample sizes of parameters were reached, as monitored in Tracer 1.7 (Rambaut et al., 2018). The sampled trees were visualized in DensiTree 2.0 (Bouckaert and Heled, 2014), applying a burnin threshold of 20 %, and a maximum-clade credibility tree was produced in the BEAST module TreeAnnotator. Second, we performed maximum-likelihood (ML) phylogenetic reconstructions on supermatrices restricted to variable sites (SNPs) with PhyML 3.1 (Guindon et al., 2010), using a GTR + G + I model and the aLRT SH-like method as an approximate estimation of node support, with 100 bootstrap replicates. Third, distance-based neighbor-joining (NJ) trees were also inferred from the SNP supermatrices in SeaView 5 (Gouy et al., 2021), with 100 bootstrap replicates. In each analysis, the dominant topology was recorded, and the tree was considered robust if all internal branches received at least 0.95/95 of node support. For visualization, we also illustrated the most plausible *Alytes* tree by a phylogenetic network built in SplitsTree 4.18.3 (Huson and Bryant, 2006), using the uncorrected P method.

In addition, we attempted to reconstruct species trees with the multispecies coalescent of SNAPP (Leaché et al., 2014b) implemented in BEAST 2. Specifically designed for phylogenetic analyses of unlinked SNP data, SNAPP is nevertheless limited by the need for each locus to be genotyped in each pre-defined group, which drastically reduces the number of informative loci between diverged species, even with relaxed bioinformatic filters. To maximize the number of informative loci and minimize computational times, we selected 25 individuals from sample set 3 above, to retain a maximum of three individuals per lineage (Table S2). We then exported three SNP datasets that differed in filtering stringency, keeping RAD loci sequenced in either all (*n-p* = 0), 20 (*n-p* = 5) or 15 individuals (*n-p* = 10), and calling a single SNP per RAD-tag (filter *-write-random-snp*). SNAPP analyses were performed on each dataset with model parameters and priors optimized following the recommendations of Leaché and Bouckaert (2018). The chain was sampled every 1,000 iterations and ran at least 1 million iterations, by which long-term stationarity was reached. Results were visualized by cloudograms in DensiTree 2.0, with a burnin of 50 %.

Finally, to allow comparison with our nuclear inferences, we reanalyzed the mitochondrial alignment of Maia-Carvalho et al. (2014), which spans 10 gene fragments for a total of 9,045 bp, by a timecalibrated phylogenetic inference in BEAST. Invariant sites were included, and the same parameters of model of sequence evolution, clock model, tree prior, and the time calibrations were applied as above.

3. Results

3.1. Genetic structure in the A. obstetricans complex

Based on 3,487 unlinked SNPs, genetic clustering of 38 individuals from the *A. obstetricans* complex roughly matched the mitochondrial lineages delimited and barcoded in previous studies (Fig. 1). In the STRUCTURE analysis, $L(P|K)$ steadily increases up to $K = 6$, above which the runs do not converge but drastically fluctuate between low likelihood solutions (Fig. S1); inspections of $K = 7$–10 runs reveal no additional patterns of structure. Consequently, the ΔK showed a strong spike for $K = 6$ ($\Delta K = 57,242$; Fig. S1). The six clusters correspond to the lineages identified as *A. a. almogavarii* (yellow in Fig. 1), *A. a. inigoi* (beige), *A. o. obstetricans* (red), *A. o. pertinax* (orange), *A. o. boscai* (blue), and *A. o.* cf. *boscai* (green). Although individuals were sampled in presumably "pure" populations, according to previous phylogeographic studies (Dufresnes and Hernandez, 2021), the analysis identified longrange admixture among the two most closely related pairs of taxa: between *A. o. obstetricans* and *A. o. pertinax* in the Cantabric coast and adjacent Pyrenees; between *A. a. almogavarii* and *A. a. inigoi* in the Central Spanish Pyrenees (including the holotype of *A. a. inigoi*, Table S1). The co-ancestry matrix and associated dendrogram obtained with fineRADstructure independently confirmed the population clustering results (Fig. S2). All six lineages of the *A. obstetricans* complex are
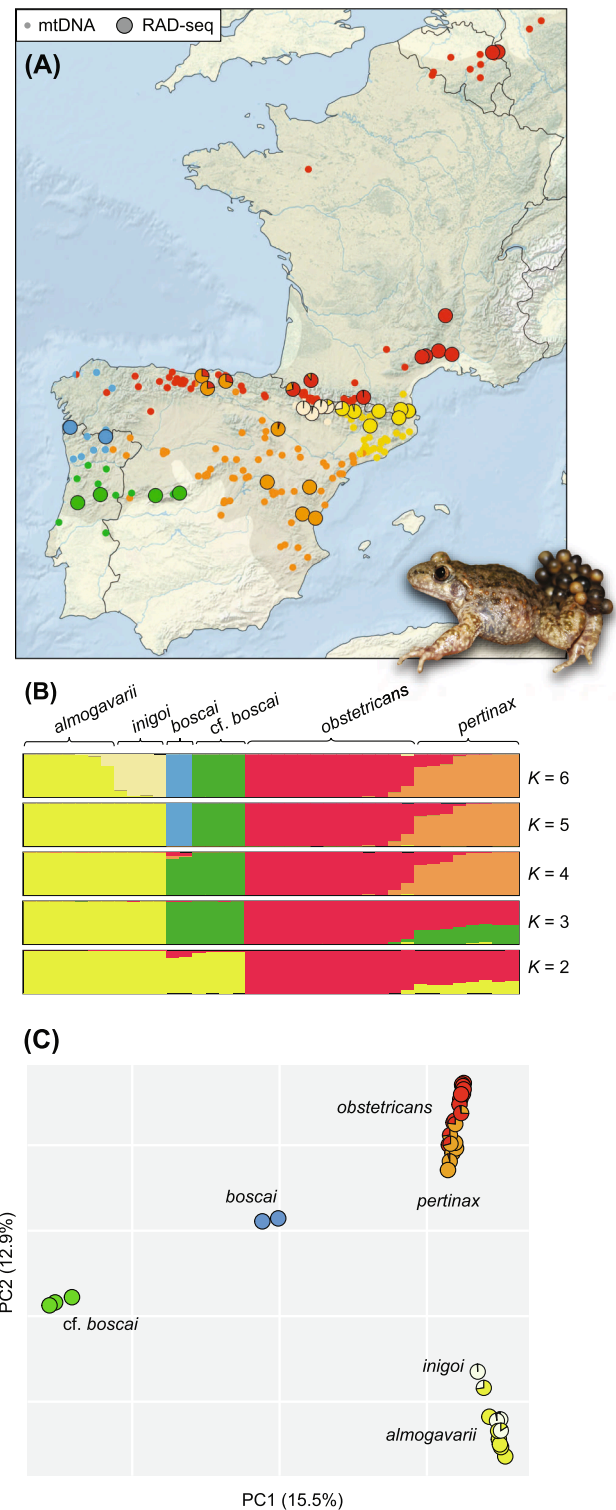

Fig. 1. Phylogeography of the *A. obstetricans* species complex (subgenus *Alytes*). (A) mtDNA barcoding (small dots) from Gonçalves et al. (2015, and references therein), Dufresnes and Martínez-Solano (2020), Lucati et al. (2022) and Vliegenthart et al. (2022, and references therein) overlayed by nuclear ancestry of localities assigned to $K = 6$ groups by STRUCTURE, estimated from 3,487 unlinked SNPs; obtained by RAD-seq; (B) Individual ancestries for $K = 2$–6 groups; (C) Principal Component Analysis (PCA) on individual allele frequencies, displayed with STRUCTURE ancestries. On the map, grey shades represent range distributions according to the IUCN Red List (IUCN SSC Amphibian Specialist Group, 2022).

distinguished, and individuals flagged as admixed were further differentiated yet grouped close to the lineages they shared most ancestry with (Fig. S2). Moreover, the PCA delimited four clusters by the first two axes (Fig. 1C): *A. o. boscai*, *A. o.* cf. *boscai*, *A. o. obstetricans/pertinax*, and *A. a. almogavarii/inigoi*, the latter pairs being arranged in mixes of pure and admixed individuals. Pairwise F_{st} confirmed the strong genetic distances between all six lineages (0.52–0.82; Table S3, Fig. S3).

It is worth noticing that low *K* values did not perfectly group the "pure" *A. o. pertinax* samples with the cluster assigned to *A. obstetricans*, but these shared ancestry with other clusters, especially *A. almogavarii* ($K = 2$) or *A. o. boscai*/cf. *boscai* ($K = 3$) (Fig. 1B). To understand whether this pattern reflects ancestral polymorphism or historical admixture, we designed ABC simulations modelling the origin of *A. o. pertinax* (Fig. 2A). A hybrid origin was retrieved as the most likely demographic scenario (Fig. 2B). Posterior probabilities of scenario choice (P_{dir} and P_{log}) were both null for the hypothesis of simple genetic divergence from *A. o. obstetricans* (scenario I). Instead, most of the best simulations were obtained under scenarios IV and V, where *A. o. pertinax* resulted from an admixture event between *A. o. obstetricans* and either *A. a. almogavarii*

($P_{dir} = 0.43$ and $P_{log} = 0.87$) or its sister lineage *A. a. inigoi* ($P_{dir} = 0.45$ and $P_{log} = 0.12$), respectively. Simulations under alternative hybrid origins contributed little in comparison (scenario II: $P_{dir} = 0.07$ and $P_{log} = 0.01$; scenario III: $P_{dir} = 0.06$ and $P_{log} = 0.00$).

3.2. Phylogenetic analyses

The 33 datasets built for the phylogenetic analyses, which differed by sampling schemes (all samples, $n = 52$; without admixed samples, $n = 45$; without admixed samples and samples with high rate of missing data, $n = 38$) and by filtering stringency (number p of samples where a locus must be present), consisted of alignments ranging from 1,473 bp (50 SNPs) to 696,719 bp (35,214 SNPs). As expected, the number of loci (and thus the length of alignments) was higher for the smaller sample sets, and when loci dropout is tolerated for a higher number of samples (n-p) (Fig. 3, Fig. S4).

