



Nuclear phylogenies and genomics of a contact zone establish the species rank of *Podarcis lusitanicus* (Squamata, Lacertidae)

Guilherme Caeiro-Dias^{a,*,1}, Sara Rocha^{b,c}, Alvarina Couto^a, Carolina Pereira^a, Alan Brelsford^{d,e}, Pierre-André Crochet^f, Catarina Pinho^a

^a Centro de Investigação em Biodiversidade e Recursos Genéticos, CIBIO/InBio, Vairão, Portugal

^b CINBIO, Universidade de Vigo, 36310 Vigo, España

^c Instituto de Investigación Sanitaria Galicia Sur, Hospital Álvaro Cunqueiro, 36213 Vigo, España

^d Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

^e Biology Department, University of California Riverside, CA, USA

^f CEFE, CNRS, Univ Montpellier, EPHE, IRD, Montpellier, France

ARTICLE INFO

Keywords:

Podarcis guadarramae
Reproductive isolation
Interspecific gene flow
Systematics

ABSTRACT

Unravelling when divergent lineages constitute distinct species can be challenging, particularly in complex scenarios combining cryptic diversity and phylogenetic discordances between different types of molecular markers. Combining a phylogenetic approach with the study of contact zones can help to overcome such difficulties. The *Podarcis hispanicus* species complex has proven to be prosperous in independent evolutionary units, sometimes associated with cryptic diversity. Previous studies have revealed that one of the species of this complex, *P. guadarramae*, comprises two deeply divergent yet morphologically indistinguishable evolutionary units, currently regarded as subspecies (*P. g. guadarramae* and *P. g. lusitanicus*). In this study we used molecular data to address the systematics of the two lineages of *Podarcis guadarramae* and the closely related *P. bocagei*. Firstly, we reconstructed the species tree of these three and two additional taxa based on 30 nuclear loci using the multispecies coalescent with and without gene flow. Secondly, we used SNPs obtained from RADseq data to analyze the population structure across the distribution limits *P. g. lusitanicus* and *P. g. guadarramae*, and for comparison, a contact zone between *P. bocagei* and *P. g. lusitanicus*. Nuclear phylogenetic relationships between these three taxa are clearly difficult to determine due to the influence of gene flow, but our results give little support to the monophyly of *P. guadarramae*, potentially due to a nearly simultaneous divergence between them. Genetic structure and geographic cline analysis revealed that the two lineages of *P. guadarramae* replace each other abruptly across the sampled region and that gene flow is geographically restricted, implying the existence of strong reproductive isolation. *Podarcis bocagei* and *P. g. lusitanicus* show a similar degree of genetic differentiation and reproductive isolation, with very low levels of admixture in syntopy. These results support that all three forms are equally differentiated and reproductively isolated. In consequence, we conclude that the two former subspecies of *Podarcis guadarramae* constitute valid, yet cryptic species, that should be referred to as *P. lusitanicus* and *P. guadarramae*.

1. Introduction

Molecular tools have been very useful to understand the origin and patterns of geographical variation of biodiversity, especially when combined with information on past geological events and bioclimatic changes (Gómez and Lunt, 2007; Schmitt, 2007). However, disentangling when divergent lineages constitute distinct species is still often

a challenge, especially in groups characterized by cryptic diversity or when cytonuclear discordances are observed (Eto et al., 2012; Phuong et al., 2014). To overcome such difficulties, genome-wide data and methods that allow phylogenetic inference under different models, naturally accounting for genomic heterogeneity and gene flow, represent a powerful alternative to the traditionally used mitochondrial markers and/or small number of nuclear loci (Collinson et al., 2017; Eto

* Corresponding author.

E-mail address: gcaeiroidias@unm.edu (G. Caeiro-Dias).

¹ (Present Address) Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, USA.

<https://doi.org/10.1016/j.ympev.2021.107270>

Received 24 August 2020; Received in revised form 19 May 2021; Accepted 26 July 2021

Available online 2 August 2021

1055-7903/© 2021 Elsevier Inc. All rights reserved.

and Matsui, 2014; Phuong et al., 2017). In addition, genome-wide analyses of patterns of recent admixture between abutting lineages can offer a direct test of reproductive isolation (Dufresnes et al., 2020), producing robust insights about species formation and delimitation.

The use of molecular markers has allowed to reassess the amount of species-level diversity in southern European refugia and to link the origin of this diversity to past climatic fluctuations and the complex topography of these regions (Abellán and Svenning, 2014; Gómez and Lunt, 2007). For example, a number of studies published during the past two decades have revealed that wall lizards inhabiting North Africa and Iberia, long classified as *Podarcis hispanicus* (Steindachner, 1870), constitute an assemblage of several genetically distinct lineages (Busack et al., 2005; Harris et al., 2002; Harris and Sá-Sousa, 2002, 2001; Kaliontzopoulou et al., 2011b; Lima et al., 2009; Oliverio et al., 2000; Pinho et al., 2008, 2007, 2006; Renoult et al., 2009; Sá-Sousa, 2000), many of which are now recognized as valid species (Busack et al., 2005; Geniez et al., 2014; Renoult et al., 2010; Speybroeck et al., 2020). This model has proven fruitful in uncovering cryptic diversity linked to independent evolutionary units (Kaliontzopoulou et al., 2011b) and documenting how cytonuclear discordances may arise (Renoult et al., 2009).

The first lineage to be widely recognized as a distinct species was *Podarcis bocagei* (Seoane 1884), the form inhabiting the north-western Iberian Peninsula, because it is sympatric with *P. gadarramae lusitanicus*, yet ecologically and morphologically distinct (Arnold et al., 1978). It took another twenty years for the form occurring along the western and south-western Iberian Atlantic coast (including the Berlengas islands, Portugal) and in the western Iberian Central System, described as a subspecies of *P. bocagei* as recently as 1981, to be elevated to species rank by Sá-Sousa (2001) and Harris & Sá-Sousa (2002) as *Podarcis carbonelli* (Pérez-Mellado, 1981).

