RESEARCH ARTICLE

Piecing the barcoding puzzle of Palearctic water frogs (Pelophylax) sheds light on amphibian biogeography and global invasions

Christophe Dufresnes1,2 | Benjamin Monod-Broca3 | Adriana Bellati4 | Daniele Canestrelli4 | Johanna Ambu1 | Ben Wielstra5,6 | Sylvain Dubey7 | Pierre-André Crochet8 | Mathieu Denoël9 | Daniel Jablonski10

1Laboratory of Amphibian Systematics and Evolutionary Research (LASER), College of Biology and the Environment, Nanjing Forestry University, Nanjing, People's Republic of China
2Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France
3Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, Villeurbanne, France
4Department of Ecological and Biological Sciences, University of Tuscia, Viterbo, Italy
5Instituto de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France
6Department of Biology, Leiden University, Leiden, The Netherlands
7Naturalis Biodiversity Center, Leiden, The Netherlands
8Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
9CEFE, CNRS, Univ Montpellier, EPHE,IRD, Montpellier, France
10Laboratory of Ecology and Conservation of Amphibians (LECA), FOCUS, University of Liège, Liège, Belgium

Funding information
National Natural Science Foundation of China, Grant/Award Number: 3211101356; Fonds De La Recherche Scientifique—FNSR, Grant/Award Number: T.0070.19

Abstract
Palearctic water frogs (genus Pelophylax) are an outstanding model in ecology and evolution, being widespread, speciose, either threatened or threatening to other species through biological invasions, and capable of siring hybrid offspring that escape the rules of sexual reproduction. Despite half a century of genetic research and hundreds of publications, the diversity, systematics and biogeography of Pelophylax still remain highly confusing, in no small part due to a lack of correspondence between studies. To provide a comprehensive overview, we gathered >13,000 sequences of barcoding genes from >1700 native and introduced localities and built multigene mitochondrial (~17 kb) and nuclear (~10 kb) phylogenies. We mapped all currently recognized taxa and their phylogeographic lineages (>40) to get a grasp on taxonomic issues, cyto-nuclear discordances, the genetic makeup of hybridogenetic hybrids, and the origins of introduced populations. Competing hypotheses for the molecular calibration were evaluated through plausibility tests, implementing a new approach relying on predictions from the anuran speciation continuum. Based on our timetree, we propose a
new biogeographic paradigm for the Palearctic since the Paleogene, notably by attributing a prominent role to the dynamics of the Paratethys, a vast paleo-sea that extended over most of Europe. Furthermore, our results show that distinct marsh frog lineages from Eastern Europe, the Balkans, the Near East, and Central Asia (P. ridibundus ssp.) are naturally capable of inducing hybridogenesis with pool frogs (P. lessonae). We identified 14 alien lineages (mostly of P. ridibundus) over ~20 areas of invasions, especially in Western Europe, with genetic signatures disproportionately pointing to the Balkans and Anatolia as the regions of origins, in line with exporting records of the frog leg industry and the stocks of pet sellers. Pelophylax thus emerges as one of the most invasive amphibians worldwide, and deserves much higher conservation concern than currently given by the authorities fighting biological invasions.

**KEYWORDS**
DNA barcoding, hybridogenesis, invasive alien species, Paratethys, phylogeography, timetree plausibility test

## INTRODUCTION

Amphibians are excellent models to study how climatic and geological processes have shaped past and present patterns of biodiversity, and how these patterns are now being affected by human activities. In the Palearctic realm (Europe, North Africa, Central and Far East Asia), the genetic diversity of numerous frogs, toads and salamanders has been intensively studied at various taxonomic and geographic scales, progressively enriching molecular databases with genetic barcodes (e.g., Velo-Antón et al., 2023). In widespread and complex groups, piecing this “barcoding puzzle” can offer exciting prospects, providing reliable temporal and spatial inferences that reconcile heterogeneous yet complementary datasets (e.g., Dufresnes et al., 2022; Dufresnes & Hernández, 2023; Li et al., 2015; Pramuk et al., 2008; Yuan et al., 2016).

While the temporal aspect of phylogeographic frameworks has become crucial to their interpretation (Avise, 2009), it often strikes as their weakest spot (Hipsley & Müller, 2014). Time trees are inherently affected by uncertainties in tree topologies and the calibration of molecular clocks, which are both insufficiently acknowledged sources of errors in the narrative of biogeographic studies. With these issues in mind, one way to assess the credibility of time trees is to perform biogeographical plausibility tests (BPT), by matching node ages against sets of geo-climatic events that could a priori explain lineage vicariance (Ehl et al., 2019; Fromhage et al., 2004).

Most amphibians are sensitive to climate cooling (their range margins typically follow isotherms of overwintering temperatures, e.g., Bulakhova et al., 2020) and do not cross marine barriers unless under exceptional circumstances (e.g., Vences et al., 2003). Geological evidence for climatic oscillations, marine transgressions/regressions and tectonic events that presumably triggered land changes (insularity, landmass connections, orogenesis and desertification) thus offer great opportunities for BPT, in complement to the fossil record (Hipsley & Müller, 2014).

Hints on the pace of speciation may also serve to gauge phylogenetic timeframes. When speciation is primarily driven by allopatric divergence, reproductive isolation evolves gradually through postzygotic genomic incompatibilities, and patterns of hybridization and admixture across secondary contact zones are largely affected by the age of the lineages involved (Dufresnes et al., 2021; Dufresnes & Crochet, 2022). Matching time trees against the expected timeframe of speciation in a given organismal group could thus provide a test to further evaluate different calibrations, which we coin the "lineage-compatibility plausibility test" (LCPT).

The spatial component of phylogeographic frameworks finds applications beyond the field of biogeography, being particularly informative for the study of biological invasions. Amphibian translocations are increasingly reported worldwide (in link to the pet trade, as pest control, garden ornaments, or as food, medicine or research resources), generating a high socioeconomic cost when introduced species become a threat to autochthonous wildlife (Measey et al., 2020; Soto et al., 2022). Establishing the identity and origin of non-native amphibians is instrumental to inform wildlife managers and conservation scientists (Dubey, Lavanchy, et al., 2019; Koster et al., 2023; Kuijt et al., 2023), but it first requires a priori knowledge of their native diversity, which is rarely trivial (Robbemont et al., 2023). For example, infamous global invaders such as the cane toad (Rhinella marina), the clawed frog (Xenopus laevis) or the Asian common toad (Duttaphrynus melanostictus) all belong to species complexes composed of multiple lineages in their home range (Furman et al., 2015; Rivera et al., 2022; Wogan et al., 2016). Hence, invasion genetics and historical phylogeography must go hand in hand to understand the onset of biological invasions.

Widely distributed from northwestern Africa to Japan, Palearctic water frogs of the genus Pelophylax offer great opportunities for large-scale studies in biogeography (Akin et al., 2010; Komaki et al., 2015; Lymberakis et al., 2007; Nicolas et al., 2015) and invasion biology (Denoël et al., 2022). From a molecular perspective,
Pelophylax may be the amphibian genus that has been the most studied in Eurasia over the past 60 years (Dufresnes & Mazepa, 2020). This acute interest initially stemmed from the discovery that some taxa have a hybrid origin and perpetuate themselves through hybridogenesis, a peculiar reproductive mode where F1 individuals discard one parental genome and transmit the other clonally, giving rise to stable “lineages” of hemiclonal hybrids (Lavanchy & Schwander, 2019). Hybridogenesis then offers rare opportunities to explore the boundaries of sexual reproduction on the edge of the speciation continuum (Berger, 1967; Dufresnes & Mazepa, 2020; Graf & Polls, 1989; Tunner, 1974).

Genetic research on Pelophylax later expanded to explore phylogeographic patterns, address taxonomic issues in species complexes (e.g., P. bedriagae) in Western and Central Asia, Akin et al., 2010; P. saharicus in the Maghreb, Nicolas et al., 2015; P. nigromaculatus and P. planci in East Asia, Komaki et al., 2015; Zhou et al., 2023), and tentatively map species assemblages where phenotypic variation challenges morphological field identification (e.g., Pagano et al., 2001; Vucic et al., 2018). More recently, molecular surveys aimed at tracing emerging biological invasions, notably by the marsh frogs (P. ridibundus complex), following uncontrolled introductions linked in part to the international trade for the frog leg industry (e.g., Bellati et al., 2023; Dufresnes et al., 2018; Dufresnes & Dubey, 2020; Holsbeek et al., 2008). Today, Pelophylax invaders are becoming a strong conservation concern, due to their broad habitat tolerance (Denoel et al., 2022), potential performances under future climate change (Padilla et al., 2023), fierce predation on native batrachofauna (Pille et al., 2021) and invertebrates (Pille et al., 2023), potential vectors of amphibian diseases such as chytridiomycosis (Balaz et al., 2014), and the risk to deregulate autochthonous hybridogenetic systems through hybridization and competition (Holsbeek & Jooris, 2010; Quilodran et al., 2015).

Decades of evolutionary, ecological, taxonomic, conservation and invasion research on Pelophylax have promoted extensive yet uncoordinated genetic barcoding efforts. As a consequence, a proper agreement concerning the level of resolution, diversity and distribution of the genus remains genuinely lacking. Mitochondrial DNA (mtDNA) has been preferentially sequenced to identify and map phylogeographic lineages, but mtDNA sequences are scattered across more than a hundred publications that vary in taxonomic scales, geographic scopes, sets of genes analyzed, and further differ in the way these lineages are delimited, labeled and taxonomically treated (e.g., Akin et al., 2010; Bellati et al., 2023; Bulbul et al., 2011; Dufresnes, Di Santo, et al., 2017; Hoffmann et al., 2015; Komaki et al., 2015; Liu et al., 2010; Lumberakos et al., 2007; Min et al., 2008; Nicolas et al., 2015; Papezlik et al., 2023; Plotner et al., 2008; Svinin et al., 2021; Toli et al., 2023; Ualiev et al., 2022; Wang et al., 2017). Moreover, mitochondrial barcoding may not be reliable due to hybridization, that cause mitochondrial transfer between sympatric and parapatric species (e.g., Doniol-Valcroze et al., 2021; Komaki et al., 2015; Liu et al., 2010), including the parental species of hybridogenetic hybrids (Plotner et al., 2008).