The three sets of phylogenetic analyses (BI, ML, NJ) on these 33 datasets provided mostly consistent results (Fig. 3). All trees featured the three main *Alytes* clades/subgenera: (1) *A. cisternasii* (subgenus
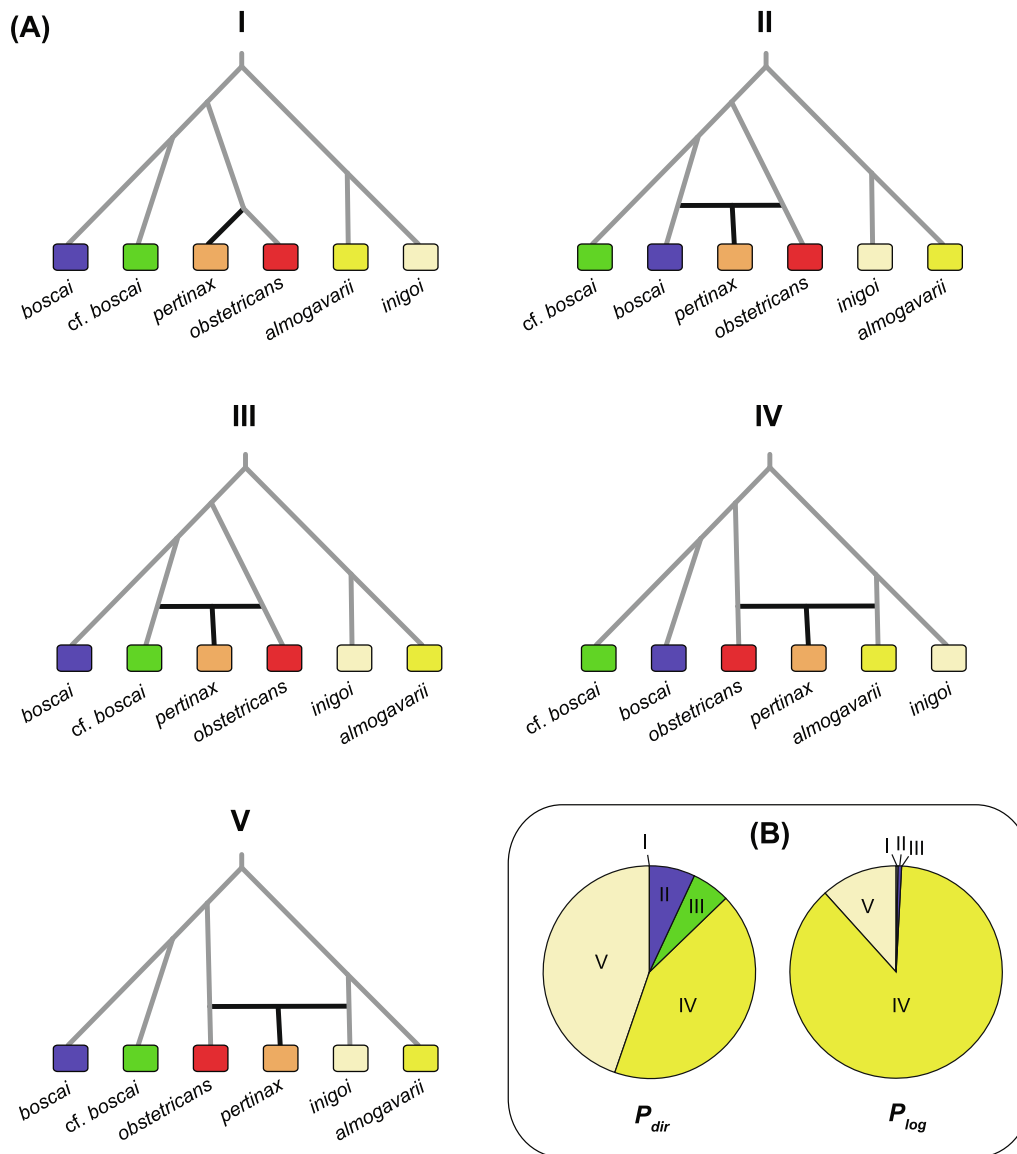

Fig. 2. Approximate Bayesian Computation (ABC) hypotheses of the origin of *A. o. pertinax* based on 3,669 unlinked SNPs. (A) Graphical displays of scenarios where *A. o. pertinax* diverged from *A. o. obstetricans* (I) or originates from admixture between *A. o. obstetricans* and one of the other lineages (II–V); (B) Proportions (posterior probabilities) of each scenario in the best simulations, as calculated with the direct (P_{dir}) and logistic (P_{log}) methods.
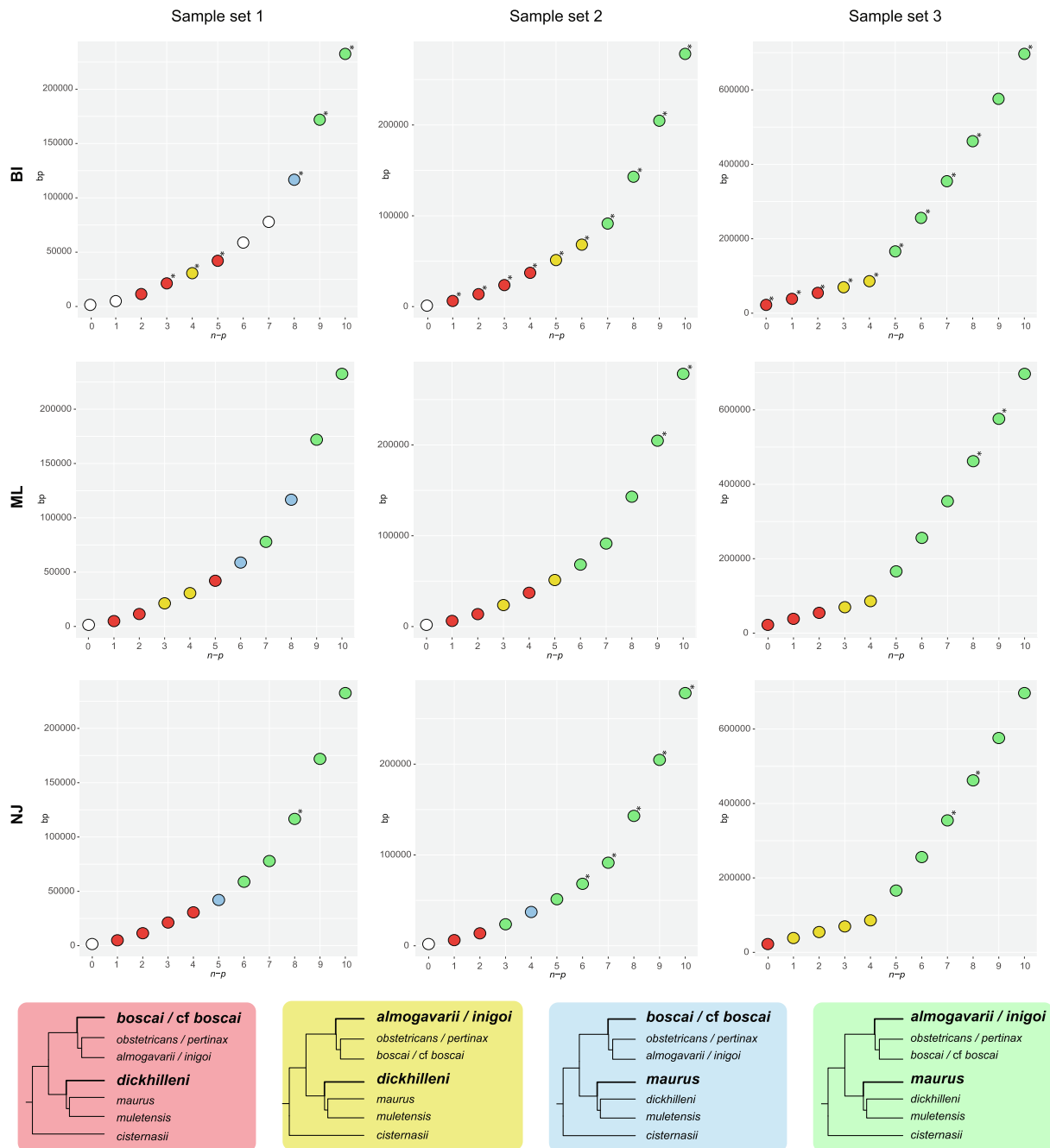
Fig. 3. Topologies obtained in respect to alignment length (bp) and the number of individuals allowed for locus dropout (*n-p*) for three defined sample sets and three methods of analyses. Sample set 1 corresponds to all 52 individuals of the study (including admixed individuals and individuals with a high proportion of missing data); sample set 2 is restricted to the 45 individuals identified as "pure"; sample set 3 is restricted to the 38 individuals identified as "pure" and with low proportions of missing data (see Methods for details). Analyses include Bayesian inference (BI) on concatenated supermatrices, and Maximum-Likelihood (ML) as well as Neighbor-Joining (NJ) on SNP alignments. The four competing topologies, as illustrated by the trees (white: no single major topology recovered), are distinguished by colors. Asterisks denote robust topologies (all posterior probabilities/bootstraps of major nodes above 0.95/95).

Ammoryctis); (2) *A. maurus/dickhilleni/muletensis* (subgenus *Baleaphryne*); (3) the *A. obstetricans* complex (subgenus *Alytes*). In the latter, three subclades were always retrieved: *A. a. almogavarii/inigoi*; *A. o. boscai/*cf. *boscai*; *A. o. obstetricans/pertinax*.

Yet, two main discrepancies, both involving closely related nodes, arose among the analyses: the relative position of *A. maurus* and *A. dickhilleni* within subgenus *Baleaphryne*, and the relative position of *A. a. almogavarii/inigoi* and *A. o. boscai/*cf. *boscai* within subgenus *Alytes*. This resulted in four competing topologies (Fig. 3), in which either *A. maurus* and *A. o. boscai/*cf. *boscai* (blue-coded topology), *A. dickhilleni*

and *A. o. boscai/*cf. *boscai* (red-coded topology), *A. dickhilleni* and *A. a. almogavarii/inigoi* (yellow-coded topology), or *A. maurus* and *A. a. almogavarii/inigoi* (green coded topology), are the first to diverge in their respective clades.