The two previous species were the first to be recognized as specifically distinct from *Podarcis hispanicus* (*sensu lato*) because they are both widely sympatric with one or several other forms of the *P. hispanicus* complex. The other species of the complex all have mainly parapatric or allopatric ranges and were thus interpreted as geographical variations of *P. hispanicus* for much longer (Mellado and Villardón, 1986; Salvador, 2000, 1986). The form occurring in the Baetic mountains and adjacent areas south of the Guadalquivir river was split from *P. hispanicus* and assigned to *Podarcis vaucheri* (Boulenger, 1905), a name previously established for the North-African populations from Morocco and Algeria (Oliverio et al. 2000; Busack et al. 2005). Busack et al. (2005) formally raised the north-eastern Iberian form that extends as far as southern France and occurs also on the Columbretes islands, Spain, to species rank but failed to determine its valid name. This was later solved by Renoult et al. (2010) who established that this species should be called *Podarcis liolepis* (Boulenger, 1905). The name *Podarcis hispanicus* (*sensu stricto*) was restricted to the lineage from south-eastern Spain by Geniez et al. (2007) on the basis of morphology but was not formally treated as specifically distinct from all other lineages in Iberia until a last systematic revision assigned valid specific names to two additional forms: *Podarcis virescens* (Geniez, Sá-Sousa, Guillaume, Cluchier & Crochet, 2014) and *Podarcis gadarramae* (Boscá, 1916), distributed from central to north-western Iberian Peninsula. Recently, (Bassitta et al., 2020) proposed the split of *P. hispanicus* into two species based on genetic and morphological information but this study did not provide a critical examination of putative reproductive barriers in the group, as well as the role of mtDNA introgression on the populations' divergence, and thus its taxonomic implications should be taken as preliminary.

The first modern reference to the distinctiveness of the form corresponding to *P. gadarramae* was made by Guillaume & Geniez (1986) on the basis of morphological features only. It was later shown to constitute a distinct evolutionary lineage (Crespo et al., 2002; Harris and Sá-Sousa, 2001; Sá-Sousa, 2000) and referred to as *P. hispanicus* type 1 before its formal revalidation by Geniez et al. (2014). Further studies using allozymes and mtDNA sequences have revealed that *P. gadarramae*

contains two deeply divergent lineages (Pinho et al., 2008, 2007, 2006) corresponding to a mitochondrial divergence estimated at around five million years ago (Kaliontzopoulou et al., 2011b). These groups, that have never been reported in sympatry, were referred to as *P. hispanicus* types 1A and 1B, and were formally described by Geniez et al. (2014) as *P. gadarramae gadarramae* (former type 1B), inhabiting Spain, mostly along the Iberian Central System Mountains, and *P. gadarramae lusitanicus* (former type 1A), inhabiting northern Portugal and north-western Spain. However, the data available at the time were not conclusive regarding their systematic status. Mitochondrial DNA data identified these two lineages as sister-taxa, with marginally lower mitochondrial divergence than indubitably distinct species pairs like *P. carbonelli* and *P. virescens* (Kaliontzopoulou et al., 2011b; Pinho et al., 2006). On the contrary, nuclear DNA sequences showed a high level of genetic divergence between *P. g. gadarramae* and *P. g. lusitanicus*, with essentially no nuclear gene flow; these two lineages were not even forming a monophyletic group (Pinho et al. 2007, 2008). Since no clear morphological features were found to separate them, precluding the analysis of morphological intergradation in contact zones, Geniez et al. (2014) cautiously suggested maintaining both *P. gadarramae* lineages as subspecies until further data would allow clarifying their evolutionary relationships and examining their interactions in contact zones.

With this study, our goal is to address the systematics of the two lineages of *Podarcis gadarramae*. We first used new sequence data from multiple nuclear loci to reconstruct species trees representing their phylogenetic relationships and test whether the two taxa represent a monophyletic group as suggested by previous studies on mtDNA. We also estimated a species tree based on the mtDNA locus alone. We sampled the margins of the known distributions of *P. g. lusitanicus* and *P. g. gadarramae* to locate a putative contact zone and assess their reproductive isolation based on population genomics analyses. For reference purpose, we also analysed a contact zone between syntopic *P. g. lusitanicus* and *P. bocagei* to formally test for reproductive isolation between these two taxa. We demonstrate that *P. g. lusitanicus* and *P. g. gadarramae* are most often recovered as non-sister taxa by phylogenetic inference under different models using nuclear loci and that the splitting events between these two and *P. bocagei* were almost simultaneous. We also found that strong reproductive barriers are acting between them. Consequently, we formally propose to treat these two forms as distinct species.

2. Material and methods

2.1. Phylogenetic relationships between *P. g. gadarramae* and *P. g. lusitanicus*

2.1.1. Sampling, DNA extraction and amplification

For phylogenetic reconstruction, we used tissue samples from previously collected specimens from CIBIO's herpetological collection. We selected 7 to 12 specimens from each of five distinct taxa: *P. g. gadarramae*, *P. g. lusitanicus*, *P. bocagei*, *P. vaucheri* and *P. muralis* (45 samples in total, Table 1). *Podarcis bocagei* was selected because it was inferred to be the sister taxa to *P. gadarramae* in mtDNA-based phylogenetic reconstructions (e.g. Kaliontzopoulou et al., 2011b). *Podarcis vaucheri* and *P. muralis* were selected as outgroups; in both of these cases we used samples from single regions (North Africa and Spain, respectively) to avoid including highly distinct lineages (Kaliontzopoulou et al., 2011b; Salvi et al., 2013). Ongoing analyses with a larger dataset including 178 individuals representing all described species in the *P. hispanicus* complex (Couto et al. unpublished) consistently recover *P. bocagei*, *P. g. gadarramae* and *P. g. lusitanicus* as a monophyletic group (see also Bassitta et al., 2020; Yang et al., 2021). In all other taxa, samples were chosen to cover as much of their known distribution as possible. All of the samples analysed were previously assigned to species based on mitochondrial DNA sequencing (Caeiro-Dias et al., 2018; Kaliontzopoulou et al., 2011b; Pinho et al., 2008, 2007).