The nuclear analyses conducted in parallel primarily aimed at distinguishing hybrid forms from their parental species, and most datasets are not combinable per design, for example, microsatellites (e.g., Dufresnes, Di Santo, et al., 2017; Papezlik et al., 2021; Sagonas et al., 2020), allozymes (e.g., Hotz et al., 2013), and RAD-sequencing (e.g., Doniol-Valcroze et al., 2021; Dubey, Maddalena, et al., 2019; Dufresnes & Dubey, 2020). The only re-usable nuclear markers are targeted gene sequences (e.g., Cuevas et al., 2022), but these also differ between geographic areas and species complexes, and each gene provides little phylogenetic resolution if not combined in multilocus analyses (e.g., Komaki et al., 2015). Therefore, basic questions that pertain to nuclear DNA, such as the phylogeographic origin(s) of distinct hybridogenetic systems, remain largely unanswered (Dubey & Dufresnes, 2017; Dufresnes & Mazepa, 2020).

More generally, previous phylogenetic studies based on various mitochondrial and nuclear datasets reached conflicting conclusions regarding species relationships and species ages (see e.g., Dubey, Maddalena, et al., 2019). Discrepancies among time trees further stem from strikingly different pre-defined calibration points. In particular, a Messinian origin (~5.3 Mya) was proposed for either the Aegean populations (P. (ridibundus) cerys; Beerli et al., 1996, based on a pioneer protein clock), the Cyprian population (P. (ridibundus) cypriensis; Plotner et al., 2012), the Cretan population (P. cretensis; Lumberakis et al., 2007), or the North-African populations (P. saharicus; Nicolas et al., 2015), despite the fact that these Mediterranean taxa did not diverge simultaneously (Ehl et al., 2019).

Incorporating the various sequence archives available for Pelophylax into a comprehensive meta-dataset can offer a global view on the biogeographical history of the genus, on the extent and diversity of human-mediated translocations, the diversity of lineages able to trigger hybridogenesis, and many other aspects related to the taxonomy and evolution of the genus. Water frogs are also particularly suitable for timetree plausibility tests. Their widespread distribution warrants many paleographic candidate events of vicariance for the BPT, both in the western (Ehl et al., 2019) and the eastern (Dufresnes & Litvinchuk, 2022) regions of the Palearctic. In addition, the outcome of hybridization between taxa has been relatively well-documented, both regarding hybridogenetic mechanisms (reviewed by Dufresnes & Mazepa, 2020) and the contact zones between several phylogeographic lineages (e.g., Dufresnes & Dubey, 2020; Hotz et al., 2013; Nagai et al., 2018; Papezlik et al., 2021; Sagonas et al., 2020; Santucci et al., 1996), thus allowing to experiment the use of the LCPT for the first time.

To provide a comprehensive overview of the genus’ evolution, diversity and distribution, we screened 131 sources (complemented by new data) to build a meta-dataset of 13,619 sequences from 1765 georeferenced localities, spanning commonly sequenced mtDNA and nuclear genes. Specifically, we aimed at mapping all species and their intraspecific lineages, reconstructing multilocus phylogenies, applying the BPT and LCPT, reassessing the lineages involved in hybridogenesis, and exploring the origins of allochthonous populations to appraise the emerging problem of water frog invasions.
2 | METHODS

2.1 | Mitochondrial barcoding

Mitochondrial barcoding information was compiled for the following six markers (and their adjacent tRNAs) abundantly available in *Pelophylax*: 16S ribosomal RNA (16S), cytochrome c oxidase subunit 1 (COX1), NADH-dehydrogenase subunit 1 (ND1), subunit 2 (ND2), subunit 3 (ND3) and cytochrome b (cyt-b). Publicly accessible sequences were harvested from GenBank or directly from the original literature as of August 2023, complemented by new sequences of 16S, COX1 and ND2, generated as in previous work (16S: Dufresnes & Hernandez, 2023; COX1: Doniol-Valcroze et al., 2021; ND2: Plötner et al., 2008). All sequences were manually aligned, trimmed, and re-arranged in Seaview 5.0 (Gouy et al., 2021). Unique haplotypes were screened and given standardized identifiers. For each gene, databases were constructed following the approach detailed in Dufresnes and Hernandez (2023). Briefly, this consists of assigning each bar-coded sequence to its corresponding haplotype and to the metadata available in the original publications and on GenBank (geographic origin, voucher specimen, sample name, original haplotype label). The information was cross-checked between sources, for example, GenBank entries, original publications (and their supporting material)—including those re-using the same samples—to avoid duplicates and to correct discrepancies. Sequences with missing data, of short length, or of bad sequencing quality or with seemingly chimeric composition, were discarded. For each gene, the diversity was visualized with phylogenetic networks in SplitsTree 4.18.3 (Huson & Bryant, 2006). Mitogroups, i.e., sets of closely related mitochondrial haplotypes, were defined from the divergence patterns at each gene and labelled according to the oldest taxonomic names available as given by Frost (2023), and/or with letters for species complexes. Mitogroup correspondence between markers was facilitated by the availability of full or fragmented mitogenomes for the majority of lineages, and by cross-examining geographic distributions. Shallow mitogroups could not always be distinguished across all six genes due to their different phylogenetic resolution. When sequences were not released, mitogroups were inferred from the corresponding samples based on the geographic and phylogenetic information provided in the original publication. Altogether, the mitochondrial barcoding data spanned 11,030 sequences taken from 125 sources and representative of 1695 localities (Table 1).

2.2 | Nuclear barcoding

Nuclear barcoding information combined publicly available sequences with new sequences from eight genes used in previous phylogenetic studies on the genus. In the Eastern Paleartctic (EP), we gathered the six genes sequenced in most *Pelophylax* species and their various intraspecific lineages by Komaki et al. (2015): cellular myelocytomatosis intron 2 (C-MYC), β-fibrinogen intron 7 (βFib), mesoderm-specific transcript intron 3 (MEST), rhodopsin intron 3 (Rhô), gene 9 intron 2 of the sex-determining region Y box (sox9) and tyrosinase (Tyr). The Tyr dataset was completed by the numerous sequences published by Liu et al. (2010) for two Asian mainland species. In the Western Paleartic (WP), we essentially obtained sequences from C-MYC, βFib and Tyr, as well as two additional genes: recombination activating gene intron 1 (RAG1) and serum albumin intron 1 (SAI-1). New sequences from C-MYC, βFib, SAI1 and Tyr were generated as in previous work (βFib and Tyr: Komaki et al., 2015; C-MYC and SAI-1: Dubey & Dufresnes, 2017). Heterozygous sequences (i.e., with ambiguity bases) were phased using PHASE 2.1 under default settings (Stephens et al., 2001). Haploype reconstructions were performed separately for each gene and subsets of samples. For each gene, databases was then built as for mtDNA, namely by gathering and cross-verifying sequence metadata, identifying unique haplotypes and haplogroups with phylogenetic networks, and labelling them according to the species and/or mitogroup they belong to, when relevant. The parental alleles of hybridogenetic hybrids were considered in the dataset when they could be properly phased by cloning (original publications) or by unambiguous assignments to haplotypes identified in the parental species. Altogether, the nuclear barcoding data spanned 2589 sequences taken from 24 sources representative of 364 localities (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>16S</th>
<th>COX1</th>
<th>ND1</th>
<th>ND2</th>
<th>ND3</th>
<th>cyt-b</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sequences</td>
<td>535</td>
<td>684</td>
<td>206</td>
<td>3155</td>
<td>2573</td>
<td>3877</td>
<td>11,030</td>
</tr>
<tr>
<td>Georeferenced sequences</td>
<td>493</td>
<td>639</td>
<td>192</td>
<td>3141</td>
<td>2558</td>
<td>3165</td>
<td>10,188</td>
</tr>
<tr>
<td>Localities</td>
<td>213</td>
<td>265</td>
<td>93</td>
<td>694</td>
<td>466</td>
<td>625</td>
<td>1695</td>
</tr>
<tr>
<td>Sources</td>
<td>40</td>
<td>36</td>
<td>17</td>
<td>34</td>
<td>31</td>
<td>60</td>
<td>125</td>
</tr>
</tbody>
</table>

**TABLE 1** Number of sequences available for each of the six mitochondrial genes analyzed for barcoding, the number of localities they originate from, and the number of contributing sources (published work and/or GenBank archives).
properly aligned (see also Dufresnes & Hernandez, 2023). Two phylogenetic analyses were performed. First, a maximum-likelihood tree reconstruction was obtained with PhyML 3.2 (Guindon et al., 2010), using the smart model selection algorithm (Lefort et al., 2017) and testing for branch support by 100 bootstrap replicates. The analysis was run on a dedicated PhyML server (http://www.atgc-montpellier.fr/phyml/). Second, a Bayesian tree reconstruction was performed in BEAST 2.6 (Bouckaert et al., 2019) in which the tRNAs, the 15 genes and the control region were partitioned to independently estimate parameters of sequence evolution (GTR+G+I) and mutation rates (strict clocks). The birth–death tree prior was chosen as a compromise between deep interspecific and shallow intraspecific lineage diversifications. The chain was run for 50 million iterations (sampling every 5000) and monitored for stationarity in Tracer 1.7.2. A maximum clade credibility tree (MCCT) was exported with the TreeAnnotator module, discarding the first 20% as burnin.

For the nuclear phylogeny, we built a supermatrix of 134 sequences representing 67 diploid individuals, spanning the eight genes gathered for barcoding (~9.8 kb). The sampling differed between the two Palearctic clades, EP lineages being well-sequenced at the six genes analyzed by Komaki et al. (2015), while WP lineages were more scarcely sequenced at five genes, two of which were virtually unavailable in the EP (RAG1 and SAI-1). In an effort to reduce the amount of missing data, the supermatrix included sequences that concatenated genes obtained from different specimens of the same lineage (as identified in the previous section). Like for the mitochondrial phylogeny, we first analyzed our nuclear supermatrix with PhyML and BEAST (partitioning the eight genes) to obtain single trees for the whole genus. However, contrary to mtDNA, where all loci are fully linked and represent the same genealogy, nuclear gene trees are expected to differ from one another. Supermatrix-based reconstructions that assume linkage across genes might thus be improper, with the most polymorphic genes biasing the topology. To circumvent this issue, we also performed multispecies coalescent analyses with *BEAST (Heled & Drummond, 2010), where gene trees are estimated independently to infer the species tree. Because this approach is sensitive to the representativeness of each taxon at each gene, separate analyses were conducted for the EP and WP species, for which different sequences sets were available. The EP subset consisted of 66 sequences representing 33 individuals (including one WP outgroup), spanning C-MYC, βFib, MEST, Rho, sox9 and Tyr (all taken from Komaki et al., 2015). The WP subset consisted of 72 sequences representing 36 individuals (including one EP outgroup), spanning C-MYC, βFib and Tyr, RAG1 and SAI-1. Sequences were pre-assigned to 9 (EP) and 23 (WP) populations corresponding to identified phylogeographic lineages (see Section 3). The birth–death tree prior was chosen for the species tree, and the chains were run for 100 million iterations (sampling every 5000). Stationarity was monitored in Tracer and the results were visualized in Densitree (Bouckaert & Heled, 2014), discarding the first 20% as burnin.