The predominant topology and its support varied mostly in respect to the number of loci of the datasets, the inclusion of admixed individuals, and the method of analysis (Fig. 3). Only one topology per dataset was retrieved in most cases, with robust support in the BI analysis. The shorter alignments (<150 kb, obtained with *n-p* ≤ 4–8) alternatively retrieved the red, yellow, or blue topologies, while the longer

alignments (≥150 kb, obtained with *n-p* ≥ 5–9) eventually recovered only the green topology. We observed the same pattern in the ML and NJ analyses, although, with these methods the ambiguous nodes were robustly supported only in the larger datasets (>50–100 kb SNPs, obtained with *n-p* ≥ 5–9). In general, the most variable results were obtained from the first sample set, which included several admixed samples. In contrast, the second and third sample sets, which differed by the inclusion of samples featuring high rates of missing data, provided mostly similar patterns.

Based on a subset of 25 samples (1–3 per lineage), SNAPP, which only considers loci sequenced across all lineages, could only exploit a limited number of informative loci, namely 220, 587 and 630 SNPs depending on filtering stringency. None of these inferences converged towards a single robustly supported tree, although the green topology (notably the basal position of *A. almogavarii/inigoi* in the *A. obstetricans* complex) was predominantly represented in the analysis of the largest dataset (Fig. S5).

From these observations, we consider the green topology as the most likely *Alytes* tree. A maximum-clade credibility consensus tree summarizing the time-constrained BI analysis is reported in Fig. 4, with terminal branches detailed in Fig. S6. Inferences obtained from the corresponding ML and NJ analyses are reported in Fig. S7 and Fig. S8, respectively. A phylogenetic network is also available in Fig. S9.

Our time tree indicates a Mid-Miocene origin for the early divergence of *A. cisternasii* (9.9 Mya), followed by the split between subgenera *Alytes* and *Baleaphryne* (7.7 Mya). Subgenus *Baleaphryne* diversified at the Mio-Pliocene transition, with the nearly simultaneous evolution of *A. maurus* (5.33 My, our calibration), *A. muletensis* and *A. dickhilleni* (4.9 My). In subgenus *Alytes*, *A. almogavarii* and *A. obstetricans* diverged in the Pliocene (3.9 Mya), and the latter was then subdivided into *A. o. obstetricans/pertinax* and *A. o. boscai*/cf. *boscai* (3.2 Mya). All three diversified at the beginning of the Pleistocene, with the split between *A. a. almogavarii* and *A. a. inigoi* (2.2 Mya), between *A. o. boscai* and *A. o.* cf. *boscai* (2.5 Mya) and between *A. o. obstetricans* and *A. o. pertinax* (2.0 Mya), noting that the latter age is not meaningful since *A. o. pertinax* most likely appeared by hybridization with *A. almogavarii* (Fig. 2).

The nuclear tree features three main discordances with the mitochondrial tree (Fig. 4). Both *A. a. almogavarii* and *A. o. boscai* carry mtDNA related to *A. o. obstetricans/pertinax*, even though these taxa show deeper nuclear divergences, being more closely related to *A. a. inigoi* and *A. o.* cf. *boscai*, respectively. According to the mtDNA tree topology and divergence estimates, the discordances suggest that *A. a. almogavarii* and *A. o. boscai* putatively lost their mtDNA through hybridization with the ancestor of *A. o. obstetricans/pertinax* during the Pleistocene, namely around 1.8 Mya for the former and 0.7 Mya for the latter (Fig. 4). The very young mitochondrial split between *A. o. obstetricans* and *A. o. pertinax*, consistent with the hybrid origin of the latter, was even more recent (0.5 Mya). In subgenus *Baleaphryne*, the mtDNA and nuclear trees suggest that the first species to diverge were *A. muletensis* and *A. maurus*, respectively, although this part of the mtDNA phylogeny is unresolved (Fig. 4). Node ages in the nuclear and mitochondrial trees are provided with their 95 % HPD in Table S4.

4. Discussion

4.1. Diversification and taxonomy in the A. obstetricans complex

Our study sheds light on several instances of phylogeographic discordance affecting the *A. obstetricans* complex (subgenus *Alytes*). First, we show that the Central Pyrenean *A. a. inigoi* represents a genuine taxon (i.e., not a ghost mitochondrial lineage) closely related to the Eastern Pyrenean *A. a. almogavarii* (Figs. 1, 4). This relationship was previously suspected by hierarchical microsatellite clustering analyses (Maia-Carvalho et al., 2018), guiding the description of *inigoi* as a subspecies of *A. almogavarii* (Dufresnes and Hernandez, 2021). Second, in Western Iberia, the northwestern and southwestern populations, both attributed to the subspecies *A. o. boscai*, were retrieved as sister lineages. In the mitochondrial tree, however, *A. almogavarii* and *A. o. boscai* are both paraphyletic (Martínez-Solano et al., 2004; Gonçalves et al., 2015; Fig. 4). It now appears clear that these taxa experienced hybridization and mitochondrial capture from *A. obstetricans* ancestors, which reset their mitochondrial divergence at two different moments of the Pleistocene (Fig. 4). Third, ABC analyses suggested that *A. o. pertinax* represents a homoploid hybrid subspecies formed by introgressive hybridization between *A. o. obstetricans* and *A. almogavarii* (Fig. 2), an event that presumably occurred during the last stages of the Pleistocene according to the weak mitochondrial divergence (Fig. 4). This striking result feeds back early genetic studies on *Alytes*: based on allozyme
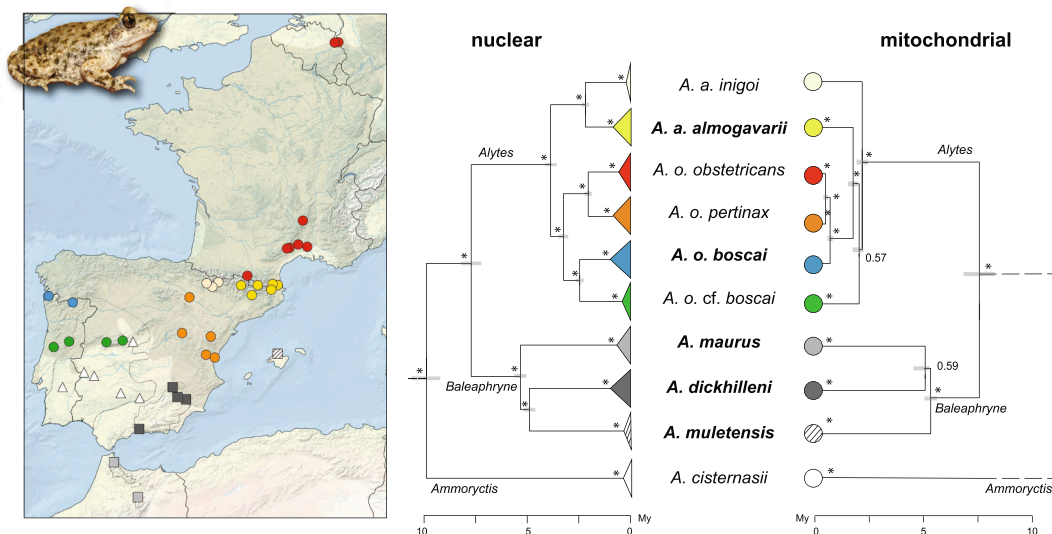

Fig. 4. Time-calibrated Bayesian phylogenies of the genus *Alytes*, based on nuclear (278,265 bp) and mitochondrial (9,045 bp) sequence alignments. The nuclear dataset consists of 45 non-admixed samples (sample set 2), considering the loci present in at least 35 of them (*n-p* = 10). Geographic origins of these samples are shown on the map (circles: subgenus *Alytes*; squares: subgenus *Baleaphryne*; triangles: subgenus *Ammoryctis*). The mitochondrial sequences were taken from Maia-Carvalho et al. (2014). Asterisks denote high posterior probabilities (>0.95). Taxa highlighted in bold feature phylogenetic position that differ between the nuclear and the mitochondrial trees. Grey bars show the 95 % HPD of the divergence time estimates.

polymorphism, García-París (1995) hypothesized the presence of *A. almogavarii* alleles as far as Central Spain (Sierra de Guadarrama), now known to be inhabited by *A. o. pertinax*. Major cyto-nuclear discordances thus appear common in refugial areas, especially in the Iberian (e.g., Renoult et al., 2009; Pöschel et al., 2018; Dufresnes et al., 2020a) and other European peninsulas (e.g., the Balkans, Wielstra and Arntzen, 2020; Dufresnes et al., 2021c), which acted both as hotspots and melting pots of genetic diversity during Quaternary climatic oscillations. Moreover, the *A. obstetricans* complex supports a prominent role of hybridization and reticulate evolution in shaping phylogeographic diversifications, as seen from the hybrid origin of an increasing number of populations/taxa revealed by multilocus studies, e.g., in toads (Dufresnes et al., 2019b; Firneno et al., 2020), snakes (Zinenko et al., 2016) and birds (Brelsford et al., 2011).

Adding to the confusion, previous nuclear analyses provided a bulk of different, weakly supported topologies in *Alytes*. For instance, ~4kb of intron sequences branched *A. a. inigoi* with the lineages assigned to *A. obstetricans* (Maia-Carvalho et al., 2014). Microsatellite analyses of northeastern Iberian populations grouped *A. a. almogavarii* and *A. a. inigoi* together but nested them within the high diversity of *A. o. obstetricans/pertinax* in this part of the range, and in a different order depending on analyses (Lucati et al., 2022). All these observations emphasize the potential unreliability of conventional markers for phylogenetic inferences, including introns when they are used in low numbers (e.g., Cruaud et al., 2014), and especially microsatellites, which are known to retain high ancestral polymorphism that confounds lineage-diagnostic alleles (van Oppen et al., 2000).