Table 1

Samples used for species tree reconstructions. All samples were previously assigned to species based on mtDNA (Caeiro-Dias et al., 2018; Kaliontzopoulou et al., 2011b; Pinho et al., 2008, 2007). Pb: *P. bocagei*, Pgl: *P. g. lusitanicus*, Pgg: *P. g. guadarramae*, PvMA: *P. vaucheri* Morocco/ Algeria lineage, Pm: *P. muralis*.

Sample Code	mtDNA	Locality	Region	Country	Latitude	Longitude
AK_3.120	Pb	Madalena	Porto	Portugal	41.10398	-8.66138
AK_3.166	Pb	Montesinho	Bragança	Portugal	41.97927	-6.79532
AK_3.221	Pb	Subportela	Viana do Castelo	Portugal	41.68743	-8.71812
DB8104	Pb	Sarria	Lugo	Spain	42.78333	-7.40000
AK_3.123	Pb	Vila Pouca de Aguiar	Vila Real	Portugal	41.44583	-7.67218
AK_3.341	Pb	Gerês	Braga	Portugal	41.71833	-8.16667
DB4292	Pb	Torneros de la Valdería	León	Spain	42.22542	-6.23940
AK_3.56	Pb	Gião	Porto	Portugal	41.31295	-8.69163
AK_3.281	Pb	Tanes	Astúrias	Spain	43.21116	-5.40253
DB8140	Pb	Taboadela	Ourense	Spain	42.23333	-7.81667
DB8415	Pgl	Los Ancares	León	Spain	42.66963	-6.72697
AK_5.23	Pgl	Moledo	Viana do Castelo	Portugal	41.83856	-8.87407
AK_5.143	Pgl	Alvão	Vila Real	Portugal	41.35000	-7.86667
AK_5.259	Pgl	Ledesma	Salamanca	Spain	41.09175	-5.99790
AK_5.150	Pgl	Vila de Rua	Viseu	Portugal	40.95000	-7.56667
AK_5.180	Pgl	Gerês	Braga	Portugal	41.71833	-8.16667
AK_5.225	Pgl	Tudera	Zamora	Spain	41.41689	-6.21043
DB8377	Pgl	Pendilhe	Viseu	Portugal	40.88333	-7.81667
DB8459	Pgg	Guadarrama	Madrid	Spain	40.68333	-4.08333
DB8903	Pgg	Torrejón de la Calzada	Madrid	Spain	40.20000	-3.80000
AK_5.194	Pgg	Ciudad Rodrigo	Salamanca	Spain	40.59295	-6.53633
AK_6.291	Pgg	Trujillo	Cáceres	Spain	39.460667	-5.881500
AK_5.206	Pgg	Alba de Tormes	Salamanca	Spain	40.825590	-5.515460
DB8446	Pgg	La Alberca	Salamanca	Spain	40.46667	-6.08333
DB8422	Pgg	Villacastín	Segóvia	Spain	40.78333	-4.41667
DB8614	Pgg	Arévalo	Ávila	Spain	41.06207	-4.72029
AK_7.300	PvMA	Midelt	Meknès-Tafilalet	Morocco	32.68237	-4.74265
AK_7.409	PvMA	Tislit Lake	Errachidia	Morocco	32.19640	-5.64293
AK_7.334	PvMA	Ketama	Al-Hoceima	Morocco	34.87823	-4.61087
DB1047	PvMA	Tizi-n-Tleta	Taroudannt	Morocco	30.78090	-7.64354
DB1449	PvMA	Ceuta	Ceuta	Spain	35.88827	-5.31616
AK_7.137	PvMA	Mischliffen	Meknès-Tafilalet	Morocco	33.40543	-5.10332
DB1140	PvMA	Imouzer-des-Glaoua	Tizi-n-Titchka	Morocco	31.30697	-7.36261
DB1587	PvMA	Imouzer Kandar to Annoceur	Fès-Boulemane	Morocco	33.62582	-4.89628
DB76	PvMA	Lac Iseli	Meknès-Tafilalet	Morocco	32.21647	-5.54972
AK_7.385	PvMA	Debdou	Oujda	Morocco	33.87247	-3.03878
AK_7.26	PvMA	Taza	Taza	Morocco	34.22119	-4.01586
AK_7.86	PvMA	N Oukaimeden	Marrakech	Morocco	31.20355	-7.86172
AK_2.100	Pm	Rio Segre	Lleida	Spain	42.36897	1.75992
DB1875	Pm	Ruta del Cares	Astúrias	Spain	43.25277	-4.84212
DB4296	Pm	Palacio del Compludo	León	Spain	42.45613	-6.45005
DB8980	Pm	Tanes	Astúrias	Spain	43.21117	-5.40253
DB4281	Pm	La Omanuela	León	Spain	42.77731	-5.97740
DB4294	Pm	Leon	León	Spain	42.59016	-5.57659
DB4288	Pm	Valdehuesa	León	Spain	42.94342	-5.31696

We selected 30 nuclear DNA loci to perform phylogenetic reconstruction: the 21 anonymous loci described in Pereira et al. (2013) plus 3 introns and 6 protein coding loci frequently used in studies of squamate phylogenetics. The list of loci and primers used in this study is provided in Appendix A, Table A1 and GenBank accession numbers can be found in Appendix A, Table A2. For *P. bocagei*, *P. vaucheri* and *P. muralis*, sequences were previously obtained and published (Andrade et al., 2019; Pereira et al., 2013), but all sequences of *P. g. guadarramae* and *P. g. lusitanicus* are new to this study. Genomic DNA from new samples was extracted using either EasySpin® Genomic DNA Tissue Kit (Citomed, Odivelas, Portugal) or QIAGEN® QIAmp Micro Kit (QIAGEN, Crawley, United Kingdom) following the supplier's protocol. The quality and quantity of extracted DNA was evaluated on 0.8% agarose gels and on a QUBIT 2.0 fluorimeter (Life Technologies, Grand Island, NY, USA). Amplification was carried out according to the protocol outlined in Pereira et al. (2013); this protocol was the same for all loci, except in the case of *Nfycint16* and *Rag2*, in which we used a nested PCR approach (although still using the same mix composition). Purification and Sanger sequencing of PCR products were carried out by Macrogen (<http://dna.macrogen.com/eng/>; Seoul, Korea) using the same primers used for amplification.