### 2.4 Time tree inferences

Because the nuclear phylogeny remains unsolved for many young lineages (see Section 3), only the mitochondrial tree was used to infer time trees. Four analyses were conducted in BEAST using most recent common ancestor (MRCA) priors according to competing hypotheses of the Mediterranean taxa that originated during the Messinian Salinity Crisis (MSC), namely P. saharicus (Nicolas et al., 2015), P. cretensis (Lymbenakis et al., 2007), P. (ridibundus) cypriensis (Plötner et al., 2012) or P. (ridibundus) cerigensis (Beerli et al., 1999). Fossil evidence indicates that the genus had already dispersed across Eurasia by the early Oligocene (Paleogene zone MP22, 30.9–32.5 Mya; Lemierre et al., 2023; Rage & Roček, 2003), suggesting that its MRCA is at least this old. In all time trees, a first calibration was thus enforced by a log-normal MRCA prior of all Pelophylax sequences, with offset = 30, M = 3 and S = 1. The second calibration (MSC split) was enforced by a normally distributed MRCA prior of mean = 5.3 and σ = 0.1 for either: (A) the split between P. (ridibundus) cypriensis and all the Near Eastern and Middle Eastern mitogroups; (B) the split between P. cretensis and the P. ridibundus complex; (C) the split between P. saharicus and P. perezi; (D) the split between P. (ridibundus) cerigensis and closely related Anatolian mitogroups. Each analysis was run for at least 20 million iterations after the chain reached stationarity (sampling every 1000 iterations) and summarized in a time-annotated MCCT.

To evaluate the most realistic timeframes, we first set up a BPT following Fromhage et al. (2004) and Ehl et al. (2019), by matching each node age against relevant paleogeographic events potentially affecting the diversification (Table 3). Second, we implemented the LCPT as a new approach that considers the patterns of hybridization and admixture between species/lineages predicted from their divergence, based on comparative hybrid zone analyses in European anurans (Dufresnes et al., 2021). Expectations include the progressive reduction of gene flow across hybrid zones between 2 My (wide admixture) and 6 My of divergence (restricted admixture), with virtually no gene flow for parapatric lineages greater than 10 My (Dufresnes et al., 2021), and only F1-like hybrids still
being potentially produced between sympatric species of up to 15–20 My of divergence (e.g., allopolyploids in *Bufotes*, Dufresnes, Mazepa, et al., 2019; sterile hybrids in *Hyla*, Drillon et al., 2019). To apply the LCPT, we reviewed documented patterns of hybridization in *Pelophylax*, notably across hybrid zones (Table 4), and assessed whether the ages of the involved lineages were consistent with the anuran timeframe of speciation.

Two additional aspects were considered to discuss the biogeographic scenario inferred from our mitochondrial time tree. First, we gathered fossil occurrences from fosFARbase (Böhme & Ilg, 2003) and mapped them for different geological periods between the Oligocene and the Miocene. Second, we identified cyto-nuclear discordances based on the resolved parts of our nuclear phylogeny as well as the phylogenomic trees published for *P. saharicus* (Donioli-Valcroze et al., 2021) and other WP species (Dubey, Maddalena, et al., 2019).

### 2.5 Origins of *Pelophylax* invasions

Based on lineage distributions, we screened for unnoticed *Pelophylax* translocations and reassessed known invasions in a comparative framework. Introduced populations, involving different taxa, have been surveyed using mitochondrial barcodes in Italy (Bellati et al., 2019, 2023; Bisconti et al., 2019; Bruni et al., 2020; Domeneghetti et al., 2013), Malta (Papežík et al., 2024), France...
TABLE 4  Pairs of Pelophylax lineages for which pattern of hybridization and admixture were documented in the wild and that could be used for the lineage-compatibility plausibility test.

<table>
<thead>
<tr>
<th>Lineage pair</th>
<th>Pattern of hybridization/admixture</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>lessonae × ridibundus</td>
<td>F1 hybridogenetic hybrids (&quot;esculentus&quot;) in most of Europe</td>
<td>Dufresnes &amp; Mazepa (2020), and references therein</td>
</tr>
<tr>
<td>bergeri × ridibundus</td>
<td>F1 hybridogenetic hybrids (&quot;hispanicus&quot;) in Italy</td>
<td>Dufresnes &amp; Mazepa (2020), and references therein</td>
</tr>
<tr>
<td>perezi × ridibundus</td>
<td>F1 hybridogenetic hybrids (&quot;grafi&quot;) in S-France and N-Iberia</td>
<td>Dufresnes &amp; Mazepa (2020), and references therein</td>
</tr>
<tr>
<td>epiroticus × kurtmueller</td>
<td>Occasional F1 hybrids and perhaps (exceptionally) F2 and backcrosses in Greece</td>
<td>Hotz et al. (2013), Sagonas et al. (2020), Sofianidou (1996)</td>
</tr>
<tr>
<td>bergeri A × bergeri B</td>
<td>Admixture over ~200km in S-Italy</td>
<td>Santucci et al. (1996)</td>
</tr>
<tr>
<td>bergeri A × lessonae</td>
<td>Admixture all over Switzerland</td>
<td>Dufresnes, Di Santo, et al. (2017), Dufresnes &amp; Dubey (2020)</td>
</tr>
<tr>
<td>brevipodus A × brevipodus B</td>
<td>Admixture over ~100km in Central Japan</td>
<td>Nagai et al. (2018)</td>
</tr>
<tr>
<td>cf. ridibundus D/F × ridibundus/ kurtmueller</td>
<td>Admixture over ~200km in NE-Greece</td>
<td>Hotz et al. (2013), Sagonas et al. (2020)</td>
</tr>
</tbody>
</table>

Note: While Pelophylax grafi hybrids bear perezi and ridibundus genomes, these two taxa may never have hybridized and P. grafi could have resulted from hybridization between P. perezi and P. esculentus.

3 | RESULTS

3.1 | Mitochondrial barcoding and phylogeny

We recovered, mapped, and inferred the phylogeny of 41 distinct Pelophylax mitogroups, including 27 in the WP (Figures 1 and 2) and 14 in the EP (Figure 3). The taxonomic arrangements followed hereafter (which mostly correspond to recent accounts) are detailed in Table S1.

In the WP, the tree is divided into five major clades, as follows:

1. A clade composed of P. perezi (SW-Europe) and P. saharicus, including different mitogroups between the Western, Central and Eastern Maghreb for the latter.
2. A clade composed of the pool frog taxa P. shajipericus (W-Balkans), P. lessonae lessonae (most of Europe) and P. lessonae bergeri (Apennine Peninsula), including different mitogroups in mainland Italy and Sicily for the latter.
3. A monotypic clade composed of P. epiroticus (SW-Balkans).
4. A monotypic clade composed of P. cretensis (Crete).
5. A clade with 17 lineages regrouped as the marsh frog P. ridibundus, including the subspecies P. ridibundus ridibundus (most of Europe), P. ridibundus kurtmuelleri (Balkans), P. ridibundus bedriagae (Levant), P. ridibundus cypriensis (Cyprus), and a subclade of 13 shallow lineages noted P. ridibundus cf. ridibundus A–M, some corresponding to described taxa, namely, P. ridibundus cerigensis (lineage E; SW-Anatolian coast and offshore islands), P. ridibundus caralitanus (lineage B; S-Anatolia), P. ridibundus terentievii (lineage M; Uzbekistan) and P. ridibundus persicus (lineage A; Iran). Their complex distributions are detailed in Figure 2, noting that some closely related mitogroups are not distinguished by some markers, i.e., ridibundus/kurtmuelleri for 16S; cf. ridibundus D/F/G for cyt-b and ND3.

In the EP, the tree is divided into four major clades as follows:

1. A monotypic clade composed of P. mongolius (Central China).
2. A monotypic clade composed of P. fukienensis, with distinct mitogroups in Taiwan and mainland China.
3. A Japanese clade composed of P. porosus porosus (NE-Honshu) and P. porosus brevipodus (rest of Honshu), with distinct brevipodus mitogroups between SW and E-Honshu (referred to as the "Nagoya" and "Okayama" forms in the literature).
FIGURE 1. Phylogenetic relationships and distribution of *Pelophylax* mitogroups in the Western Palearctic (WP). Taxonomic names are given, when available (Table S1). The map combines barcoding information from six genes; shaded areas: presumably native distributions (some eastern ranges may reflect old introductions; Kuzmin, 1999); arrows: introduced populations genotyped. Inset maps zoom in the Balkan Peninsula (top left), Kamchatka Peninsula (top right) and Azores (bottom left). Mitogroup distributions in *Pelophylax ridibundus* are detailed in Figure 2. The tree shows the Bayesian analysis of full and partial mitogenomes (~16.8 kb); see Figure S1 for terminal branches and for the Maximum-Likelihood analysis; node circles illustrate branch support; the grey line (bottom of the tree) indicates the placement of *saharicus* B based on nuclear phylogenomics (Doniol-Valcroze et al., 2021; see also Figure 4). Native ranges were adapted from IUCN Red list (2023) modified according to occurrence data from GBIF (www.gbif.org) and MNHN (https://inpn.mnhn.fr).
4. A widespread clade regrouping eight mitogroups that belong to the *P. nigromaculatus* and *P. plancyi* complexes (noted A–G). Their taxonomic identity is blurred by rampant cyto-nuclear discordances (see below) and high lineage diversity, with some taxa matching more than one mitogroup (e.g., *nigromaculatus*: A, B, F, H; *reinhardtii*: C and G; *plancyi*: E and G) and several taxa belonging to the same mitogroup (G: *chosenicus*, *plancyi* and *reinhardtii*).

The Bayesian and Maximum-Likelihood inferences are largely congruent (Figure S1). Both provided a robust topology, except for the position of *P. ridibundus bedriagae* relative to *P. ridibundus ridibundus/kurtmuelleri*, the position of *P. fukienensis* relative to *P. mongoliensis*, and for the shallowest *P. ridibundus* cf. *ridibundus* and *P. nigromaculatus/plancyi* lineages (Figure S1).

#### 3.2 | Nuclear barcoding and phylogeny

Thirty two nuclear lineages were retrieved, including 23 in the WP (Figures 4 and 5) and 9 in the EP (Figure 6). Most lineages correspond to previously identified mitogroups and share similar distributions (Figures 1–3; Table S1) but a few closely related mitogroups are not distinguished by the nuclear genes.