Several pending issues regarding the taxonomy of the *A. obstetricans* taxonomy can now be clarified. In their comparison of anuran amphibian hybrid zones, Dufresnes et al. (2021b) showed that speciation progresses gradually with genetic divergence, with half of the lineages studied showing signs of reproductive isolation by ~ 3 My. The nuclear divergence we estimated between *A. almogavarii* and *A. obstetricans* (~3.9 My) is thus fully consistent with their narrow area of admixture in Catalonia (Dufresnes and Martínez-Solano, 2020) and supports the decision to split them as two distinct species (Speybroeck et al., 2020). Following this rationale, *A. a. inigoi* should retain its subspecies status, as per its more recent divergence (2.2 My) and extensive gene flow with *A. a. almogavarii* (Fig. 1). In fact, we detected introgression in populations that were supposedly pure according to microsatellite analyses (Maia-Carvalho et al., 2018; Lucati et al., 2022). Even the holotype of *A. a. inigoi* featured *A. a. almogavarii* alleles (15 %), which causes a nomenclatural problem (see also the case of *Triturus arntzeni*, Wielstra and Arntzen, 2014). The International Code for Zoological Nomenclature states that "A species-group name established for an animal later found to be a hybrid [Art. 17] must not be used as the valid name for either of the parental species, even if it is older than all other available names for them." (Article 23.8) – noting that taxa of hybrid origin (e.g., hybrid or hybridogenetic species) make exceptions (Article 17.2). The availability of *A. a. inigoi* thus lies in what the Code implies by the vague term "hybrid". Because in the taxonomic context, "hybrid" generally refers to F1 offspring, we assume that partial traces of introgression do not compromise the use of this name for the Central-Pyrenean lineage. However, a neotype collected from truly pure populations of *A. a. inigoi* (e.g., further west in its range) may be designated to affirm the name to this taxon, should it be needed by future amendments of the Code. Anyhow, this case exemplifies an urgent need for clarification of the Code regarding the definition and status of hybrid specimens, especially the maximum proportion of foreign ancestry tolerated for name-bearing types to remain valid. Given the overwhelming role of hybridization in evolution, and the increased resolution offered by genomics to detect it, we predict this issue to become more frequent than currently appreciated.

With 3.2My of divergence from *A. o. obstetricans/pertinax*, the Western Iberian *A. o. boscai/cf. boscai* clade falls in the middle of the grey zone of speciation *sensu* Dufresnes et al. (2021b). Although none of

our specimens showed traces of introgression, these originate far from the area of parapatry. Using microsatellites, Maia-Carvalho et al. (2018) did report numerous populations of intermediate ancestry in Central Spain that could indicate admixture between *A. o. pertinax* and *A. o.* cf. *boscai*. In time, investigations of their hybrid zones with genomic data will inform on the taxonomic status of the Western Iberian clade, and here we provisionally maintain these populations as conspecific of *A. obstetricans*. In any case, the deep phylogenetic divergence of the southwestern Iberian lineage (*A. o.* cf. *boscai*) calls for its taxonomic description, and to assess the transition with *A. o. boscai*, which presumably extends along the Douro River in northern Portugal (Maia-Carvalho et al., 2018).

Finally, our data highlighted admixture over several hundred kilometers between *A. o. obstetricans* and *A. o. pertinax*, which is consistent with their recent origin and subspecies rank. In Cantabria, two microsatellite studies identified an endemic genetic cluster, assigned to *A. o. obstetricans* as per mtDNA barcoding (B1b in Maia-Carvalho et al., 2018; PEU in Lucati et al., 2022). From our data, however, these populations consist of *A. o. pertinax* recently admixed by *A. o. obstetricans*, and no additional structure was detected (Fig. 1, Fig. S1). The naturally high genetic variation of admixed populations, even more so if one parental taxon has a reticulate origin, combined with the potentially unstable demographic history of the Cantabrian populations during the Quaternary (generating then fixing private alleles), could have thus artefactually mimicked genetic structure in the microsatellite data.

4.2. Evolutionary relationships and biogeography in Alytes

Phylogenetic relationships among the dry-adapted subgenus *Baleaphryne* have been a matter of debate (Arntzen and García-París, 1995; Altaba, 1997; Arntzen & García-París, 1997), in part due to the remarkable morphological and behavioral differences of the Mallorcan *A. muletensis* compared to the continental species *A. maurus* (North Africa) and *A. dickhilleni* (southwestern Spain), and the initial misplacement of *A. maurus* in subgenus *Alytes* based on allozymes (Arntzen and Szymura, 1984). Our phylogenomic analyses confirmed that all three species diverged nearly simultaneously (hence the difficulty to retrieve a single topology), which is very consistent with the refilling of the Mediterranean Sea that isolated the Balearic Archipelago, Africa, and Iberia at roughly the same time (Arntzen and García-París, 1995, 1997). Furthermore, we confirm that *A. cisternasii* was the first extant species to arise, justifying its recognition as a separate subgenus, *Ammoryctis* (Dubois, 1987), but not necessarily as a separate genus (as in Dubois et al., 2021) given its relatively recent divergence (~10–16 My, Fig. 4) and close resemblance with other midwife toads in terms of shape and mating calls.

None of the *Baleaphryne* or *Ammoryctis* species show substantial phylogeographic structure despite their currently fragmented ranges (Fig. 4, see also De Pous et al., 2013 for *A. maurus*). The taxonomy of these species should thus remain stable, and the present disconnections must be relatively recent. Accordingly, projections of climatic niches identified continuous suitable areas across expected ranges (Beukema et al., 2013; Maia-Carvalho et al., 2018; Rodríguez-Rodríguez et al., 2020; Donaire-Barroso et al., 2022). Rather than phylogeographic breaks, the genetic structure retrieved by conservation genetic studies in *A. dickhilleni* (Dias et al., 2015), *A. cisternasii* (Gonçalves et al., 2009), and *A. muletensis* (Kraaijeveld-Smit et al., 2005) more likely reflect recent population subdivisions during the Holocene and the Anthropocene, potentially associated to the declines of these threatened species.

Proposed biogeographical scenarios for the diversification of *Alytes* are mostly consistent with our phylogenomic results (Arntzen & García-París, 1995; Fromhage et al., 2004; Martínez-Solano et al., 2004). The early split of *A. cisternasii* was tentatively attributed to the opening of the Betic sea strait (Betic crisis), which isolated southern from northern Iberia around 16–14 Mya, connecting the former to Africa (Arntzen and

García-París, 1995). This hypothesis initially relied on allozyme- and mitochondrial-based phylogenies, where the split of *A. cisternasii* was dated around 13 and 16 My, respectively (Maia-Carvalho et al., 2014), in line with the rich fossil record of this epoch (Bastir et al., 2014). Our time-calibrated nuclear tree points to a more recent origin for this species, namely around ~ 10 Mya, which can be attributed to two alternative geoclimatic events (Ehl et al., 2019): the accretion of the Rift-Betic Cordillera some 10 Mya (Lonergan and White, 1997), or the fragmentation of the insular Betic region from 12 to 6 Mya (momentarily connecting the region to Africa). The latter event could also explain the split between ancestral *Baleaphryne* (presumably in Africa) and *Alytes* (presumably in northern Iberia), which we dated to 7.7 Mya. *Baleaphryne* would have later spread across Iberia and the Balearic Islands during the Messinian Salinity Crisis (5.96–5.33 Mya) and evolved in vicariance when the Gibraltar Strait reopened. Subgenus *Alytes* later diversified throughout northern Iberia during the Plio-Pleistocene, presumably as the climate cooled down and promoted a typical refugia-within-refugia dynamic (Gómez and Lunt, 2007; Abellán and Svenning, 2014).

4.3. Phylogeographic analyses with RAD-seq

The genus *Alytes* makes the case of how genomic phylogeography can help settling on the evolution and taxonomy of species complexes, especially when the patterns of structure and diversity could not be comprehensively retrieved by limited sets of mtDNA and nuclear loci due to limited phylogenetic resolution combined with pervasive historical and present hybridization (see also e.g., Firneno et al., 2021). Nevertheless, by comparing the outcomes of numerous phylogenetic datasets generated from our RAD-seq library, we also highlighted that the results remained sensitive to the sampling scheme, and most importantly, the number of loci included in respect to the proportion of missing data (Fig. 3). While RAD-seq analyses can answer diverse questions in molecular ecology (Davey and Blaxter, 2010; Peterson et al., 2012), their use in phylogeography still lacks a proper practical framework (Shafer et al., 2017), and some of the lessons learnt here might be helpful to design future surveys.

A notorious issue in phylogenetic analyses is the inclusion of unexpectedly introgressed samples. These may branch with intermediate positions between their parental clades, thus giving false impressions of divergence, and can further skew the topology of the tree by affecting the reconstruction of ancestral state characters (e.g., Schilling et al., 2018; Wang et al., 2019; Appelhans et al., 2020; Firneno et al., 2020). The issue is exemplified by the analyses of our sample set 1 in Fig. 3, which includes *A. a. almogavarii/inigoi* and *A. o. obstetricans/pertinax* recently admixed individuals, and where the correct topology was robustly retrieved only in larger alignments/SNP matrices compared to the other sample sets. A key step in any phylogenetic study is to select samples representative of the populations of interest that should be free of recent admixture, i.e., located far away from transition zones. When dealing with species complexes however, the targeted lineages are usually young, their genomes have remained mostly compatible, and depending on opportunities for secondary contacts, they may admix over hundreds of kilometers (e.g., Dufresnes et al., 2019a; Rivera et al., 2022). In this respect, we have shown how conventional barcoding information may be misleading to choose reference populations, both due to cyto-nuclear discordance and the difficulty to quantify admixture from few loci that were not designed for lineage-diagnosticity (e.g., microsatellites). We thus strongly encourage to first exploit genomic datasets to verify the nature of samples by population genetic analyses, before including them in phylogenomic inferences.