2.1.2. Alignments and post-processing

All gene sequences were checked, edited and then aligned using the program Sequencher v.4.1.4. (Gene Codes Corporation). Several sequences were heterozygous for insertion/deletion polymorphisms, and we used the method outlined by Flot et al. (2006) to resolve them. For the majority of the loci alignment was not trivial as they contained a considerable amount of indels, so we used the automated method implemented in PRANK v.1.40110 (Löytynoja and Goldman, 2008), that takes into account the evolutionary distances between sequences while also recognizing insertions and deletions as distinct evolutionary events. A few final adjustments, when considered necessary, were made by hand. Because many of the alignments had large regions with indels, plus some highly variable regions, which can be problematic for phylogenetic analyses, trimAL v.1.4 (Capella-Gutiérrez et al., 2009) was used to remove large indels and poorly aligned regions. The heuristic "automated1" option was used to automatically decide the best method among those available in the program to trim each specific alignment.

The Bayesian algorithm implemented in the program PHASE v.2.1.1 (Stephens et al., 2001) was used to recover gametic phases, assisted by the known haplotype phases determined using the Flot et al. (2006) method. All ingroup taxa were grouped in a combined dataset for phasing but outgroups were phased independently. The input files were prepared using DNAsp v.5.0 (Librado and Rozas, 2009) with minor

modifications by hand. Each dataset was analysed using the general model for recombination rate (-MR) (Li and Stephens, 2003) with 1000 steps for burnin, one of thinning interval and 1000 main iterations. Each analysis was repeated five times with different random seeds. We accepted reconstructions that were common to all 5 runs and had over 0.75 of posterior probability. All other ambiguities were considered as unresolved and replaced by “N” in the alignments.

Once clean alignments were obtained, we assessed whether data were affected by recombination using RDP v.3.44 (Martin et al., 2010) using three methods: RDP, GENECONV (Sawyer, 1989) and MaxChi (Smith and Smith, 1998). We used the option “*automask*” for an optimal recombination detection, setting the cut-off *p*-value to 0.001. Following author recommendations (Martin, pers. comm), if recombination was not inferred for the three methods simultaneously, we assumed recombination-free alignments.

2.1.3. Species-tree estimation (under multispecies coalescent and isolation-and-migration models)

Using the nuclear loci, we estimated the relationships between the different species (and subspecies of *P. gadarramae*) using the multispecies coalescent model as implemented in *BEAST in BEAST version 2.5.2 (Bouckaert et al., 2019), as well as using gene trees and ASTRAL-III (Rabiee et al., 2019; Zhang et al., 2018). Models for sequence evolution for each locus were estimated using jModeltest 2 (Darriba et al., 2012) and implemented accordingly (using the closest model available) in *BEAST. Substitution rates were co-estimated relative to locus Pod16 (fixed to 1) under strict-clock models. A Birth and Death prior was used for the species-tree and the “linear and constant root” model for the population sizes. Two independent runs were performed, each with 1.725×10^9 steps along the MCMC and sampling every 50,000 steps. After checking for stationarity and convergence of the runs with Tracer v.1.7.1 (Rambaut et al., 2018), the first 9,500 trees produced were discarded as burn-in for each run. The remaining 25,000 trees from each run were used (both independently and combined) to produce a consensus (maximum clade credibility) tree, with median node heights, using LogCombiner v.2.5.2 and TreeAnnotator v.2.5.2. From the combined 50,000 trees, the first 10,000 were also used to visualize tree congruency in DensiTree v.2.2.6 (Bouckaert et al., 2014), i.e., the consensus species-tree and the species-tree distribution. Additionally, we estimated the gene trees (and 100 bootstraps) for each locus using RaxML-NG (Kozlov et al., 2019) and used the ML trees (both the original ones as well as trees with branches with support lower than 10 collapsed, following Zhang et al. (2017)) to build an unrooted species tree in ASTRAL. Support was assessed by computing local posterior probabilities (Sayyari and Mirarab, 2016) and quartet frequencies were used to perform a polytomy test in the two internal branches (Sayyari and Mirarab, 2018).

The model assumed by these approaches to reconstruct species relationships, the multispecies coalescent, assumes that there is no gene flow between taxa. This is not an adequate model in the case of *Podarcis* wall lizards, in which past and present gene flow has been documented (Caeiro-Dias et al., 2020; Pinho et al., 2009, 2008; Yang et al., 2021), and may mislead species-tree estimation (Jiao et al., 2020). Therefore, we also reconstructed the species tree using the BEAST2 modules DENIM (Jones, 2019) and AIM (Müller et al., 2018), which incorporate both incomplete lineage sorting and gene flow in the model. Both implementations co-estimate migration rates, effective population sizes and species trees but differ in how gene flow is modelled, and in the amount of gene flow tolerated by the model (lower in DENIM). We ran DENIM with the same set up for site and clock models as described above. For coalescence and migration, we tested different priors on popPriorScale (0.005 and 0.05), as well as on GammaComponent.1, which establishes the prior for migration rates. The gamma parameter for migration rate is determined by shape \times scale. We left shape parameter as 1 in all analyses and tested 4 different values for scale (the equivalent to mean migration rates of 0.0001, 0.001, 0.01 and 0.1). We used the “flexible”

model (which allows for differences in migration rate along the tree) in final analyses, although we also tested the “simple” model in preliminary runs. In the “flexible” model we set the relatedness factor to 0.5 and the migration decay scale to 0.01, as suggested by Jones (2019). As recommended in the software manual, we turned CollapseWeight to zero and unchecked the estimate option for this parameter. We set the priors for CollapseWeight, bdcGrowthRate.t:Species and popPriorScale as recommended in the program manual, and left the other priors as default. Final runs included from 0.85 to 2.10×10^9 steps each, sampled every 50,000 steps, and we performed 2 – 5 replicates per data set. To visualize the resulting species trees, we proceeded in the same way as described above for *BEAST.