In contrast to mtDNA, phylogenetic relationships are not robustly resolved. The WP species tree obtained with the multispecies coalescent distinguishes the main clades and provide a generally consistent topology compared to the mitochondrial tree, but with little support for the position of *P. shapifericus*, *P. epeiroticus* and *P. cretensis*, as well as for the *P. ridibundus* lineages (Figure 4). The EP species tree is generally more robust, but uncertainties remain regarding the position of *P. plancyi* and *P. fukienensis* (Figure 6). The supermatrix trees obtained by Bayesian and Maximum-Likelihood inferences propose alternative, although not better supported topologies within the marsh frog clade (Figure S2). In the EP, both analyses robustly recover *P. fukienensis* and *P. plancyi* as sister clades (Figure S2). However, the supermatrix nuclear trees (Figure S2) should be interpreted with caution, as they improperly assume linkage among sequences from different genes, and the most polymorphic ones then disproportionately drive the phylogenetic reconstruction. This issue affects the WP dataset particularly because the most sampled gene (SA1-1) is more variable than others.

Despite these uncertainties, the nuclear trees unambiguously depart from the mitochondrial tree in several respects, confirming the discordances noted by previous studies. First, in *P. saharicus*, the Algerian lineage B shares nuclear alleles with the Tunisian lineage C (we cannot distinguish them with the genes analyzed, Figure 4) and both differ from the Moroccan lineage A—a pattern also retrieved by phylogenomic analyses (Doniol-Valcroze et al., 2021). Second, as shown by Komaki et al. (2015), *P. plancyi* is highly differentiated from *P. nigromaculatus* (Figure 6), despite bearing similar mtDNA.

In contrast to the mitochondrial tree, *P. nigromaculatus*, Korean and Russian populations (bearing the early-diverged mitogroup A) group together with the Japanese populations (mitogroups B, F, H—*P. nigromaculatus nigromaculatus*), rather than the Chinese populations (mitogroups C and G), which formed a separate nuclear lineage (*P. nigromaculatus reihardtii*) (Figure 6). Other discordances, notably within the WP marsh frog clade, may pertain to low phylogenetic power. For instance, our nuclear species tree does not branch *P. ridibundus kurtmuelleri* with *P. ridibundus ridibundus* (Figure 4), while the latter two are sister taxa in phylogenomic (Dubey, Maddalena, et al., 2019) and mitochondrial trees (Figure 1).

Finally, the *P. ridibundus* alleles trapped in hybridogenetic hybrids corresponded to several lineages (Figures 4 and 5): *ridibundus* in Western and Northern Europe; *kurtmuelleri* in Hungary; cf. *ridibundus* G in Russia; and cf. *ridibundus* X in Central and S-Italy. The latter was retrieved in marsh frogs from the Balkans and Central Asia and did not correspond to a particular mitogroup.
FIGURE 3 Phylogenetic relationships and distribution of *Pelophylax* mitogroups in the Eastern Palearctic (EP). Taxonomic names are given, when available (Table S1). The map combines barcoding information from six genes; shaded areas: native distributions (see the legend of Figure 1 for details); arrows: introduced populations genotyped. The inset map zooms in on Xinjiang Uyghur region (NW-China). The tree shows the Bayesian analysis of full and partial mitogenomes (~16.8 kb); see Figure S1 for terminal branches and for the Maximum-Likelihood analysis; node circles illustrate branch support; the grey lines emphasize the cyto-nuclear discordance in *Pelophylax nigromaculatus/plancyi* by showing the nuclear topology retrieved by the multispecies coalescent (Figure 5; see also Figure S2 and Komaki et al., 2015).
FIGURE 4  Phylogenetic relationships based on the multispecies coalescent and distribution of Pelophylax nuclear lineages in the Western Palearctic. Taxonomic names are given, when available. The species tree is based on five genes (~4.9 kb); thin lines: all sampled trees; thick lines: consensus tree. Barcoding information are combined on the map; shaded areas: presumably native distributions (see the legend of Figure 1 for details); arrows: introduced populations genotyped. The inset map zooms in Kamchatka Peninsula (top right). Letters (map) and asterisks (tree) identify the lineages retrieved in hybridogenetic hybrids. The distributions of the Pelophylax ridibundus lineages are detailed in Figure 5.
3.3 | Timeframe of diversification

The four time-calibrated mitochondrial phylogenies, each enforcing a Messinian origin to a different taxon, recovered contrasting timeframes of diversification (Figure S3). Calibration D (MSC split of cerigensis) yielded an unrealistically old time-tree, with the root age estimated to the Early-Cretaceous (136 My) and Early/Mid-Miocene divergences (>10 My) for almost all lineages, including closely related ones, and was not considered in the plausibility tests. Calibrations B (MSC split of cypriensis) and C (MSC split of perezi) proposed essentially similar timeframes, except for the nodes involved in the calibration, while calibration A (MSC split of cypriensis) provided
generally older ages. All three were compared with the BPT and the newly proposed LCPT.

The BPT pointed toward calibrations B and C as the most plausible (Figures 7 and 8). Out of the 39 major nodes recovered, 35 (90%) had median ages corresponding to one or more paleographic events, and 38 (97%, B) or 37 (95%, C) had overlapping 95% highest posterior density with them. For calibration A, these numbers dropped to 28 (72%) and 29 (74%), respectively. Some major splits are explained by calibrations B and C only, for example, *perezi/saharicus* (Betic crisis with B; MSC with C); the *saharicus/cf. saharicus* diversification (Rif corridor opening with B; Plio-Pleistocene climatic oscillations with C); the split between pool and marsh frogs (Badenian Salinity Crisis); the origin of *P. shaipericus* (MSC); the divergence of the main EP clades (monsoon intensification); the mainland/Japan dispersification of *P. nigromaculatus nigromaculatus* (volcano-tectonics in SW-Japan + occasional landbridges); and the surges of all cf. *ridibundus* and *nigromaculatus/plancyi* lineages (dramatic environmental changes linked to the Quaternary glaciations). We further note that while the mitochondrial age estimate of Taiwanese *P. fukienensis* (7.0–7.5 Mya) under calibrations B and C is inconsistent with the emergence of Taiwan (~5 My), the different position(s) of this species in the nuclear phylogenies (Figure 5; Figure S2) always implies a younger origin (see also Komárek et al., 2015).

The LCPT also points toward calibrations B and C as the most plausible (Figure 9). Seven nodes correspond to the split between eight pairs of lineages for which patterns of hybridization/admixture in the wild had been documented (Table 4), and their ages are all coherent with the speciation continuum documented in Palearctic anurans (Dufresnes et al., 2021). Specifically, calibration B and C recovered 17–20 My (cf. *ridibundus/perezi*) and 13 My of divergence (cf. *ridibundus/lessonae*) between the genomes of hemiclines, which compares to the divergence of other non-recombining anuran hybrids (Drlion et al., 2019; Dufresnes, Mazepa, et al., 2019). Instead, calibration A suggested up to 30 My for *P. ridibundus/perezi*; we are not aware of fertile hybrids between species of such deep divergence in amphibians. Calibrations B and C further retrieved ≤3.5 My of age for all the widely admixing taxa, which accordingly corresponds to the lower edge of the grey zone of anuran speciation, that is, the window of divergence where differing lineages admix (Dufresnes et al., 2021). Under calibration A, two of them (cf. *ridibundus* D/F vs. *ridibundus/kurtmuelleri* and *lessonae vs. bergeri*) received older age estimates (5.4 and 4.2 My) respectively that rather correspond to the upper edge of the grey zone, where lineages exhibit steep hybrid zones with restricted admixture. Finally, all three calibrations are in principle coherent with the interactions between *P. epeiroticus* and *P. ridibundus kurtmuelleri*. These taxa occasionally hybridize in areas of sympathy, and while only sterile F1s have been usually documented (Sofianidou, 1996), allozyme and microsatellite allele sharing characteristics of F2s and backcrosses (suggestive of genetic admixture) were suspected in a few individuals (Hotz et al., 2013; Papežík et al., 2021; Sagonas et al., 2020). Hybridization with or without restricted admixture is compatible with either calibration A (13.8 My), B (8.3 My) and C (8.5 My).

Finally, calibrations B and C are also consistent with the fossil record: (1) the restriction of the Upper Oligocene fossils to NW/W-Europe, as shown by the species *P. meriani* and *P. aquensis* (documented from 21.5 to 24.5 Mya), suggests that WP ancestor(s) had not yet diversified across the whole WP by then; (2) the appearance of the genus in Iberia by the Early Miocene, notably with the species *P. pueyoi* (documented from 21.0 to 6.2 Mya) suggests that the *perezi/saharicus* clade emerged around this epoch; (3) the nearly absent records in SE-Europe and Anatolia until the Middle/Late Miocene suggest that marsh frogs only evolved by this epoch, as exemplified by the species *P. barani* (reported from 10.6 to 9.5 Mya).

The similar timetrees obtained with calibrations B and C are thus the most plausible, and distinguishing them is hardly possible with the present BPT and LCPT. The few discrepancies between these timetrees are explainable by alternative events: the split between the *P. perezi* and the *P. saharicus* clades is consistent with either the Betic crisis ~16 My (B) and the MSC ~5.3 My (C); the *P. saharicus/cf. saharicus* diversification is consistent with the opening of the Rif corridor ~8 My (B) and Plio-Pleistocene climatic changes ≤3.5 My (C). We note that by combining nuclear and mitochondrial sequences under a community level BPT approach, Eh et al. (2019) estimated these nodes to be 13.2 and 6.4 Mya, which are closer to calibration B. This timetree may thus be more accurate, and is accordingly presented with a biogeographic interpretation in Figure 10.

### 3.4 Water frog introductions

Our barcoding framework covered 19 invaded areas (16 in the WP, 3 in the EP) involving at least 14 lineages from 5 species (*P. ridibundus*, *P. cf. saharicus*, *P. shaipericus*, *P. perezi*, *P. nigromaculatus*), and perhaps a sixth one (*P. lessonae*, but see below). Five lineages, all from *P. ridibundus*, are featured in the stocks of frog sellers (*ridibundus*, *kurtmuelleri*, *bedriagae*, *cf. ridibundus* B and F). In complement to the literature, our analyses provide original insights on the following cases.