In our comparative phylogenetic analyses, the size of the dataset, which expectedly increased exponentially with the amount of locus dropout allowed by bioinformatic filters (e.g., Wagner et al., 2013), was a major factor affecting the topology (Fig. 3). Nearly complete but comparatively small supermatrices (<150 kb of concatenated RAD tags)

retrieved alternative trees, which were often robust in the BI analyses (see e.g., Leaché et al., 2015). Because these trees appear well constructed, there is a risk to retain an erroneous phylogeny if longer alignments are not considered due to their larger amounts of missing data. The influence of missing data on phylogenetic reconstruction has been intensively discussed in the past (e.g., Wiens, 2006; Rubin et al., 2012; Huang and Knowles, 2016; Crotti et al., 2019; Cerca et al., 2021), but it seems marginal for genomic datasets, where the number of polymorphic sites is always high (Eaton et al., 2017; Huang and Knowles, 2016; Rancilhac et al., 2019). For instance, RAD-seq studies on plants (Karbstein et al., 2020) and fishes (Tripp et al., 2017) have shown that adequate phylogenetic resolution can be obtained even with extreme amounts of missing data (respectively 70 and 90 %). Accordingly, our results were similar with (sample set 2) and without (sample set 3) individuals that bear outlier proportions of missing data, noting that these remain in relatively low proportion overall (<20 %), even with the most lenient thresholds (Fig. S4). We thus recommend exploring several filtering combinations when building phylogenies from RAD-seq sequences, both to maximize the number of included loci and ensure that results hold across datasets.

The phylogenetic method chosen may also be a source of discordances between RAD-seq trees (Rancilhac et al., 2019; Widhelm et al., 2019). Here, all three supermatrix-based analyses converged on a single topology with sufficiently large datasets (≥150 kb) (Fig. 3). The inferences were only inconclusive for the multispecies coalescent (Fig. S5), potentially because of the comparatively much lower number of loci informative in SNAPP – where any loci not sequenced in all pre-defined groups are automatically discarded. Obtaining enough RAD-seq loci conserved across a whole diversification can be challenging due to locus dropout, which increases with lineage divergence as mutations affect enzyme restriction sites. While the present genomic data was seemingly too limiting for SNAPP (≤630 SNPs shared across *Alytes* lineages), the approach performed appropriately in other anurans where at least three times more markers could be obtained with the same RAD-seq protocol (e.g., 3,157 SNPs in *Rana*, Dufresnes et al., 2020a, 2,448 and 9,097 SNPs in *Discoglossus* and *Pelodytes*, respectively, Dufresnes et al., 2020c, 2,302 SNPs in *Hyla/Dryophytes*, Dufresnes et al., 2021d). Depending on the prerequisites of analyses, obtaining enough loci to reach robust phylogenomic inferences may thus still remain challenging even with genomic data.

RAD-seq was originally developed for intraspecific population genetics (Baird et al., 2008), and the implementation of this approach in multispecies phylogenies is not universally popular (Leaché et al., 2015; McKain et al., 2018; Hühn et al., 2022). Of particular concern is the reliability to retrieve homologous loci across diverged species, and in sufficient numbers for meaningful results (Paris et al., 2017). As such, some recent phylogeographic studies use alternative methods of genome reduction, such as target enrichment (Mamanova et al., 2010). Despite these concerns, methodological surveys suggest that key bioinformatic parameters involved in the building of RAD tag assemblies have limited effects on downstream phylogenetic analyses (Clugston et al., 2019; Scornavacca et al., 2020). As we show, the effect of SNP calling parameters also becomes negligible as long as large datasets are retrieved (see also Rancilhac et al., 2019). RAD-seq thus remains a flexible technique for phylogeographic analyses, with limited costs and optimization efforts.

While benefiting from an unprecedented resolution, species trees built from genomic datasets may be particularly sensitive to historical gene flow (Leaché et al., 2014a; Jiao et al., 2020). On the one hand, the detection of fine levels of introgression and cyto-nuclear discordance contributes empirical evidence that many young taxa have regularly experienced hybridization during their progressive differentiation (Cairns et al., 2021; Yang et al., 2021). These findings have been changing our perception of gene flow in speciation – from an antagonistic force to an inevitable phenomenon promoting diversification through reinforcement, adaptive introgression (Feder et al., 2012) and

even the rise of new taxa (e.g., *A. o. pertinax* here). On the other hand, the fact that many extant species carry foreign alleles may skew tree topologies, perhaps in a similar way as contemporary hybrids (Fig. 3). Yet, historical admixture is not as trivial to detect (e.g., Caeiro-Dias et al., 2021) and there is no straightforward approach to account for gene flow in phylogenomics. Available methods typically compare gene trees to the species tree, but gene trees cannot be inferred from short DNA sequences such as *de novo* assembled RAD tags. Besides, the mitochondrial tree is typically used as the species tree, which can be deceitful when cyto-nuclear discordance occurs. Inferences more appropriate for genomic SNPs, such as site frequency spectrum or Approximate Bayesian Computation (ABC) are restricted to test specific hypotheses involving a few lineages (as we did here for the origin of *A. o. pertinax*; see also Firneno et al., 2020). Phylogenetic relationships among species-rich diversifications might thus remain difficult to settle upon even in the genomic era (Caeiro-Dias et al., 2021), and midwife toads will be worth reinvestigating with complementary approaches, such as transcriptomes or long-read genome assemblies.

CRediT authorship contribution statement

Johanna Ambu: Methodology, Formal analysis, Investigation, Resources, Writing – original draft, Visualization. **Íñigo Martínez-Solano:** Resources, Writing – review & editing. **Tomasz Suchan:** Formal analysis, Resources, Writing – review & editing, Visualization. **Axel Hernandez:** Resources, Writing – review & editing. **Ben Wielstra:** Resources, Writing – review & editing. **Pierre-André Crochet:** Resources, Writing – review & editing. **Christophe Dufresnes:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107783.

References

Abellán, P., Svenning, J.C., 2014. Refugia within refugia–patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. Biol. J. Linn. Soc. 113 (1), 13–28. https://doi.org/10.1111/bij.12309.

Altaba, C.R., 1997. Phylogeny and biogeography of midwife toads (*Alytes, Discoglossidae*): a reappraisal. Bijdragen tot de Dierkunde 66 (4), 257–262. https://doi.org/10.1163/26660644-06604006.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17 (2), 81–92. https://doi.org/10.1038/nrg.2015.28.

Appelhans, M.S., Paetzold, C., Wood, K.R., Wagner, W.L., 2020. RADseq resolves the phylogeny of Hawaiian *Myrsine* (Primulaceae) and provides evidence for hybridization. J. Syst. Evol. 58 (6), 823–840. https://doi.org/10.1111/jse.12668.

Arntzen, J.W., 2022. The Midwife Toad Challenge After (Half) a Century. Front. Ecol. Evol. 5 https://doi.org/10.3389/fevo.2022.819542.

Arntzen, J.W., García-París, M., 1995. Morphological and allozyme studies of midwife toads (genus *Alytes*), including the description of two new taxa from Spain. Bijdragen tot de Dierkunde 65 (1), 5–34. https://doi.org/10.1163/26660644-06501002.

Arntzen, J.W., García-París, M., 1997. Phylogeny and biogeography of midwife toads (*Alytes*, Discoglossidae): a rebuttal. Bijdragen tot de Dierkunde 66 (4), 263–268. https://doi.org/10.1163/26660644-06604007.

Arntzen, J.W., Szymura, J.M., 1984. Genetic differentiation between African and European midwife toads (*Alytes*, Discoglossidae). Bijdragen tot de Dierkunde 54 (2), 157–162. https://doi.org/10.1163/26660644-05402001.

Avise, J., 1994. Molecular Markers, Natural History and Evolution. Springer, New York, NY. 14–511. https://doi.org/10.1007/978-1-4615-2381-9.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., Fay, J.C., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3 (10), e3376. https://doi.org/10.1371/journal.pone.0003376.

Baldassarre, D.T., White, T.A., Karubian, J., Webster, M.S., 2014. Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution 68, 2644–2657. https://doi.org/10.1111/evo.12457.

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13 (4), 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x.

Barrios, V., Olmeda, C., Ruiz, E., Ayres, C., Bentata, V., Böll, S., Borgula, A., Bosch, J., Bosman, W., Duguet, R., Goffart, P., Kordges, T., Schley, L., Schmidt, B., Teixeira, J., Uthleb, H., & Verbist, V. (2012). Action Plan for the Conservation of the Common Midwife Toad (*Alytes obstetricans*) in the European Union. European Commission. EU Species Action Plan. The N2K Group.

Bastir, M., Böhme, M., Sanchiz, B., 2014. Middle Miocene remains of Alytes (Anura, Alytidae) as an example of the unrecognized value of fossil fragments for evolutionary morphology studies. J. Vertebr. Paleontol. 34 (1), 69–79. https://doi.org/10.1080/02724634.2013.794813.

Beukema, W., De Pous, P., Donaire-Barroso, D., Bogaerts, S., Garcia-Porta, J., Escoriza, D., Arribas, O.J., El Mouden, E.H., Carranza, S., 2013. Review of the systematics, distribution, biogeography and natural history of Moroccan amphibians. Zootaxa 3661 (1), 1–60. https://doi.org/10.11646/zootaxa.3661.1.1.

Bonnet, T., Leblois, R., Rousset, F., Crochet, P.-A., 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution 71 (9), 2140–2158. https://doi.org/10.1111/evo.13296.

Bosch, J., Martínez-Solano, I., García-París, M., 2001. Evidence of a chytrid fungus infection involved in the decline of the common midwife toad (*Alytes obstetricans*) in protected areas of central Spain. Biol. Conserv. 97 (3), 331–337.

Bosch, B., Mora-Cabello de Alba, A., Marquínez, S., Price, S.J., Thumsová, B., Bielby, J., 2021. Long-term monitoring of amphibian populations of a national park in Northern Spain reveals negative persisting effects of *Ranavirus*, but not *Batrachochytrium dendrobatidis*. Frontiers in Veterinary Science 8, 659. https://doi.org/10.3389/fvets.2021.645491.

Bouckaert, R., Heled, J., 2014. DensiTree 2: Seeing trees through the forest. bioRxiv, 012401 https://doi.org/10.1101/012401.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., Pertea, M., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15 (4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Brandt, A.L., Ishida, Y., Georgiadis, N.J., Roca, A.L., 2012. Forest elephant mitochondrial genomes reveal that elephantid diversification in Africa tracked climate transitions. Mol. Ecol. 21 (5), 1175–1189. https://doi.org/10.1111/j.1365-294X.2012.05461.x.