In AIM, models and clock-rates were implemented as for *BEAST and DENIM. Priors for migration were defined by two parameters: MigIndicatorSum.Species and migRates.Species, controlling respectively the number of migration routes (i.e., between how many “branches” is gene flow observed) and the amount of gene flow. A range of priors was tested in both parameters following the authors’ recommendations. For the migration routes we tested Poisson priors with Lambda = 0.693, 2 and 3, resulting in distributions, for which 50% of the probability is for 0, 2 and 3 active migration routes, with tails reaching larger values, respectively. These were all combined with priors for the amount of gene flow taken from a lognormal distribution with means 3.3 and 1.65, reflecting we expect about 1 in every 10 (3.3) or 20 (1.65) lineages to have a migration event over the course of the tree (implemented as advised in <https://taming-the-beast.org/tutorials/AIM-Tutorial/> based on the species tree height inferred in *BEAST and DENIM (0.03); Barido-Sottani et al., 2018). Inference was also made without data (from prior only) to test the effect of priors on final estimates. All models were run for over 500 million generations, sampling every 100,000. Tracer, TreeAnnotator and DensiTree were used to inspect convergence, posterior distributions and to summarize the trees. AIM annotator was used to obtain summaries of the ranked species trees, migration estimates and posteriors (<https://github.com/genomescale/starbeast2/tree/master/src/starbeast2/aimannotator>). Bayes Factors for non-zero migration distributions were calculated as in Müller et al. (2017), using the migration estimates of the most frequent ranked trees from the run with the highest posterior (after 25% burnin), computed as $BF = (p^*(1-q)) / ((q^*(1-p))$, with *p* being the posterior (number of estimates above zero) and *q* the prior support for gene flow (i.e., the Poisson distribution Lambda / the total number of possible migration routes; 3/32 in this case).

To assess if incomplete lineage sorting (ILS) could lead to conflict between the mitochondrial gene tree and the species tree, we also estimated a species-tree under the multispecies coalescent using one mtDNA gene only, a 661 bp ND4 gene fragment, from 7 to 12 individuals per species (Appendix A, Table A3) using *BEAST. Unlike the mtDNA gene tree, the inference of a species tree (even if based on a single gene) uses the MSC, therefore accounting for effective population size and for the possibility of incomplete lineage sorting (see recommendation for this analysis in Drummond and Bouckaert (2015)). Ploidy was set to 0.5; site-model was estimated in jModeltest2 as above (GTR + I); mean substitution and clock rates were fixed to 1 as we were not interested in dating the tree. A relaxed uncorrelated lognormal model was used for the clock, with the default unscaled standard deviation (0.3). The analytical population size integration was used for the population model and trees were inferred under both Yule and birth–death priors, using 50 million generations MCMC chains. BEAST co-estimates gene-tree and species-tree distributions, and we summarized both as above; we present them when we discuss the results for the mtDNA locus.

2.2. Assessment of reproductive isolation and current levels of gene flow between *P. g. lusitanicus*, *P. g. gadarramae* and *P. bocagei*

2.2.1. Sampling, RADseq sequencing and SNP dataset

To evaluate the current levels of gene flow between *P. g. gadarramae* and *P. g. lusitanicus* we sampled several populations around the known

limits of their distributions, between the summer of 2014 and the winter of 2015. Previous field work and sequencing of mtDNA had documented the presence of either lineage in two populations located 55 km apart on either side of the border between Portugal and Spain (populations SAB with *P. g. lusitanicus* and CRD with *P. g. guadarramae* in Fig. 1; Caeiro-Dias et al., 2018). We collected samples from these two populations and from six other localities in between them (contact zone 1: CZ1), for a total of 58 samples from eight locations (inset in Fig. 1 and Table 2). The sampling scheme aimed to capture all the individuals that were seen, avoiding bias towards sex or age. Lizards were captured with a noose and kept in individual cloth bags until they were processed. All individuals sampled were geo-referenced and photographed. A small tail tip was collected and immediately stored in 96% ethanol for subsequent DNA extraction. Most animals were released the same day in the place of capture.

We also analysed a contact zone between *P. g. lusitanicus* and *P. bocagei* (contact zone 2: CZ2) together with CZ1 to compare the strength of reproductive isolation and distinctiveness between the three species in the two contact zones. We used 83 individuals from a previously known contact zone (CZ2 in Fig. 1; Gomes et al., 2016; Kaliontzopoulou et al., 2012) and 20 reference individuals from 11 localities (the five green squares outside the inset and the six red triangles in Fig. 1; see also Table 2), retrieved from the collections of both CIBIO (Vairão, Portugal) and BEV (CEFE, Montpellier, France).