Near Belgorod, at the Russia-Ukraine border, two cyt-b sequences generated by Sumida et al. (2000) unambiguously branch with mitogroup *cf. ridibundus* K, which is naturally restricted to Cilicia in S-Turkey; its presence in Eastern Europe implies an unnoticed introduction, noting that the same haplotype was reported in the introduced populations from Switzerland (Dubey et al., 2014). Second, Piton et al. (2008) and Akın et al. (2010) published ND2 and ND3 sequences from the Libyan coast (Cyrena) that correspond to *kurtmuelleri* mitotypes found throughout the Balkan Peninsula and offshore islands (e.g., Kythira), Libya is supposedly inhabited only by *P. saharicus* (Blackburn et al., 2022), so *P. ridibundus kurtmuelleri* must have been introduced. Third, the new CO1X sequences confirm that the marsh frogs invasive in S/W-France are composed of at least three *P. ridibundus* lineages, namely *cf. ridibundus C, kurtmuelleri* and *ridibundus* (as reported locally, e.g., Doniol-Valcroze et al., 2021; Dufresnes, Denoël, et al., 2017), and that introduced populations in the Azores (Spigoni et al., 2012) and the Canary Islands belong to *P. perezi* (Figure 1).
FIGURE 7  Biogeographic plausibility test in the Western Palearctic. Estimated node ages and their confidence intervals (error bars), based on calibrations A, B and C, are shown for the regions where they underly vicariance, overlayed with relevant paleographic events affecting these regions (described in Table 3). Calibrations B and C yield more explainable node ages, that is, that can be attributed to candidate paleographic events. Node labels are shown on the summary tree; grey line: nuclear branching in Pelophylax saharicus (Figure 4, Doniol-Valcroze et al., 2021).
 Altogether, the most frequently introduced lineages are kurtmuelleri (seven confirmed areas), ridibundus (four areas), cf. ridibundus B/C/D/F from W-Anatolia (eight areas), and cf. ridibundus K from Cilicia (four areas) (Table S5). N-Italy hosts the most diverse population of non-native water frogs (eight lineages), followed by France, Belgium and Switzerland (≥5 lineages each). Beyond lineages, mapping the native distributions of introduced mitotypes brought little additional information on the putative origins of these introductions in most instances (Figures S4 and S5). This is because of the high number of alien mitotypes, their absence, or, on the contrary, their fast occurrence across native ranges, the use of different markers in invasive and native ranges (e.g., Belgium), or the fact that some sequences were not made public (e.g., P. nigromaculatus in China).

Hypothetical translocation pathways inferred from the patterns of diversity and distributions are summarized in Figure 11. The large invasions involve multiple sources in Europe, Asia Minor and even North Africa. The origins of some lineages can be traced with a certain accuracy given the restricted distribution of the mitogroups/mitotypes involved, for example, Tunisia/E-Algeria (cf. saharicus in France), the Levantine coast (pediagrae in N-Italy and Belgium), the Volga drainage (cf. ridibundus G mitogroup + ridibundus/cf. ridibundus G nuclear alleles in Kamchatka), S-Turkey (cf. ridibundus B in Cyprus); the Sichuan Basin (reinhardtii/plancyi mitogroups G/C in Tibet). In contrast, some introduced mitotypes from the widespread kurtmuelleri/ridibundus/cf. ridibundus lineages are found in most of south and eastern Europe and extend as far as the Baltic Sea, the Caucasus and the Volga River drainage. Three regions are recurrently highlighted as the potential sources of several independent introductions: (1) the eastern Adriatic coast, which hosts P. shqipericus, and the P. ridibundus kurtmuelleri mitotypes found in Aspromonte (S-Italy), Sardinia, and the Po Valley; (2) the SE-Balkans and adjacent W-Anatolia, which hosts the P. ridibundus cf. ridibundus C/D/F lineages and the P. ridibundus ridibundus/kurtmuelleri mitotypes introduced in France and Switzerland; and (3) the Cilician region, where P. ridibundus cf. ridibundus K is naturally restricted.

Finally, the updated distributions of the kurtmuelleri and bergeri lineages may question the introduced status of some populations. On the Adriatic coast, P. ridibundus kurtmuelleri extends naturally as far north as Istria in W-Croatia and its alleles also regionally occur eastward outside the southern Balkans, namely within the range of P. ridibundus ridibundus in the Pannonian plains, close to the Black Sea and Baltic coasts, as well as the Dnieper Basin (Figures 2 and 5). In pool frogs, P. leseanae bergeri mtDNA was revealed in southeastern France, not far from the Apennine populations, therefore, putatively connecting the native distribution of this lineage in Peninsular Italy with parts of its presumed introduced range in France and Switzerland (Figure S6).

4 | DISCUSSION

In this study, we retrieved 13,619 barcode sequences from 1765 localities of Pelophylax water frogs and inferred multilocus
mitochondrial and nuclear phylogenies to retrace patterns of evolution, diversity and distribution across both native and introduced ranges. These resources were combined with timetree plausibility tests, introducing a novel approach, which showcases the genus as a biogeographic paradigm for the Palearctic batrachofauna. Our framework offers preliminary insights on the origins of hybridogenic hybrids and emphasizes the remarkably diverse composition of exotic populations, which combined, constitute one of the largest amphibian invasions worldwide and should be acknowledged accordingly in conservation policies.

4.1 Toward integrative tests of timetree plausibility

The drastically different timeframes obtained for the same diversifications (Kumar et al., 2022) stresses the need to ascertain the plausibility of time trees before drawing biogeographic conclusions. Plausibility tests can help disentangling competing timeframes, but they remain rarely implemented. In effect, biogeographic interpretation usually relies on a single pre-defined calibration scheme without considering alternative scenarios. This issue is also characteristic of the *Pelophylax* phylogeographic literature (e.g., Akin et al., 2010; Beerli et al., 1996; Dubey & Dufresnes, 2017; Komaki et al., 2015; Lymberakis et al., 2007; Nicolas et al., 2015; Plötner et al., 2012), which we have re-examined at the global scale using the BPT, an approach initially proposed by Fromhage et al. (2004) for Mediterranean amphibians and later extended to the whole herpetofauna community (Ehl et al., 2019).

Additional ways to authenticate time trees may be helpful. For instance, comparing substitution rates to values taken from the literature may flag questionable molecular clocks (Hipsley & Müller, 2014). Yet, published rates must be compared with caution as these are often obtained from different “local clocks” that were independently inferred between organismal groups and calibration settings (Ehl et al., 2019). As a complement, here we have incorporated specific knowledge of the anuran speciation continuum to predict realistic age boundaries with the LCPT, which effectively complemented the BPT in identifying the most proper calibrations (Figures 7–10). The LCPT could be similarly applied to other vertebrates that follow the same mode of speciation, when post-zygotic barriers accumulate gradually with genomic divergence (i.e., the “mass of genes” model of hybrid incompatibilities, Dufresnes et al., 2021), providing some independent a priori knowledge of the timeframe of this relationship.

Despite these efforts, distinguishing between calibrations B and C (which yielded similar timeframes overall) was challenging. Accordingly, the BPT loses informativeness when too many candidate paleographic events are available, thus offering equally plausible explanations for competing age estimates. In the Northern Hemisphere, this is particularly true of young lineages, for which the Quaternary climatic oscillations can always be invoked as a cause of phylogeographic structure (Figures 7 and 8). The LCPT may be particularly helpful in such situations. Here, a decisive test to disentangle between calibrations B and C would be to measure reproductive isolation (hybridization and introgression) between *P. perezi* and *P. cf. saharicus* in S-France (where the latter was introduced; Doniol-Valcroze et al., 2021; see below), as well as in the contact zones between the *saharicus* lineages in Algeria, and for which different outcomes are expected given the contrasting levels of divergence predicted by each calibration.

Whatever their biogeographic or biological basis, time tree plausibility tests are only as reliable as the trees themselves. Our mitochondrial tree, although robust and including all known lineages, features a few instances of cyto-nuclear discordance that imply to critically reassess the biogeographical significance of corresponding nodes (Figures 1 and 3). Here, the alternative nuclear topologies support rather than contradict our results. In the *P. saharicus* complex, the basal split of the Moroccan lineage suggested by the nuclear tree...
FIGURE 10  Mitochondrial time tree for Pelophylax (calibration B) and biogeographic scenario. Mitogroups are color-coded by geographic regions (closely related mitogroups are collapsed). Node values and bars indicate average node ages and their 95% highest posterior density. Dashed lines show genuine discordances from the nuclear tree. Right panels show paleomaps of Eurasia since the Late Eocene (Scotese, 2016), with pre-Pliocene fossil record (crosses) (Böhme & Ilg, 2003). The oldest Pelophylax fossils, as well as species known only from the fossil record are emphasized on the tree. The maps and tree are annotated with the following biogeographic hypotheses: dispersal from Eastern Palearctic (EP) to Western Palearctic (WP) at least 30 Mya (a); WP/EP disconnection by the Paratethys basin and aridification of Central Asia (b); expansion of the WP ancestors (c) and early diversification by the anoxic Paratethys sea (d); expansion of EP ancestors and pool/marsh frog ancestors (WP) during the Miocene climatic optimum (e) and European diversification following the Badenian Salinity Crisis (f); differentiation between Pelophylax perezi and P. saharicus ancestors in Iberia during the Betic Crisis (16–14 Mya), and subsequent isolation of the latter in N-Africa (g); Late-Miocene diversification of the mainland EP clade following the intensification of the East Asian monsoon (h), and disconnection of Japan by volcano-tectonic activities (i); opening of the Rif corridor in N-Africa (8–5.6 Mya), causing the split between P. saharicus A and P. cf. saharicus B/C (j); separation of the Balkans and Anatolia by completion of the Mid-Aegean trench (13–9 Mya), isolating P. epeiroticus (k); Messinian Salinity Crisis (MSC), connecting then disconnecting Crete and the SW-Adriatic coast from adjacent lands, simultaneously isolating P. cretensis and P. shapiericus (l); diversifications throughout the Palearctic due to Plio-Pleistocene climate cooling/drying and sea resurgences, alternating range connections via land bridges (m) and range disconnections into multiple glacial refugia isolated by orography (n) and newly formed deserts (o).
### TABLE 5 Diversity of exotic taxa/haplogroups detected across 19 introduced ranges.

| Frog farms | cf. saharicus C | perezi | bergeri A | shujpericus | bedriagae | ridibundus | kurtmuelleri | cf. ridibundus B | cf. ridibundus C | cf. ridibundus D | cf. ridibundus F | cf. ridibundus G | (cf. ridibundus C/D/F/G) | cf. ridibundus K | reicheldeii | Total |
|------------|------------------|--------|-----------|------------|-----------|------------|--------------|------------------|-----------------|-----------------|-----------------|-----------------|------------------|---------------------|--------------|--------|-------|
| Azores     |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              | 8      |
| Canaries   |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| N-Italy    |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Aspromonte |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Sardinia   |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Malta      |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| S/W-France |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Switzerland|                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Belgium    |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Luxemburg  |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| N-Croatia  |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Poland     |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Kaliningrad|                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Cyprus     |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Libya      |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Belgorod   |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Kamchatka  |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Tibet      |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Uygur      |                  |        |           |            |           |            |              |                  |                 |                 |                 |                 |                   |                     |              |        |
| Total      | 8                | 1      | 3         | 1          | 5^b       | 5^b        | 3            | 10^b            | 1                | 1                | 1                | 3                | 1                 |                     |              |        |

**Note:** Some ranges correspond to several independently studied populations.

- ^a^ Old introductions.
- ^b^ Introduced/native status unclear (see main text).
- ^c^ Based on only SAI-1 size polymorphism and to be confirmed.
FIGURE 11 Tracing *Pelophylax* invasions with genetic barcodes—a summary. Composition of the invasive populations (major mitotypes, as in Figure 1) and their potential approximate origins according to our analyses (see also Table 5). Arrows highlight translocation pathways based on the native distributions of mitotypes and mitogroups (Figures S4 and S5); dash lines emphasize inaccuracy. Cases that may not reflect actual introductions are not reported (Table 5).