Brelsford, A., Mila, B., Irwin, D.E., 2011. Hybrid origin of Audubon's warbler. Mol. Ecol. 20 (11), 2380–2389. https://doi.org/10.1111/j.1365-294X.2011.05055.x.

Brelsford, A., Dufresnes, C., Perrin, N., 2016. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. Heredity 116 (2), 177–181. https://doi.org/10.1038/hdy.2015.83.

Caeiro-Dias, G., Rocha, S., Couto, A., Pereira, C., Brelsford, A., Crochet, P.-A., Pinho, C., 2021. Nuclear phylogenies and genomics of a contact zone establish the species rank of *Podarcis lusitanicus* (Squamata, Lacertidae). Mol. Phylogenet. Evol. 164, 107270 https://doi.org/10.1016/j.ympev.2021.107270.

Cairns, N.A., Cicchino, A.S., Stewart, K.A., Austin, J.D., Lougheed, S.C., 2021. Cytonuclear discordance, reticulation and cryptic diversity in one of North America's most common frogs. Mol. Phylogenet. Evol. 156, 107042 https://doi.org/10.1016/j.ympev.2020.107042.

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. https://doi.org/10.1111/mec.12354.

Cerca, J., Maurstad, M.F., Rochette, N.C., Rivera-Colón, A.G., Rayamajhi, N., Catchen, J.M., Struck, T.H., Freckleton, R., 2021. Removing the bad apples: A simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. Methods Ecol. Evol. 12 (5), 805–817.

Chan, K.O., Hertwig, S.T., Neokleous, D.N., Flury, J.M., Brown, R.M., 2022. Widely used, short 16S rRNA mitochondrial gene fragments yield poor and erratic results in phylogenetic estimation and species delimitation of amphibians. BMC Ecology and Evolution 22 (1), 1–9. https://doi.org/10.1186/s12862-022-01994-y.

Chenuil, A., 2006. Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects. Genetica 127 (1), 101–120. https://doi.org/10.1007/s10709-005-2485-1.

Clugston, J.A., Kenicer, G.J., Milne, R., Overcast, I., Wilson, T.C., Nagalingum, N.S., 2019. RADseq as a valuable tool for plants with large genomes—A case study in cycads. Mol. Ecol. Resour. 19 (6), 1610–1622. https://doi.org/10.1111/1755-0998.13085.

Cornuet, J., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J., Estoup, A., 2014. DYIABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30 (8), 1187–1189. https://doi.org/10.1093/bioinformatics/btt763.

Crotti, M., Barratt, C.D., Loader, S.P., Gower, D.J., Streicher, J.W., 2019. Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). Zool. Scr. 48 (2), 157–167. https://doi.org/10.1111/zsc.12335.

Cruaud, A., Gautier, M., Galan, M., Foucaud, J., Sauné, L., Genson, G., Dubois, E., Nidelet, S., Deuve, T., Rasplus, J.Y., 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Mol. Biol. Evol. 31 (5), 1272–1274. https://doi.org/10.1093/molbev/msu063.

Currat, M., Ruedi, M., Petit, R.J., Excoffier, L., 2008. The hidden side of invasions: massive introgression by local genes. *Evolution: International Journal of Organic*. Evolution 62 (8), 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x.

Dai, C., Wang, W., Lei, F., 2013. Multilocus phylogeography (mitochondrial, autosomal and Z-chromosomal loci) and genetic consequence of long-distance male dispersal in Black-throated tits (*Aegithalos concinnus*). Heredity 110 (5), 457–465. https://doi.org/10.1038/hdy.2012.114.

Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9 (5–6), 416–423. https://doi.org/10.1093/bfgp/elq031.

De Pous, P., Metallinou, M., Donaire-Barroso, D., Carranza, S., Sanuy, D., 2013. Integrating mtDNA analyses and ecological niche modelling to infer the evolutionary history of *Alytes maurus* (Amphibia; Alytidae) from Morocco. The Herpetological Journal 23 (3), 153–160.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009.

Dias, G., Beltrán, J.F., Tejedo, M., Benítez, M., Miras, E.G., Ferrand, N., Gonçalves, H., 2015. Limited gene flow and high genetic diversity in the threatened Betic midwife toad (*Alytes dickhilleni*): evolutionary and conservation implications. Conserv. Genet. 16 (2), 459–476. https://doi.org/10.1007/s10592-014-0672-2.

Donaire-Barroso, D., Hernandez, A., Escoriza, D., 2022. Shedding light on a threatened midwife toad: fine-scale niche modelling of the Moroccan endemic *Alytes maurus* (Pasteur and Bons 1962). Afr. J. Herpetol. 71 (1), 83–93. https://doi.org/10.1080/21564574.2021.2014989.

Dubois, A., 1987. Miscellanea taxinomica batrachologica (I). *Alytes*. Paris 5, 7–95.

Dubois, A., Ohler, A., Pyron, R.A., 2021. New concepts and methods for phylogenetic taxonomy and nomenclature in zoology, exemplified by a new ranked cladonomy of recent amphibians (Lissamphibia). Megataxa 5 (1), 1–738. https://doi.org/10.11646/megataxa.5.1.1.

Dufresnes, A., Hernandez, A., 2021. Phylogeographic advances in midwife toads (*Alytes*) support the existence of a novel taxon endemic to the Central Pyrenees. J. Zool. Syst. Evol. Res. 59 (8), 2170–2179. https://doi.org/10.1111/jzs.12564.

Dufresnes, C., Mazepa, G., Rodrigues, N., Brelsford, A., Litvinchuk, S.N., Sermier, R., Lavanchy, G., Betto-Colliard, C., Blaser, O., Borzée, A., Cavoto, E., Fabre, G., Ghali, K., Grossen, C., Horn, A., Leuenberger, J., Philips, B.C., Saunders, P.A., Savary, R., Maddalena, T., Stöck, M., Dubey, S., Canestrelli, D., Jeffries, D.L., 2018. Genomic evidence for cryptic speciation in tree frogs from the Apennine Peninsula, with description of Hyla perrini sp. nov. Front. Ecol. Evol. 144. https://doi.org/10.3389/fevo.2018.00144.

Dufresnes, C., Strachinis, I., Tzoras, E., Litvinchuk, S.N., Denoël, M., 2019a. Call a spade a spade: taxonomy and distribution of *Pelobates*, with description of a new Balkan endemic. ZooKeys 859, 131. https://doi.org/10.3897/zookeys.859.33634.

Dufresnes, C., Mazepa, G., Jablonski, D., Oliveira, R.C., Wenseleers, T., Shabanov, D.A., Auer, M., Ernst, R., Koch, C., Ramírez-Chaves, H.E., Mulder, K.P., Simonov, E., Tiutenko, A., Kryvokhyzha, D., Wennekes, P.L., Zinenko, O.I., Korshunov, O.V., Al-Johany, A.M., Peregontsev, E.A., Masroor, R., Betto-Colliard, C., Denoël, M., Borkin, L.J., Skorinov, D.V., Pasynkova, R.A., Mazanaeva, L.F., Rosanov, J.M., Dubey, S., Litvinchuk, S., 2019b. Fifteen shades of green: the evolution of *Bufotes* toads revisited. Mol. Phylogenet. Evol. 141, 106615.

Dufresnes, C., Nicieza, A.G., Litvinchuk, S.N., Rodrigues, N., Jeffries, D.L., Vences, M., Perrin, N., Martínez-Solano, Í., 2020a. Are glacial refugia hotspots of speciation and cytonuclear discordances? Answers from the genomic phylogeography of Spanish common frogs. Mol. Ecol. 29 (5), 986–1000. https://doi.org/10.1111/mec.15368.

Dufresnes, C., Litvinchuk, S.N., Rozenblut-Kościsty, B., Rodrigues, N., Perrin, N., Crochet, P.-A., Jeffries, D.L., 2020b. Hybridization and introgression between toads with different sex chromosome systems. Evol. Lett. 4 (5), 444–456.

Dufresnes, C., Pribille, M., Alard, B., Gonçalves, H., Amat, F., Crochet, P.-A., Dubey, S., Perrin, N., Fumagalli, L., Vences, M., Martínez-Solano, I., 2020c. Integrating hybrid zone analyses in species delimitation: lessons from two anuran radiations of the Western Mediterranean. Heredity 124 (3), 423–438. https://doi.org/10.1038/s41437-020-0294-z.

Dufresnes, C., Suchan, T., Smirnov, N.A., Denoël, M., Rosanov, J.M., Litvinchuk, S.N., 2021a. Revisiting a speciation classic: Comparative analyses support sharp but leaky

transitions between *Bombina* toads. J. Biogeogr. 48 (3), 548–560. https://doi.org/10.1111/jbi.14018.

Dufresnes, C., Brelsford, A., Jeffries, D.L., Mazepa, G., Suchan, T., Canestrelli, D., Nicieza, A., Fumagalli, L., Dubey, S., Martínez-Solano, Í., Litvinchuk, S., Vences, M., Perrin, N., Crochet, P.-A., 2021b. Mass of genes rather than master genes underlie the genomic architecture of amphibian speciation. Proc. Natl. Acad. Sci. 118 (36), e2103963118 https://doi.org/10.1073/pnas.2103963118.

Dufresnes, C., Probonas, N.M., Strachinis, I., 2021c. A reassessment of the diversity of green toads (Bufotes) in the circum-Aegean region. Integrative Zoology 16 (3), 420–428. https://doi.org/10.1111/1749-4877.12494.

Dufresnes, C., Brelsford, A., Baier, F., Perrin, N., Parsch, J., 2021d. When sex chromosomes recombine only in the heterogametic sex: heterochiasmy and heterogamety in *Hyla* tree frogs. Mol. Biol. Evol. 38 (1), 192–200.

Dufresnes, C., Martínez-Solano, Í., 2020. Hybrid zone genomics supports candidate species in Iberian *Alytes obstetricans*. Amphibia-Reptilia 41 (1), 105–112. https://doi.org/10.1163/15685381-20191312.