We performed genomic DNA extractions and checked DNA quantity and quality as explained above in Section 2.1.1. We obtained ddRAD sequence data using modifications to protocols from Parchman et al. (2012), Peterson et al. (2012) and Purcell et al. (2014). The complete protocol is described by Brelsford et al. (2016). Fifty-three samples were included in one library containing a total of 665 samples, as described in Caeiro-Dias et al. (2020). Five other samples were included in another library containing a total of 749 samples following the same protocol. Raw sequence reads are deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA665746. Raw sequences were demultiplexed and SNPs called and filtered as described in Caeiro-Dias et al. (2020). Briefly, individual raw reads were demultiplexed using the `process_radtags` module of Stacks v.2.2 (Catchen, 2013), the optimal *de novo* assembly parameters for our data set were tested following the protocol described in Rochette and Catchen (2017) adapted to Stacks version used, prior to final *de novo* read alignment. The resulting dataset after variants calling was filtered using the populations module from Stacks to remove those with >0.7 maximum observed heterozygosity and subsequently with `vcftools` v.0.1.15 (Danecek et al., 2011) to discard loci with depth coverage less than 8, with alleles with minimum frequency (*maf*) lower than 0.05 and present in less than 80% of the markers for each dataset. We then performed two additional filtering steps using a custom Python script (available at https://github.com/catpinho/filter_RADseq_data) to remove loci exhibiting >8 SNPs per RAD tag and to keep only one SNP *per* locus, choosing the SNPs that maximize the frequency differences between reference populations. Finally, individuals with >50% of missing data were discarded. A final general quality filter was performed by applying the same `vcftools` filters as before to ensure that the final SNPs fulfil such criteria after the last three filters. SNP calling and filtering processes were performed three independent times to obtain three distinct datasets: one including all samples from the three species together (complete dataset); a second one for the transect across *P. g. lusitanicus* and *P. g. guadarramae* distribution limits, including the known reference populations (inset in Fig. 1; CZ1 dataset), and a third one with samples from *P. g. lusitanicus* and *P. bocagei* contact zone including reference samples (from locations outside the inset in Fig. 1; CZ2 dataset).

2.2.2. Analysis of contact zones

The genomic variability among individuals was first assessed with a principal component analysis (PCA) on the SNP dataset in R v.4.0.3 (R Core Team, 2020) using the `adegenet` package v.2.0.1 (Jombart 2008;

Jombart and Ahmed 2011). We performed three PCA analyses, one for each dataset (complete, CZ1 and CZ2). As PCA cannot handle missing data, missing data values had to be replaced. The commonly used “mean method” replaces missing data using the mean allele frequencies estimated from the whole sample, which may increase similarity of individuals where the proportion of missing data is relatively high, mimicking the signal of introgression. To ensure that we are not artificially increasing the similarity of samples in the contact zones, we replaced missing data using the Breiman’s regression random forest algorithm (Breiman, 2001) implemented in R package `randomForest` v.4.6–14 (Liaw and Wiener, 2002). The values of our missing data in each of the three datasets were predicted from 500 independently constructed regression trees and 50 bootstrap iterations with default bootstrap sample size.

We then used Structure v.2.3.4 (Pritchard et al., 2000) to evaluate the proportion (*Q*) of each individual’s genome originating from each of the parental species using the three SNP datasets (complete, CZ1 and CZ2). We will refer to the proportion of *P. g. lusitanicus* (Q_L) to describe the results. Structure was run with $K = 3$ for the complete dataset and $K = 2$ for the CZ1 and CZ2 datasets since we were interested in detecting admixture between these species. Runs were performed five times independently with one million repetitions, a burn-in of 250,000, assuming admixture, independent allelic frequencies and with a prior of individual ancestry of 0.5, following Wang (2017). For analysis with the CZ1 and CZ2 datasets, 90% posterior probability intervals (CI) for each individual were also estimated in each independent run and used to discriminate parental genotypes (individuals with 90% CI of Q_L overlapping 0 or 1) and admixed genotypes (individuals with 90% CI of Q_L non-overlapping 0 or 1). Structure Harvester web v.0.6.94 (Earl and VonHoldt 2012) was used to visualize the likelihood of the data. Runs that maximized the likelihood for each dataset were retained and are presented in the Results section.

Lastly, we evaluated the strength of reproductive isolation and determined the centre (*c*) and width (*w*) of the hybrid zone by estimating the geographic cline for the hybrid index (*HI*). We used the R package HZAR (Derryberry et al., 2014), which provides functions to fit allele frequency or *HI* data to equilibrium geographic cline models (Barton and Gale, 1993; Gay et al., 2008; Szymura and Barton, 1991, 1986) using the Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm. For each sample location we calculated the distance from the SAB population. Those distances were estimated using the spherical law of cosines formula after a linear interpolation of latitudes and original longitudes, to create a linear transect between populations SAB and CRD. We used *Q* values estimated with Structure as *HI* and these values were fitted to 16 equilibrium geographic cline models using HZAR version 0.2–5 (Derryberry et al., 2014). All models estimated cline centre (as the distance from SAB population, *c*) and width ($1/\text{maximum slope}$, *w*). Additionally, distinct models could estimate different combinations of the distance from the cline centre to the tail (δ) and the tail slope (τ): no tail (none), right tail only (right), left tail only (left), symmetric tails (mirror), or both tails separately (both); and whether they estimate (free), did not estimate (none) or fix at 0 and 1 (fixed) the *HI* at cline ends (p_{\min} and p_{\max}). For this study right tail means that the model estimates δ and τ to southwest (in the direction of *P. g. lusitanicus*) and left tail to northeast (in the direction of *P. g. guadarramae*). We performed three independent runs of one million MCMC iterations, a burnin of 100,000 and sampling every 10 iterations for each model and checked for convergence. For each locus, the model with the lowest AIC score was selected as the best-fitting model. We then estimated the CI as the region delimited by the maximum and minimum values of the 95% credible cline region (Derryberry et al., 2014).