(Doniol-Valcroze et al., 2021) is coherent with the hypothesis that the opening of the Rif corridor during the Miocene initiated the diversification near the Moroccan/Algerian border. Instead, the basal split of the Tunisian lineage suggested by the mtDNA tree (Figure 1) implies an unknown phylogeographic break further east (see also Beddek et al., 2018). In China, the potentially younger age of *P. fukiensis* than suggested by the mitochondrial phylogeny makes the fairly recent (Pliocene) orogenesis of Taiwan a plausible hypothesis of vicariance. In Korea and Japan, the old mitochondrial divergence but young nuclear differentiation is coherent with and alternance of historical biogeographic barriers and dispersal in the region (see below).

As these examples show, carefully considering discordant phylogenetic patterns independently retrieved by mitochondrial and nuclear markers can thus improve the interpretations of time trees, whereas combining them into a single species tree would instead produce “chimeric” topologies. Hence, while complicating phylogenetic inferences, the widespread cyto-nuclear discordance documented in many amphibians (e.g., Dufresnes, Mazepa, et al., 2019; Wielstra et al., 2017; Zielinski et al., 2013) also represents opportunities for historical biogeography, by giving hints on the age of past secondary contacts (e.g., the divergence of captured mitochondria), or informing on previous vicariance events that only left trace in the mitochondrial ghost lineages, e.g., Wielstra et al., 2021) due to subsequent lineage fusion (Garrick et al., 2019).

4.2 | The Paratethys: An underappreciated driver of amphibian vicariance?

The phylogeography of nearly all Palearctic amphibians has been relatively well-examined within their western and eastern ranges, and specific regions therein (e.g., Busack, 1986; Dufresnes & Litvinchuk, 2022; Ehl et al., 2019; Oosterbroek & Arntzen, 1992), but causes for the early emergence of distinct clades between these ranges received a more sporadic attention. Ranid frogs originate from Asia (Yuan et al., 2016) and putatively gained access to Europe in the Early Oligocene (~30Mya), a timing confirmed by German and French fossils from the Paleogene zone MP22 (32.5–30.9 Mya, Lemierre et al., 2023). This timeframe suggests that the closing of the Turgai Sea—a large body of salt water that previously fragmented Eurasia since the Cretaceous (Akhmet'ev, 2011; Briggs, 1995; Duellman & Trueb, 1994)—opened the route for colonization (a in Figure 10). The lack of secondary dispersal events emphasizes that European and Asian faunal exchanges remained limited overall, potentially due to the aridification of Central Asia and the westward retreat of the Paratethys Sea (b in Figure 10), two side effects of the first uplift of the Qinghai-Tibetan Plateau (QTP, ~55–35 Mya) (Bougeois et al., 2018; Meijer et al., 2019; Sun et al., 2020).

In comparison, the six other amphibian groups with extant representatives in both the EP and WP bear various timeframes of divergence: ~12.5 My in *Bufo* (Garcia-Porta et al., 2012), ~14 My in *Bombina* (Pabijan et al., 2013), ~23 My in *Hyla* (Li et al., 2015), ~25 My in *Rana* (Yuan et al., 2016), >40 My in Pleurodelinae newts (Arntzen et al., 2015; Kieren et al., 2018; Malacuso et al., 2022). Some of these splits may thus have similar drivers as *Pelophylax* (e.g., *Hyla, Rana*), while others imply older and younger connections, for example, across the Turgai Sea at moments when it was briefly permeable (e.g., Pleurodelinae), and during the Miocene climatic optimum, after which the aridification of Central Asian deserts intensified (e.g., *Bufo, Bombina*).

The peculiar paleography and biogeography of the Paratethys domain offers fascinating prospects. The sea was putatively formed in the Early Oligocene (~34 Mya), when it became isolated from the Peri-Tethys (the northern part of the Tethys Ocean) and the Mediterranean region due to a global drop in sea level combined with the simultaneous orogenesis of Euro-Anatolian mountains.
ranges (Alps, Carpathians, Dinarides, Taurus, Elburz) (Meulenkamp & Sissingh, 2003; Palcu & Krijgsman, 2023). The Paratethys then consisted of a series of deep interconnected basins over nearly 4000 km, from the Alps to the Urals. Due to its isolation, the sea eventually turned into an giant anoxic lake, until a marine transgression re-established connection with Mediterranean waters (~15 Ma) (the Badenian Flooding) (Bartol et al., 2014; Palcu et al., 2019). This ventilation was brief as the central basins disconnected again and shortly experienced an evaporation event, the Badenian Salinity Crisis (~13.8-13.4 Ma) (De Leeuw et al., 2010). The Paratethys subsequently shrunk into brackish mega-lakes, of which the Black Sea, Caspian Sea and Aral Sea are its present-day remnants (Palcu et al., 2021; Popov et al., 2006). From a climatic perspective, the progressive retreat of the Paratethys has led to profound changes in Earth’s pressure system and atmospheric circulations and was directly linked to the establishment and intensification of the East Asian monsoon (Ramstein et al., 2012), increasing humidity and aridity in coastal and inland regions of China, respectively (Sun, Gong, et al., 2015; Sun, Ma, et al., 2015, 2015; Zhang et al., 2007).

Like the Mediterranean Sea, the dynamics of the Paratethys Sea must have significantly contributed to amphibian vicariance since the Oligocene. In a meta-analysis of allozyme data available at the time, Oosterbroek and Arntzen (1992) elegantly linked patterns of divergence in several terrestrial animals to the paleography of the Paratethys, with emphasis on the nearly continuous disconnections between Asia Minor, Eastern, and Western Europe during most of the Miocene (see the paleographic maps in Meulenkamp & Sissingh, 2003). In Pelophylax, the Paratethys likely constrained the distributions of the European ancestors, notably between proto-marsh and proto-pool frogs, whose divergence coincides with the Badenian events—when the sea briefly switched from an impassable barrier to a gateway between northwestern and southeastern European ranges (e in Figure 10). The same ancient phylogeographic break seems to exist in many other widespread anurans: the potential paleodistribution and divergence time of the two main clades of WP Rana (R. iberica/temporaria vs. R. graeca/macromenis), Pelobates (P. cultripes vs. P. balcanicus/syracus), Pelodytes (P. caucasicus vs. P. punctatus/ibericus) all roughly coincide with a putative Badenian vicariance (~15–13 Ma) in the community-clock of Ehl et al., 2019.

In the EP, extant clades emerged during the Late Miocene (~9 Ma), potentially linked to the re-intensification of the East Asian monsoon (dated to ~9.5 Ma by Sun, Ma, et al., 2015; h in Figure 10), which itself was associated with the retreat of Paratethys alongside phases of QTP growth (Zhang et al., 2007, 2021). Although it is more frequently invoked for plant biogeography (e.g., Ma et al., 2019), strengthening of the monsoon regime may be a key driver of amphibian divergence in the EP, especially in Chinese species groups with inland and coastal lineages that likely diversified during the Late Miocene (see numerous examples in Dufresnes & Litvinchuk, 2022). Hence, the Paratethys domain appears a major biogeographic determinant of Palearctic amphibians at the global scale, both directly as a fluctuating physical barrier, and indirectly due to its influence on the Asian paleo-climate.

### 4.3 | Other biogeographic drivers

In the WP, the dynamic paleography of the Mediterranean region and its consequences on herpetofaunal evolution have been extensively discussed (e.g., Ehl et al., 2019). Our Pelophylax timetree could relate four well-documented episodes of marine transgressions/subsidence to explain up to seven Mediterranean splits, including some that previously remained enigmatic. In the western parts of the basin, the split between P. perezi and P. saharicus retrieved by our preferred calibration B matched the Betic Crisis, when Iberia was fragmented by marine transgressions (Lonergan & White, 1997) and large saline lakes (Anadón et al., 1989) that supposedly led to the vicariance of many freshwater organisms (Altaba, 1998), including amphibians (e.g., Alytes isteris; Altaba, 1997; Ambu et al., 2023). South-Iberia then momentarily detached from Europe by geotectonic activity and multiple sea straits, and eventually connected to Africa, which could explain the dispersal and subsequent isolation of proto-P. saharicus into their present range (g in Figure 10). This two-step vicariance process may similarly account for the appearance of other amphibians, for example, the tree frogs Hyla meridionalis/carthaginensis, the salamander Salamandra algira, and the newts Pleurodeles nebulosus/vojaci, which also correspond to Middle-Miocene relics now naturally isolated and fragmented across N-Africa (Ehl et al., 2019).

In the southern Balkans, the deep divergence of P. epeiroticus coincides with the formation of the mid-Aegean trench (k in Figure 10), which separated southeastern Europe from Anatolia (Dermitzakis & Papanikolaou, 1981). Southern Greece, particularly the Peloponnese, host old herpetofauna endemics that may have a similar origin, for example, Anguis cephallonica (Gvoždík et al., 2023), Anatolian/Balkan divergences dated to the Late-Miocene are typically found in amphibians (e.g., Pelobates syriacus/balkanicus, Dufresnes, Strachinis, et al., 2019). Further north, the E-Adriatic P. shipericus may have split from Apennine pool frog ancestors during the MSC (l in Figure 10). Its closest relatives (P. lessonae lessonae/bergeri) are now distributed in peninsular Italy, and likely persisted there since the Pliocene, as moist paleo conditions favored long-term refugial maintenance for amphibians (Macaluso et al., 2023). Ancient cross-Adriatic connections are emphasized by the phylogenetic relatedness between other Apennine and Balkan taxa for example, in Triturus newts (T. carnifex/macedonius, also dated to the MSC, Wielstra et al., 2019), Rana brown frogs (R. graeca/dalmatina, Yuan et al., 2016), or by evidence from Messinian hybridization events between them, for example, in Anguis slow worms (A. veronensis/cephallonica, Gvoždík et al., 2023).