Earl, D.A., VonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4 (2), 359–361. https://doi.org/10.1007/s12686-011-9548-7.

Eaton, D.A., Spriggs, E.L., Park, B., Donoghue, M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. 66 (3), 399–412. https://doi.org/10.1093/sysbio/syw092.

Edwards, S.V., Shultz, A.J., Campbell-Staton, S.C., 2015. Next-generation sequencing and the expanding domain of phylogeography. Folia Zool. 64 (3), 187–206. https://doi.org/10.25225/fozo.v64.i3.a2.2015.

Edwards, S.V., Potter, S., Schmitt, C.J., Bragg, J.G., Moritz, C., 2016. Reticulation, divergence, and the phylogeography–phylogenetics continuum. Proc. Natl. Acad. Sci. 113 (29), 8025–8032. https://doi.org/10.1073/pnas.1601066113.

Ehl, S., Vences, M., Veith, M., 2019. Reconstructing evolution at the community level: a case study on Mediterranean amphibians. Mol. Phylogenet. Evol. 134, 211–225. https://doi.org/10.1016/j.ympev.2019.02.013.

Feder, J., Egan, S., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28, 342–350. https://doi.org/10.1016/j.tig.2012.03.009.

Firneno Jr, T.J., O'Neill, J.R., Portik, D.M., Emery, A.H., Townsend, J.H., Fujita, M.K., 2020. Finding complexity in complexes: assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. Mol. Ecol. 29 (18), 3543–3559. https://doi.org/10.1111/mec.15496.

Firneno Jr, T.J., O'Neill, J., Itgen, M.W., Kihneman, T.A., Townsend, J.H., Fujita, M.K., 2021. Delimitation despite discordance: Evaluating the species limits of a confounding species complex in the face of mitonuclear discordance. Ecol. Evol. 11 (18), 12739–12753.

Fromhage, L., Vences, M., Veith, M., 2004. Testing alternative vicariance scenarios in Western Mediterranean discoglossid frogs. Mol. Phylogenet. Evol. 31 (1), 308–322. https://doi.org/10.1016/j.ympev.2003.07.009.

Garcia-Cisneros, A., Palacín, C., Ventura, C.R.R., Feital, B., Paiva, P.C., Pérez-Portela, R., 2018. Intraspecific genetic structure, divergence and high rates of clonality in an amphi-Atlantic starfish. Mol. Ecol. 27 (3), 752–772. https://doi.org/10.1111/mec.14454.

García-París, M., 1995. Variabilidad genética y distribución geográfica de *Alytes obstetricans almogavarii* en España. Revista Española de Herpetología 9, 133–138.

García-París, M., Arntzen, J.W., 2002. *Alytes dickhilleni* (Arntzen and García-París 1995). In: Pleguezuelos, J.M., Márquez, R., Lizana, M. (Eds.), : *Atlas y Libro Rojo de los Anfibios y Reptiles de España*. Dirección General de Conservación de la Naturaleza, Asociación Herpetológica Española, Madrid, pp. 76–78.

Gómez, A., Lunt, D.H., 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Phylogeography of Southern European Refugia. Springer, Dordrecht, pp. 155–188.

Gonçalves, H., Martínez-Solano, Í., Ferrand, N., García-París, M., 2007. Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, *Alytes*): deep coalescence or ancestral hybridization? Mol. Phylogenet. Evol. 44 (1), 494–500. https://doi.org/10.1016/j.ympev.2007.03.001.

Gonçalves, H., Martínez-Solano, Í., Pereira, R.J., Carvalho, B., García-París, M., Ferrand, N., 2009. High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent, multiple refugia: evidence from mtDNA, microsatellites and nuclear genealogies. Mol. Ecol. 18 (24), 5143–5160. https://doi.org/10.1111/j.1365-294X.2009.04426.x.

Gonçalves, H., Maia-Carvalho, B., Sousa-Neves, T., García-París, M., Sequeira, F., Ferrand, N., Martínez-Solano, Í., 2015. Multilocus phylogeography of the common midwife toad, *Alytes obstetricans* (Anura, Alytidae): contrasting patterns of lineage diversification and genetic structure in the Iberian refugium. Mol. Phylogenet. Evol. 93, 363–379. https://doi.org/10.1016/j.ympev.2015.08.009.

Goudet, J., 2005. Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5 (1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Gouy, M., Tannier, E., Comte, N., Parsons, D.P., 2021. Seaview version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. Methods Mol. Biol. 2231, 241–2610. https://doi.org/10.1007/978-1-0716-1036-7_15.

Griffiths, R. A., Garcia, G., & Oliver, J. (2008). Re-introduction of the Mallorcan midwife toad, Mallorca, Spain. Pp 54-57 in: *Global re-introduction perspectives: reintroduction case-studies from around the globe*. IUCN/SSC Re-introduction Specialist Group, Abu Dhabi.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. PhyML 3.0: new algorithms, methods and utilities. Syst. Biol. 59 (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270 (1512), 313–321.

Hodel, R.G., Chen, S., Payton, A.C., McDaniel, S.F., Soltis, P., Soltis, D.E., 2017. Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Sci. Rep. 7 (1), 1–14. https://doi.org/10.1038/s41598-017-16810-7.

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

Hühn, P., Dillenberger, M.S., Gerschwitz-Eidt, M., Hörandl, E., Los, J.A., Messerschmid, T.F., Paetzold, C., Rieger, B., Kadereit, G., 2022. How challenging RADseq data turned out to favor coalescent-based species tree inference. A case study in *Aichryson* (Crassulaceae). Mol. Phylogenet. Evol. 167, 107342 https://doi.org/10.1016/j.ympev.2021.107342.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23 (2), 254–267. https://doi.org/10.1093/molbev/msj030.

IUCN SSC Amphibian Specialist Group. (2022). *Alytes obstetricans*. The IUCN Red List of Threatened Species 2022: e.T178591974A89699462. https://dx.doi.org/10.2305/IUCN.UK.2022-1.RLTS.T178591974A89699462.en.

Jiao, X., Flouri, T., Rannala, B., Yang, Z., Foster, P., 2020. The impact of cross-species gene flow on species tree estimation. Syst. Biol. 69 (5), 830–847.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24 (11), 1403–1405.

Karbstein, K., Tomasello, S., Hodač, L., Dunkel, F.G., Daubert, M., Hörandl, E., 2020. Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (Ranunculaceae). Taxon 69 (6), 1191–1220. https://doi.org/10.1002/tax.12365.

Kimball, R.T., Guido, M., Hosner, P.A., Braun, E.L., 2021. When good mitochondria go bad: Cyto-nuclear discordance in landfowl (Aves: Galliformes). Gene 801, 145841. https://doi.org/10.1016/j.gene.2021.145841.

Kraaijeveld-Smit, F.J.L., Beebee, T.J.C., Griffiths, R.A., Moore, R.D., Schley, L., 2005. Low gene flow but high genetic diversity in the threatened Mallorcan midwife toad *Alytes muletensis*. Mol. Ecol. 14 (11), 3307–3315. https://doi.org/10.1111/j.1365-294X.2005.02614.x.

Kuhl, H., Frankl-Vilches, C., Bakker, A., Mayr, G., Nikolaus, G., Boerno, S.T., Klages, S., Timmermann, B., Gahr, M., Battistuzzi, F.U., 2021. An unbiased molecular approach using 3′-UTRs resolves the avian family-level tree of life. Mol. Biol. Evol. 38 (1), 108–127.

Lange, L., Brischoux, F., Lourdais, O., 2022. Benefits of paternal thermoregulation: male midwife toads select warmer temperature to shorten embryonic development. Behav. Ecol. Sociobiol. 76 (4), 1–10. https://doi.org/10.1007/s00265-022-03155-z.

Leaché, A. D., & Bouckaert, R. R. (2018). Species trees and species delimitation with SNAPP: a tutorial and worked example. Available at: http://evomicsorg.wpeng ine.netdna-cdn.com/wpcon tent/uploads/2018/01/BFD-tutor ial-1.pdf.

Leaché, A.D., Harris, R.B., Rannala, B., Yang, Z., 2014a. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63 (1), 17–30. https://doi.org/10.1093/sysbio/syt049.

Leaché, A.D., Fujita, M.W., Minin, V.N., Bouckaert, R.R., 2014b. Species delimitation using genome-wide SNP data. Syst. Biol. 63 (4), 534–542. https://doi.org/10.1093/sysbio/syu018.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D., Linkem, C.W., 2015. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biol. Evol. 7 (3), 706–719. https://doi.org/10.1093/gbe/evv026.

Lee-Yaw, J.A., Grassa, C.J., Joly, S., Andrew, R.L., Rieseberg, L.H., 2019. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytol. 221 (1), 515–526. https://doi.org/10.1111/nph.15386.

Lonergan, L., White, N., 1997. Origin of the Betic-Rif mountain belt. Tectonics 16 (3), 504–522. https://doi.org/10.1029/96TC03937.

Lucati, F., Miró, A., Bosch, J., Caner, J., Jowers, M.J., Rivera, X., Donaire-Barroso, D., Ventura, M., 2022. Contrasting patterns of genetic admixture explain the phylogeographic history of Iberian high mountain populations of midwife toads. bioRxiv. https://doi.org/10.1101/2022.05.12.490531.

Maia-Carvalho, B., Gonçalves, H., Ferrand, N., Martínez-Solano, Í., 2014. Multilocus assessment of phylogenetic relationships in *Alytes* (Anura, Alytidae). Mol. Phylogenet. Evol. 79, 270–278. https://doi.org/10.1016/j.ympev.2014.05.033.

Maia-Carvalho, B., Vale, C.G., Sequeira, F., Ferrand, N., Martínez-Solano, Í., Goncalves, H., 2018. The roles of allopatric fragmentation and niche divergence in intraspecific lineage diversification in the common midwife toad (*Alytes obstetricans*). J. Biogeogr. 45 (9), 2146–2158. https://doi.org/10.1111/jbi.13405.

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., Takezaki, N., 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. Mol. Biol. Evol. 35 (5), 1284–1290.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. Nat. Methods 7 (2), 111–118. https://doi.org/10.1038/nmeth.1419.