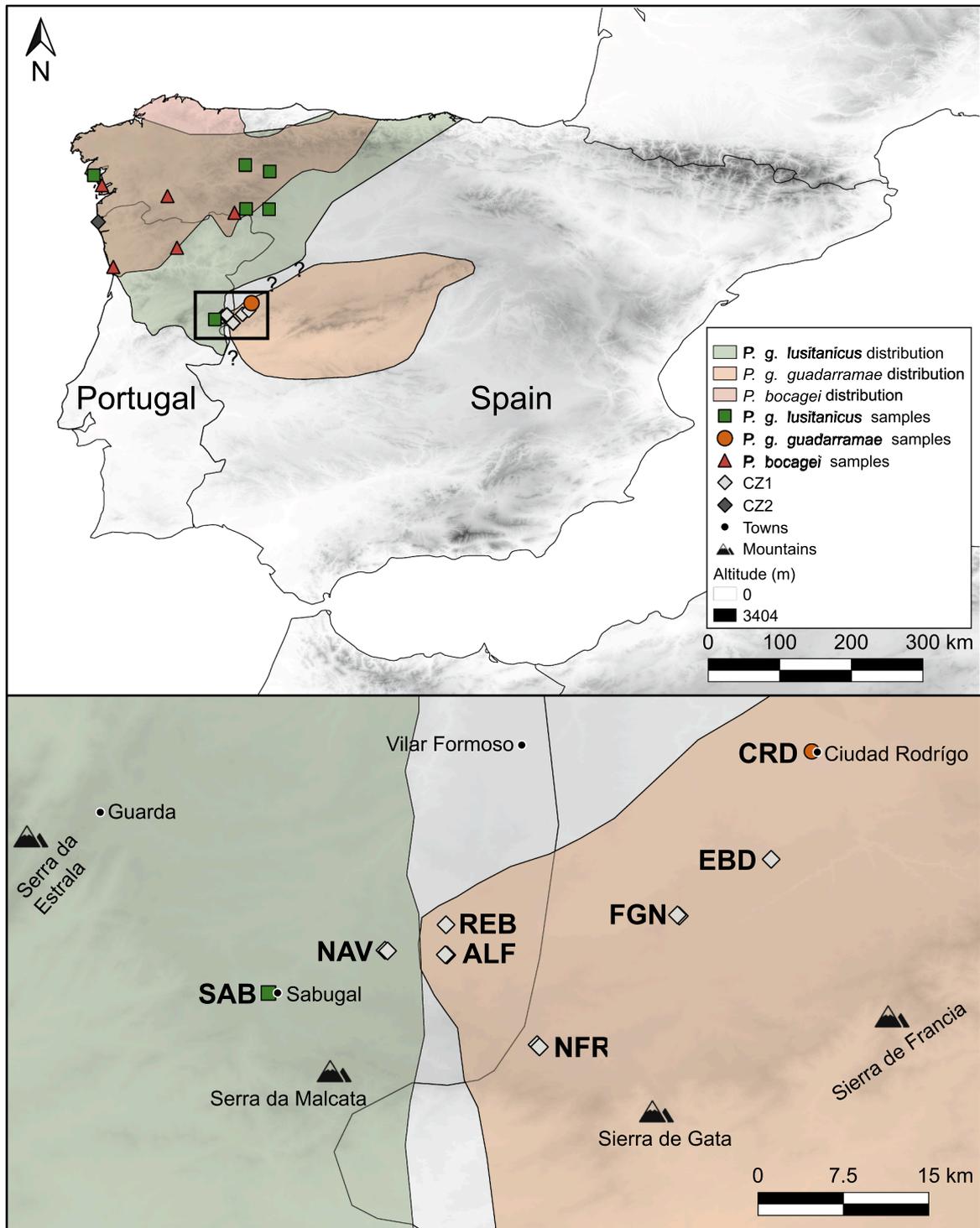


Fig. 1. Map of the Iberian Peninsula (on top) showing the known distribution (based on published and unpublished information) of *P. g. lusitanicus* (green), *P. g. guadarramae* (orange) and *P. bocagei* (red) and sampled locations for contact zone analyses. Green squares are reference *P. g. lusitanicus*, orange circle is the reference *P. g. guadarramae* and red triangles are reference *P. bocagei*. Dark grey diamond represents the location of the contact zone between *P. g. lusitanicus* and *P. bocagei*. Question marks between the distributions of *P. g. lusitanicus* and *P. g. guadarramae* denote areas where distribution boundaries are unknown (but see Caeiro-Dias et al., (2018) for more details on species distributions). Darker background represents higher altitudes. The inset (on the bottom) highlights the transect between *P. g. lusitanicus* and *P. g. guadarramae* (CZ1); showing in more detail sampled locations (grey diamonds). SAB: Sabugal, NAV: Nave, ALF: Alfaiates, REB: Rebolosa, NFR: Navasfrías, FGN: Fuenteguinaldo, EBD: El Bodón, CRD: Ciudad Rodrigo. SAB and CRD are the two locations where *P. g. lusitanicus* and *P. g. guadarramae*, respectively, were previously known. Small dots show the locations of main towns in the region, black triangles represent the main mountains and black line the border between Portugal and Spain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Number of individuals from each location included in the three SNP datasets (complete, CZ1 and CZ2) used for the analysis of population structure and contact zone between *P. g. lusitanicus* and *P. g. guadarramae* and the contact zone between *P. lusitanicus* and *P. bocagei*.

Location name (acronym)	Sample nr	Dataset	Country	Latitude	Longitude
Sabugal (SAB)	19/19	Complete/ CZ1	Portugal	40.35132	-7.09484
Nave (NAV)	1/1	Complete/ CZ1	Portugal	40.39528	-6.97480
Nave (NAV)	1/1	Complete/ CZ1	Portugal	40.39463	-6.97220
Alfaiates (ALF)	1/1	Complete/ CZ1	Portugal	40.39044	-6.91111
Alfaiates (ALF)	1/1	Complete/ CZ1	Portugal	40.39080	-6.91279
Rebolosa (REB)	1/1	Complete/ CZ1	Portugal	40.42154	-6.91203
Navafrias (NFR)	2/2	Complete/ CZ1	Spain	40.29816	-6.81823
Navafrias (NFR)	8/9	Complete/ CZ1	Spain	40.29585	-6.81586
Fuenteguinaldo (FGN)	6/6	Complete/ CZ1	Spain	40.42866	-6.67607
Fuenteguinaldo (FGN)	1/1	Complete/ CZ1	Spain	40.43170	-6.67377
El Bodón (EBD)	4/4	Complete/ CZ1	Spain	40.48896	-6.57727
Ciudad Rodrigo (CRD)	8/7	Complete/ CZ1	Spain	40.60006	-6.53548
Montesinho (MON)	2/2	Complete/ CZ2	Portugal	41.97927	-6.79532
Vila Pouca de Aguiar (VPA)	3/3	Complete/ CZ2	Portugal	41.44583	-7.67218
Porto (POR)	10/10	Complete/ CZ2	Portugal	41.15334	-8.64341
Taboadela (TAB)	2/2	Complete/ CZ2	Spain	42.23333	-7.81667
Sansexo (SAN)	1/1	Complete/ CZ2	Spain	42,40000	-8.81670
Moledo (MOL)	82/85	Complete/ CZ2	Portugal	41.83836	-8.87340
Isla de Rua (IDR)	5/5	Complete/ CZ2	Spain	42.55010	-8.94000
Santa Eulalia (SEU)	3/3	Complete/ CZ2	Spain	42.03222	-6.26830
Ungilbe (UNG)	2/1	Complete/ CZ2	Spain	42.03764	-6.61951
Ocera (OCE)	1/1	Complete/ CZ2	Spain	42.70600	-6.63000
La Silva (LSI)	1/1	Complete/ CZ2	Spain	42.60868	-6.25881