On the continent, most phylogeographic breaks reflect Plio-Pleistocene landscape changes linked to the climatic oscillations (m, n, and o in Figure 10), which promoted the high diversity now found across the Mediterranean Peninsulas under the classical "refugia within refugia" model, as seen for the diversification of P. lessonae
in the Apennine Peninsula and of *P. ridibundus* in the Balkan Peninsula and Anatolia. In Central Asia, the Pleistocene emergence of several cf. *ridibundus* mitogroups parallels the setup of the dune system in deserts that now efficiently act as dispersal barriers, for example, the Karakum and Kyzylkum (Atamuradov, 1994), as in *Bufo* green toads (Dufresnes, Mazepa, et al., 2019). In N-Africa, however, the *Mio-Pliocene split found within *P. saharicus* may rather be the cause of a tectonic event, namely the opening of the Rif corridor, as also hypothesized for *Discoglossus* painted frogs (Pabijan et al., 2012).

In the EP, in addition to climatic changes (see above), the complex phylogeographic patterns recovered in *Pelophylax* are congruent with the geotectonic history. During the Miocene, Japan progressively became an island arc, as southwestern lands pulled away from the mainland (opening the Sea of Japan) and central/northern parts rose from the seabed by orogenesis and accretion (Barnes, 2003; Isozaki, et al., 2011). The archipelago eventually became isolated following intense Plio-Pleistocene volcano-tectonic activity along the Median Tectonic Line—notably in the Hohi volcanic zone of Central Kyushu (0.5–1.5 Ma; Kamata & Kodama, 1999)—and more recently by the opening of the Korean Strait (i in Figure 10). However, the Strait remains relatively shallow (~100m deep on average) and was mostly exposed during glacial maxima (Lee et al., 2022), offering recurrent opportunities for dispersal in many amphibians (Dufresnes & Litvinchuk, 2022). The alternation of isolation vs. connectivity episodes within the last millions of years likely drove the spread of closely related mitogroups now found thousands of kilometers apart, in Japan and China, and can even explain some of the cytonuclear discordances. In particular, the mitochondrial divergence in Korean *P. nigromaculatus* suggests a Pliocene event of vicariance from Japan, followed by a more recent colonization (homogenizing the nuclear genomes of Japanese and Korean populations). Within Japan, notably on Honshu, tectonic activities promoted volcanism, tectonic basin expansion, and fault subduction (e.g., Fossa Magna; Barnes, 2003) that geographically associate with the subdivisions reported within *P. nigromaculatus* and *P. porus*, as in other frogs (e.g., Zhangixalus, Dufresnes et al., 2022) and toads (e.g., *Bufo*, Igawa et al., 2006).

The origin of *P. fukienensis* remains unsettled given the distinct outcomes of different analyses but vicariance following the orogeny further suggests lesser divergence between cf. *X. lessonae* and *P. grafi* mate exclusively with one parent (*P. lessonae bergeri* and *P. perezi*), and always transmit a *ridibundus* clonal genome. From these observations, it has been assumed that only *P. ridibundus* is to discard one genome and duplicate the other, so homologous chromosomes can properly segregate during meiosis and then be transmitted clonally. The genome exclusion is not random, and depends on the parental species involved (Doležálková-Kaštánková et al., 2021; Lavanchy & Schwander, 2019). To date, three main hybridogenetic systems are distinguished and labelled according to the taxonomic names of their hybrids, i.e., the *P. esculentus esculentus* system (*P. l. lessonae × P. r. ridibundus*), the *P. esculentus hispanicus* system (*P. l. bergeri × P. r. ridibundus*) and the *P. grafi* system (*P. perezi × P. r. ridibundus*).

The widespread *P. esculentus esculentus* independently emerged multiple times in Central, Northern and Western Europe (Hoffmann et al., 2015), and is perpetuated by the exclusion of either the *ridibundus* or the *lessonae* genome (reviewed by Dufresnes & Mazepa, 2020). *Pelophylax esculentus hispanicus* and *P. grafi* mate exclusively with one parent (*P. lessonae bergeri* and *P. perezi*), and always transmit a *ridibundus* clonal genome. From these observations, it has been assumed that only *P. ridibundus* is to discard one genome and duplicate the other, so homologous chromosomes can properly segregate during meiosis and then be transmitted clonally. The genome exclusion is not random, and depends on the parental species involved (Doležálková-Kaštánková et al., 2021; Lavanchy & Schwander, 2019). To date, three main hybridogenetic systems are distinguished and labelled according to the taxonomic names of their hybrids, i.e., the *P. esculentus esculentus* system (*P. l. lessonae × P. r. ridibundus*), the *P. esculentus hispanicus* system (*P. l. bergeri × P. r. ridibundus*) and the *P. grafi* system (*P. perezi × P. r. ridibundus*).

By combining recently published nuclear barcoding information, our phylogeographic framework somewhat calls to re-appraise these assumptions. Nuclear alleles from four different *P. ridibundus* lineages were retrieved in hybridogenetic hybrids. In Italy, Dubey and Dufresnes (2017) showed that the *ridibundus* genomes of *P. esculentus esculentus* and *P. esculentus hispanicus* formed distinct SAI-1 lineages and microsatellite haplotypes, reflecting their independent origins. The SAI-1 alleles found in *P. esculentus hispanicus* are particularly intriguing. They branch with Western Anatolian cf. *ridibundus* nuclear lineages rather than the European *ridibundus*, yet are strongly divergent from any reported haplotype (Plötner et al., 2009, 2012, 2015), begging the question whether it belongs to a different, presumably extinct “ghost” taxon preserved only in the hemiclines (Dubey & Dufresnes, 2017). This “ghost”, characterized here as cf. *ridibundus* X (Figure 5), was found to be identical or highly similar to new sequences from the Balkans and Central Asia (Ualiyeva et al., 2022). The inclusion of additional genes in the nuclear phylogeny further suggests lesser divergence between cf. *ridibundus* X
and *ridibundus* (or from any other lineages of the complex) than previously suspected based on SAI-1 alone (Figure 4). The marsh frog lineage that sired *P. esculentus hispanicus* is thus very much "alive," although its evolutionary significance and history is hard to apprehend, given the complex nuclear phylogeography of *P. ridibundus* (Figure 4).

Most of the *P. esculentus esculentus* sequences phased and analyzed contained a marsh frog copy of the *ridibundus* lineage and pool frog copy of the *lessonae* lineage, as expected. However, the marsh frog copy of a few individuals corresponded to other lineages, namely *kurtmuelleri* in the Balkans (Herczeg et al., 2016) or cf. *ridibundus* G in Eastern Europe (Miura et al., 2021). Miura et al. (2021) demonstrated that the cf. *ridibundus* G genome triggered the elimination of the *lessonae* genome in hybrids (as in regular *P. esculentus*).

Hybridogenesis was also suspected in southern France between introduced *P. ridibundus kurtmuelleri* and native *P. perezi*, based on peculiar nuclear and mitochondrial genotypic combinations (Dufresnes, Denoël, et al., 2017). Hence, the capability of marsh frogs to trigger genome elimination seems not to be limited to *P. ridibundus ridibundus* sensu stricto but also extends to at least some of its related lineages.

The occurrence of alternative versions of known hybridogenetic systems offers exciting prospects, for instance to uncover the genomic/epigenetic factors involved in meiotic distortion processes that contribute to initiate and/or resist genome elimination. Drivers of genome elimination are expressed in most *P. ridibundus* ridibundus (and its related lineages) and some *P. lessonae lessonae*. Drivers involved in the resistance of genome elimination are unexpressed in most *P. lessonae lessonae, P. lessonae bergeri* and *P. perezi*, and are probably expressed in *P. shapericus* and *P. epeiroticus*—the latter two hybridize with *P. ridibundus kurtmuelleri* (in the wild) and *P. ridibundus ridibundus* (experimentally) without hybridogenesis (Hotz et al., 1985; Hotz & Uzzell, 1983). These observations could serve to formulate predictions in pairs of lineages that could then be experimentally crossed. More generally, our results emphasize the remarkable diversity of *Pelophylax* hybrids, with a large array of parental combinations capable of producing gametes, despite the inherent challenges of succeeding meiosis with incompatible genomes (Chapman & Burke, 2007).

### 4.5 | Palearctic water frogs as emerging global invaders

By mapping all water frogs analyzed with DNA barcodes in the scientific literature, we better capture the extent and diversity of allochthonous populations. In Europe, the spread of exotic *Pelophylax* lineages has become nationwide in France (Dubois, 1983; Pagano et al., 2001), Belgium (Holsbeek et al., 2010), Switzerland (Dufresnes et al., 2018), and new populations keep been regularly reported across additional regions (e.g., Cyprus, Plötner et al., 2015; Sardinia, Bellati et al., 2019; N-Italy (Bellati et al., 2023); Croatia, Jelić et al., 2022; Malta, Papežík et al., 2024). *Pelophylax* aliens have also been known from England (Wicherley et al., 2003; Zeisset & Beebee, 2003) and Spain (Arano et al., 1995). Further east, marsh frogs established and dispersed in new areas from the West and East Siberia, the Urals, and the Altai (Kuzmin, 1999). Some archipelagos also have a long history of hosting introduced *Pelophylax*, such as *P. perezi* in the Balearic (Silva-Rocha et al., 2018), the Canary (Gervais, 1836), Madeira and the Azores Islands (Malkmus, 2004).

In the latter, the shallow COX1 divergence between the Azores and Iberian populations (Doniol-Valcroze et al., 2021) runs against a natural colonization of the Atlantic islands (e.g., through rafting), which was initially considered based on patterns of allozyme differentiation (Balletto et al., 1990). The introduced populations are composed of multiple *P. ridibundus* lineages and additional species (*P. shapericus, P. saharicus, P. perezi*) which altogether represent a third of the WP phylogeographic diversity.

Ecological surveys are starting to demonstrate the negative impact of marsh frogs on native wildlife (especially other amphibians, Denoël et al., 2022; Pille et al., 2021), and this impact may intensify with climate change, given the broad thermal tolerance of some alien lineages (Padilla et al., 2023). Water frogs thus strike as one of the most diverse and successful amphibian invaders worldwide. In fact, their patterns of invasion share many similarities with notorious global invaders, for example, the clawed frog (*X. laevis*), the American bullfrog (*L. catesbeianus*), the cane toad (*R. marina*) and the Asian common toad (*D. melanostictus*). Like *Pelophylax*, these species can rapidly colonize large areas and feature distinct phylogeographic lineages (and thus geographic origins) among introduced populations (De Busschere et al., 2016; Kamath et al., 2016; Reilly et al., 2017; Vences et al., 2017), sometimes even representing several cryptic species (e.g., Mittan-Moreau et al., 2022).