Martínez-Gil, H., Martínez-Freiría, F., Perera, A., Enriquez-Urzelai, U., Martínez-Solano, Í., Velo-Antón, G., Kaliontzopoulou, A., 2022. Morphological diversification of Mediterranean anurans: the roles of evolutionary history and climate. Biol. J. Linn. Soc. 135 (3), 462–477.

Martínez-Solano, Í., Gonçalves, H.A., Arntzen, J.W., García-París, M., 2004. Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). J. Biogeogr. 31 (4), 603–618. https://doi.org/10.1046/j.1365-2699.2003.01033.x.

McKain, M.R., Johnson, M.G., Uribe-Convers, S., Eaton, D., Yang, Y., 2018. Practical considerations for plant phylogenomics. Appl. Plant Sci. 6 (3), e1038 https://doi.org/10.1002/aps3.1038.

Mongiardino Koch, N., Coppard, S.E., Lessios, H.A., Briggs, D.E., Mooi, R., Rouse, G.W., 2018. A phylogenomic resolution of the sea urchin tree of life. BMC Evol. Biol. 18 (1), 1–18. https://doi.org/10.1186/s12862-018-1300-4.

Moore, R.D., Griffiths, R.A., Román, A., 2004. Distribution of the Mallorcan midwife toad (*Alytes muletensis*) in relation to landscape topography and introduced predators. Biol. Conserv. 116 (3), 327–332. https://doi.org/10.1016/S0006-3207(03)00202-7.

Morales-Briones, D.F., Liston, A., Tank, D.C., 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). New Phytol. 218 (4), 1668–1684. https://doi.org/10.1111/nph.15099.

Pabijan, M., Crottini, A., Reckwell, D., Irisarri, I., Hauswaldt, J.S., Vences, M., 2012. A multigene species tree for Western Mediterranean painted frogs (*Discoglossus*). Mol. Phylogenet. Evol. 64 (3), 690–696. https://doi.org/10.1016/j.ympev.2012.05.017.

Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. Methods Ecol. Evol. 8 (10), 1360–1373. https://doi.org/10.1111/2041-210X.12775.

Penna, M., Velásquez, N.A., Bosch, J., 2015. Dissimilarities in auditory tuning in midwife toads of the genus *Alytes* (Amphibia: Anura). Biol. J. Linn. Soc. 116 (1), 41–51. https://doi.org/10.1111/bij.12563.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., Orlando, L., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7 (5), e37135. https://doi.org/10.1371/journal.pone.0037135.

Pöschel, J., Heltai, B., Graciá, E., Quintana, M.F., Velo-Antón, G., Arribas, O., Valdeón, A., Wink, M., Fritz, U., Vamberger, M., 2018. Complex hybridization patterns in European pond turtles (Emys orbicularis) in the Pyrenean Region. Sci. Rep. 8 (1) https://doi.org/10.1038/s41598-018-34178-0.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155 (2), 945–959. https://doi.org/10.1093/genetics/155.2.945.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. https://doi.org/10.1093/sysbio/syy032.

Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M.R., Elmer, K.R., Vences, M., Steinfartz, S., 2019. Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADSeq data analysis. Mol. Phylogenet. Evol. 133, 189–197. https://doi.org/10.1016/j.ympev.2019.01.003.

Recuero, E., Buckley, D., García-París, M., Arntzen, J.W., Cogălniceanu, D., Martínez-Solano, I., 2014. Evolutionary history of *Ichthyosaura alpestris* (Caudata, Salamandridae) inferred from the combined analysis of nuclear and mitochondrial markers. Mol. Phylogenet. Evol. 81, 207–220. https://doi.org/10.1016/j.ympev.2014.09.014.

Renoult, J.P., Geniez, P., Bacquet, P., Benoit, L., Crochet, P.-A., 2009. Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. Mol. Ecol. 18 (20), 4298–4315. https://doi.org/10.1111/j.1365-294X.2009.04351.x.

Rivera, D., Prates, I., Firneno Jr, T.J., Trefaut Rodrigues, M., Caldwell, J.P., Fujita, M.K., 2022. Phylogenomics, introgression, and demographic history of South American true toads (*Rhinella*). Mol. Ecol. 31 (3), 978–992. https://doi.org/10.1111/mec.16280.

Rodríguez-Rodríguez, E.J., Beltrán, J.F., Tejedo, M., Nicieza, A.G., Llusia, D., Márquez, R., Aragón, P., 2020. Niche models at inter-and intraspecific levels reveal hierarchical niche differentiation in midwife toads. Sci. Rep. 10 (1), 1–11. https://doi.org/10.1038/s41598-020-67992-6.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., Kolokotronis, S.-O., 2012. Inferring phylogenies from RAD sequence data. PLoS One 7 (4), e33394. https://doi.org/10.1371/journal.pone.0033394.

Rubinoff, D., Holland, B.S., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54 (6), 952–961.

Schilling, M.P., Gompert, Z., Li, F.W., Windham, M.D., Wolf, P.G., 2018. Admixture, evolution, and variation in reproductive isolation in the *Boechera puberula* clade. BMC Evol. Biol. 18 (1), 1–14. https://doi.org/10.1186/s12862-018-1173-6.

Scornavacca, C., Delsuc, F., & Galtier, N. (2020). *Phylogenetics in the genomic era*. 568 pp.

Seixas, F.A., Boursot, P., Melo-Ferreira, J., 2018. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. Genome Biol. 19 (1), 1–20. https://doi.org/10.1186/s13059-018-1471-8.

Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., Wolf, J.B.W., Gilbert, M., 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol. Evol. 8 (8), 907–917.

Speybroeck, J., Beukema, W., Dufresnes, C., Fritz, U., Jablonski, D., Lymberakis, P., Martínez-Solano, Í., Razzetti, E., Vamberger, M., Vences, M., Vörös, J., Crochet, P.-A., 2020. Species list of the European herpetofauna–2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica. Amphibia-Reptilia 41 (2), 139–189. https://doi.org/10.1163/15685381-bja10010.

Toews, D.P., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21 (16), 3907–3930. https://doi.org/10.1111/j.1365-294x.2012.05664.x.

Tripp, E.A., Tsai, Y.H.E., Zhuang, Y., Dexter, K.G., 2017. RAD seq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. Ecol. Evol. 7 (19), 7920–7936. https://doi.org/10.1002/ece3.3274.

van Oppen, M.J., Rico, C., Turner, G.F., Hewitt, G.M., 2000. Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. Mol. Biol. Evol. 17 (4), 489–498. https://doi.org/10.1093/oxfordjournals.molbev.a026329.

Vliegenthart, C., van de Vrede, M., den Boer, I., Gilbert, M.J., Lemmers, P., France, J., de Visser, M.C., Struijk, R.P.J.H., Wielstra, B., 2023. The limits of mtDNA analysis for determining the provenance of invasive species: a midwife toad example. Amphib.-Reptilia 44 (1), 27–33.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22 (3), 787–798. https://doi.org/10.1111/mec.12023.

Wang, D., Wang, Z., Kang, X., Zhang, J., 2019. Genetic analysis of admixture and hybrid patterns of *Populus hopeiensis* and *P. tomentosa*. Sci. Rep. 9 (1), 1–13. https://doi.org/10.1038/s41598-019-41320-z.

Widhelm, T.J., Grewe, F., Huang, J.P., Mercado-Díaz, J.A., Goffinet, B., Lücking, R., Moncada, B., Mason-Gamer, R., Lumbsch, H.T., 2019. Multiple historical processes obscure phylogenetic relationships in a taxonomically difficult group (Lobariaceae, *Ascomycota*). Sci. Rep. 9 (1), 1–16. https://doi.org/10.1038/s41598-019-45455-x.

Wielstra, B., 2019. Historical hybrid zone movement: More pervasive than appreciated. J. Biogeogr. 46 (7), 1300–1305. https://doi.org/10.1111/jbi.13600.

Wielstra, B., Arntzen, J.W., 2014. Kicking *Triturus arntzeni* when it's down: large-scale nuclear genetic data confirm that newts from the type locality are genetically admixed. Zootaxa 3802 (3), 381–388. https://doi.org/10.11646/zootaxa.3682.3.5.

Wielstra, B., Arntzen, J.W., 2020. Extensive cytonuclear discordance in a crested newt from the Balkan Peninsula glacial refugium. Biol. J. Linn. Soc. 130 (3), 578–585. https://doi.org/10.1093/biolinnean/blaa062.

Wiens, J.J., 2006. Missing data and the design of phylogenetic analyses. J. Biomed. Inform. 39 (1), 34–42. https://doi.org/10.1016/j.jbi.2005.04.001.

Yang, W., Feiner, N., Pinho, C., While, G.M., Kaliontzopoulou, A., Harris, D.J., Salvi, D., Uller, T., 2021. Extensive introgression and mosaic genomes of Mediterranean endemic lizards. Nat. Commun. 12, 2762. https://doi.org/10.1038/s41467-021-22949-9.

Zardi, G.I., Nicastro, K.R., McQuaid, C.D., Castilho, R., Costa, J., Serrão, E.A., Pearson, G.A., 2015. Intraspecific genetic lineages of a marine mussel show behavioural divergence and spatial segregation over a tropical/subtropical biogeographic transition. BMC Evol. Biol. 15 (1), 1–11. https://doi.org/10.1186/s12862-015-0366-5.

Zieliński, P., Nadachowska-Brzyska, K., Wielstra, B., Szkotak, R., Covaciu-Marcov, S.D., Cogălniceanu, D., Babik, W., 2013. No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). Mol. Ecol. 22 (7), 1884–1903. https://doi.org/10.1111/mec.12225.

Zinenko, O., Sovic, M., Joger, U., Gibbs, H.L., 2016. Hybrid origin of European vipers (*Vipera magnifica* and *Vipera orlovi*) from the Caucasus determined using genomic scale DNA markers. BMC Evol. Biol. 16, 76. https://doi.org/10.1186/s12862-016-0647-7.

Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. 17 (9), 2107–2121. https://doi.org/10.1111/j.1365-294X.2008.03737.x.