The numerous lineages identified in introduced populations underlie the multiple origins of alien individuals, as well as the high diversity of their native ranges. Most exotic lineages found in Western Europe naturally inhabit the Balkans and Anatolia, two hotspots of phylogeographic diversity for the genus (Akin et al., 2010; Papežík et al., 2023). Tracing most introductions back to these two regions is remarkably coherent with the trading record of the frog leg industry (Altherr et al., 2022; Auliya et al., 2023; Holsbeek et al., 2008). Frog legs are now being mostly imported frozen from outside Europe (notably Asia), but Albania and Turkey remain the two biggest suppliers of *Pelophylax* frogs for the European market (notably Belgium, France, Switzerland and Italy), where they are typically caught from wild populations (Auliya et al., 2023) and imported alive to be transformed locally (Neveu, 2004). The annual Turkish exports amount 14–35 million individuals, and more than a third of captures originate from Adana Province (Çiçek et al., 2020), which is the region of Cilicia inhabited by the narrowly distributed but widely introduced cf. *ridibundus* K lineage (Figure 2). Parts of these stocks also originate from illegal hunting in nearby Syria (Altherr et al., 2022), home of *P. ridibundus bedriagae* (barcoded in Italy and Belgium). Although not yet detected, this lineage is expected to be also present in France, as frog leg producers frequently shipped frogs from Egypt during
the second half of the 20th century (Dubois, 1983). Accordingly, genetic analyses of the stocks bred by farms and sellers revealed a mix of Eastern Mediterranean lineages (Bellati et al., 2023; Holsbeek et al., 2008, 2010).

In China, the frog trade, both for consumption and traditional medicine (Alves et al., 2013) is also a potential vector of translocations. There are thousands of farms distributing live frogs to an even higher number of food sellers across the whole country, opening countless opportunities for escapes and invasions. Although the exploitation of Hoplobatrachus has been particularly highlighted (Alves et al., 2013), Pelophylax are also harvested (IUCN SSC Amphibian Specialist Group, 2020). It would thus not be surprising if, like in Europe, future prospections would find that water frog diversity and distribution patterns in Asia were also human-mediated. For instance, recent research on Giant salamanders showed that farming was the cause of multiple unsuspected introductions (Yan et al., 2018).

These observations unequivocally incriminate commercial imports as the primary source of introductions (see also Dubois, 1983). This is nicely illustrated by the spread of *P. ridibundus* in Switzerland during the second half of the 20th century, which started by independent clusters located near airports and breeding facilities (Dufresnes et al., 2018). In W-Europe, a few frog farms continue to develop, notably by producing their own breeds that are advertised as local delicacy. For instance, some companies use the race “Rivan 92”, which was engineered through selective breeding so the frogs can be fed fish granulates rather than live insects (Neveu, 2009), and which initially descent from individuals caught in Britany (France) but are of uncertain taxonomic nature (Neveu, 2004).

Besides farming and trading activities, it is likely that a large number of introductions also originate for the purpose of ornamentation in garden ponds. Water frogs were largely sold as ornamental species by local sellers or provided with aquatic vegetation (Kok et al., 2002; Percsy & Percsy, 2002). Given the widespread popularity of stocking garden ponds with ornamental species, this may represent a cause of multiple invasions, often likely cryptic (P.-A.C., M.D., personal observation). In addition, water frogs are also popular educational and research models (alongside Xenopus), and escapes from university and schools are common, for example, in France and Switzerland (P.-A.C., C.D., personal observation).

Although the situation seems beyond repair in several countries, the issue of *Pelophylax* invasions remains largely ignored by conservation authorities—water frogs are typically overlooked in bioinvasion reports (e.g., Renault et al., 2021). We see several reasons for this lack of awareness. Exotic water frogs resemble native ones and have been present for decades in some areas (Dubois, 1983), hence they do not attract much attention compared to more spectacular species (e.g., *X. laevis* and *L. catesbeianus* in France and Italy). In addition, many introduced populations are protected by national decree because they belong to species that supposedly occur naturally in other parts of the same country (e.g., *P. ridibundus* in France). Such a “conservation paradox” (sensu Koster et al., 2023) does not only safeguards invasive species, it also prevents conservation biologists from studying invasions by making it more difficult to obtain research permits.

The unstable taxonomy of *Pelophylax* is also certainly not favorable to the implementation of long-term conservation policies, which rely on scientific (taxonomic) names. Taxonomic uncertainties even benefit frog exporting companies, which abusively use generic names (e.g., “*P. esculentus*”) to escape the legislation (Altherr et al., 2022). Furthermore, several exotic lineages await a formal description more than 15 years after their discovery, being instead collectively designated as “*P. cf. bedriagae*” or “*P. cf. ridibundus*”. At the same time, taxonomic accounts disagree on the validity and species/subspecies rank of many described taxa, depending on the applied concepts and criteria (e.g., *P. ridibundus/kurtmuelleri; P. ridibundus/ cosyriensis/caralitanus/cerigensis; P. plancyi/hubeiensis/chosenicus*; see Dufresnes & Litvinchuk, 2022; Frost, 2023; Speybroeck et al., 2020).

Solving the taxonomy of water frogs thus appears urgent if distinct lineages are to be independently treated in national and international legislation, which will be necessary to ban invasive lineages from being imported, to eradicate them in invaded areas (when feasible), and to protect them in their native range, where unsustainable harvest is causing significant population declines (Çiçek et al., 2020). The ambition of reaching a stable taxonomy in *Pelophylax* will require a tremendous amount of dedicated taxonomic work, ideally based on next-generation phylogeography (i.e., implementing genomic data), to describe unnamed lineages, clarify nomenclatural ambiguities, and delimit species under standardized criteria (e.g., Dufresnes et al., 2023). For the time being, we tentatively list species and subspecies based on previous taxonomic discussions and our updated phylogeographic overview, which should offer a basis for coming updates (Table S1). In addition, our mitochondrial haplotype database, which spans six barcoding genes, is designed so that it can be easily reused and complemented by future invasion genetic studies.

While most water frog populations with a foreign genetic signature have a human-mediated origin, our results also call for cautious interpretations in a few cases. First, rather than multiple translocations, postglacial expansions could have also naturally carried *kurtmuelleri* mtDNA and nuclear haplotypes over Central and Northern Europe (Kolenda et al., 2017; Litvinchuk et al., 2020), as seen in some anurans of Balkan origin (e.g., *Bufotes viridis*, Höglund et al., 2022). This scenario is possible given that *P. ridibundus ridibundus* and *P. ridibundus kurtmuelleri* seem to represent shallow refugial lineages of the same biological species (Papežík et al., 2023) that remain highly similar at the genomic scale (Dubey, Lavanchy et al., 2019) and may thus hybridize without reproductive barriers.

Likewise, the presence of the Italian pool frog lineage *bergeri* A in Switzerland and France was attributed to ancient translocations on the basis of the large distributional gap between the closest native (N-Apennines; Capula, 2006) and putatively introduced records (near Lyon; Dufresnes, Di Santo, et al., 2017), which made natural colonization an unlikely hypothesis (Dubey et al., 2014; Dufresnes & Dubey, 2020; Dufresnes, Di Santo, et al., 2017). We showed here that the *bergeri* mitogroup extends in at least parts of this gap along the Ligurian and Eastern French Mediterranean coastline, although at this stage, it remains undetermined whether these populations are indigenous and if they belong to pool frogs, marsh frogs or their
hybridogenetic hybrids—mitochondrial transfer among all three is common (Plötner et al., 2008). From a biogeographic perspective, it is conceivable that P. lessonae bergeri briefly expanded out of NW Italy after the last glacial maximum, after which it was replaced by other Pelophylax species (e.g., P. perezi and now P. ridibundus), given that this Italo-French route is one of Europe's post-glacial recolonization's paradigms (Taberlet et al., 1998).

4.6 | Conclusions

Despite well-known limitations (Rubinoff et al., 2006), single gene barcoding shall continue to provide essential information, even in the era of genomics (Blair, 2023). By combining decades of barcoding efforts in a meta-analysis framework, we revisit the diversification of Palearctic water frogs in space and time to appraise its evolution, diversity and biological invasions. This overview should motivate future surveys to illuminate the evolutionary relevance of the documented phylogroups, their systematics, their interactions in secondary contact zones, and hopefully contributes to a better consideration of their emerging invasions by stakeholders. Our study also reaches the limits of molecular analyses based on a few genes, and further investigations should rely on genomic data, starting by a phylogenomic species tree and multilocus analyses of hybridization and gene flow in hotspots of phylogeographic diversity for the genus, namely the Balkans, Anatolia, Central Asia, Japan and mainland China. As its complexity unfolds after over half a century of forensic research (Berger, 1967), Pelophylax hybridogenesis still holds many fascinating secrets that are bound to challenge our perception of biological notions as fundamental as sexual reproduction and speciation.

AUTHOR CONTRIBUTIONS

Christophe Dufresnes: Conceptualization; data curation; formal analysis; investigation; methodology; writing – original draft.
Benjamin Monod-Broca: Data curation; investigation. Adriana Bellati: Resources; writing – review and editing. Daniele Canestrelli: Resources; writing – review and editing. Johanna Ambu: Investigation. Ben Wielstra: Conceptualization; writing – review and editing. Sylvain Dubey: Resources; writing – review and editing. Pierre-André Crochet: Conceptualization; resources; writing – review and editing. Daniel Jablonski: Conceptualization; investigation; resources; writing – review and editing.

ACKNOWLEDGEMENTS

We are thankful to G. Mazepa and S. N. Litvinchuk for assistance with some of the Pelophylax barcoding literature, as well as P. Geniez for curating new samples. M.D. is a Research Director at F.R.S.–FNRS (Fonds de la Recherche Scientifique). This study was partly funded by a PDR FNRS grant (T.0070.19 to M.D.) and a RFIS-NSFC grant (3211101356 to C.D.).

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

All datasets and metadata have been deposited on the Zenovo platform (Dufresnes, 2024).

ORCID

Christophe Dufresnes https://orcid.org/0000-0002-8497-8908
Benjamin Monod-Broca https://orcid.org/0000-0002-9133-5548
Adriana Bellati https://orcid.org/0000-0002-7259-7569
Daniele Canestrelli https://orcid.org/0000-0001-9351-4972
Johanna Ambu https://orcid.org/0000-0002-8072-7677
Ben Wielstra https://orcid.org/0000-0002-7112-5965
Sylvain Dubey https://orcid.org/0000-0002-0422-3960
Pierre-André Crochet https://orcid.org/0000-0003-4342-9432
Mathieu Denoël https://orcid.org/0000-0002-3586-8323
Daniel Jablonski https://orcid.org/0000-0002-5394-0114

REFERENCES


SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.