3. Results

3.1. Phylogenetic relationships between *P. g. lusitanicus*, *P. g. guadarramae* and *P. bocagei*

The 30 nuclear gene fragments were obtained for 45 individuals distributed across the distribution range of the three taxa (Appendix A, Table A2, see GenBank accession numbers therein). No robust evidence of recombination was found for any alignment (i.e., none presented positive results for the three methods used). For the analyses of nuclear loci under the multispecies coalescent (MSC), the two runs with *BEAST were highly consistent, producing species trees with identical topology. Effective Sample Sizes (ESSs) were >200 for most parameters in the model. The consensus tree (Fig. 2) recovers *P. muralis* as sister to all other taxa, as expected, followed by the split of *P. vaucheri*. *Podarcis guadarramae* *guadarramae* is the next split while *P. g. lusitanicus* is recovered as the sister taxa of *P. bocagei*, rendering *P. guadarramae* paraphyletic. Species trees certainty was high since all topologies matched the consensus topology and node lengths had little variation

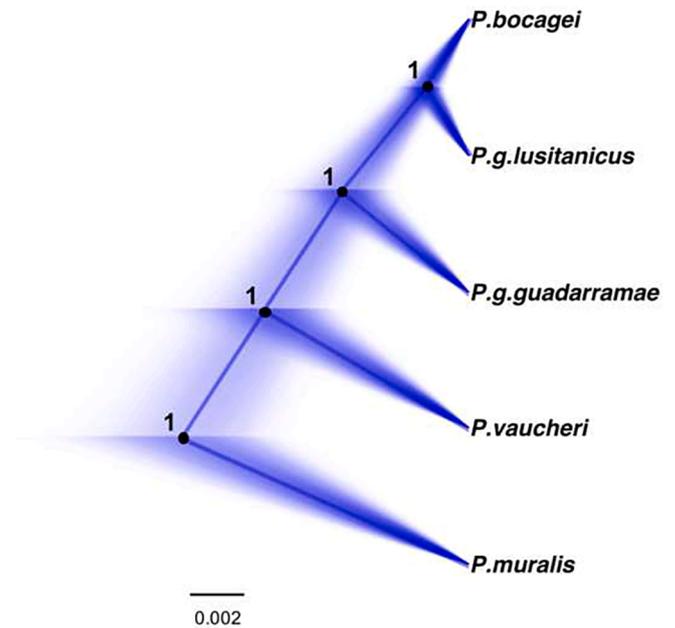


Fig. 2. Species tree distribution inferred with *BEAST based on 30 partial nuclear gene sequences of the five *Podarcis* species studied. The maximum clade credibility consensus tree after combining all the trees from each run is shown as a thicker line. Numbers at nodes represent Bayesian posterior probabilities. The trees represented with thinner lines represent 10,000 trees subsampled from the combined pool.

(Fig. 2). This was also the topology recovered by ASTRAL, using both sets of gene-trees, although in this case support for internal branches is much lower (Fig. 3). Importantly, the quartet support measures around the two internal edges (1 and 2 in Fig. 3) show some deviations from the proportions expected by ILS alone ($q_1 > 0.33$; $q_2 \sim q_3 < 0.33$; Sayyari and Mirarab, 2016), especially for edge 1, and a polytomy cannot be rejected according to the experimental test implemented in ASTRAL, with *p*-values of 0.68 and 0.69, respectively, using the ML trees.

When incorporating gene-flow, runs using the simple model in DENIM generally failed to converge independently of run length, so we focused on the more realistic flexible model. Using this model, preliminary analyses using *popPriorScale* = 0.05 or 0.005 generated similar patterns, so we chose 0.005 for subsequent runs. Patterns of run convergence were not similar for different migration priors. For migration priors of 0.0001 and 0.001 all the replicate runs converged, as shown by high ESS across all parameters (>200), except for the prior. For a migration prior of 0.01, three out of five replicates converged with similarly high ESS. For the highest migration prior - 0.1 - we failed to obtain any convergent run among the five trials and combining different runs did not improve the estimates. This may suggest that our data set is not informative enough for DENIM to explore scenarios of higher migration rates or that these values are already at the boundaries of the migration rates DENIM can work with, since the model that DENIM uses only holds for low migration (Jones, 2019). Although migration count estimates increased for most loci with increasing priors on migration rate, all the runs that converged returned a similar species tree, consistent with the species tree recovered with MSC methods. This may suggest that gene flow (e.g., between the sympatric *P. bocagei* and *P. g. lusitanicus*) is not the single factor responsible for the apparent paraphyly of *P. guadarramae*, but the impossibility of exploring scenarios of larger migration rates using this program makes us cautious about these results.

AIM analyses, however, resulted in different inferences: five of the six prior schemes converged within our computing time boundaries, with very good ESS's (in general > 300, only > 100 for a few specific parameters). Overall, from the prior combinations tested, for each pair

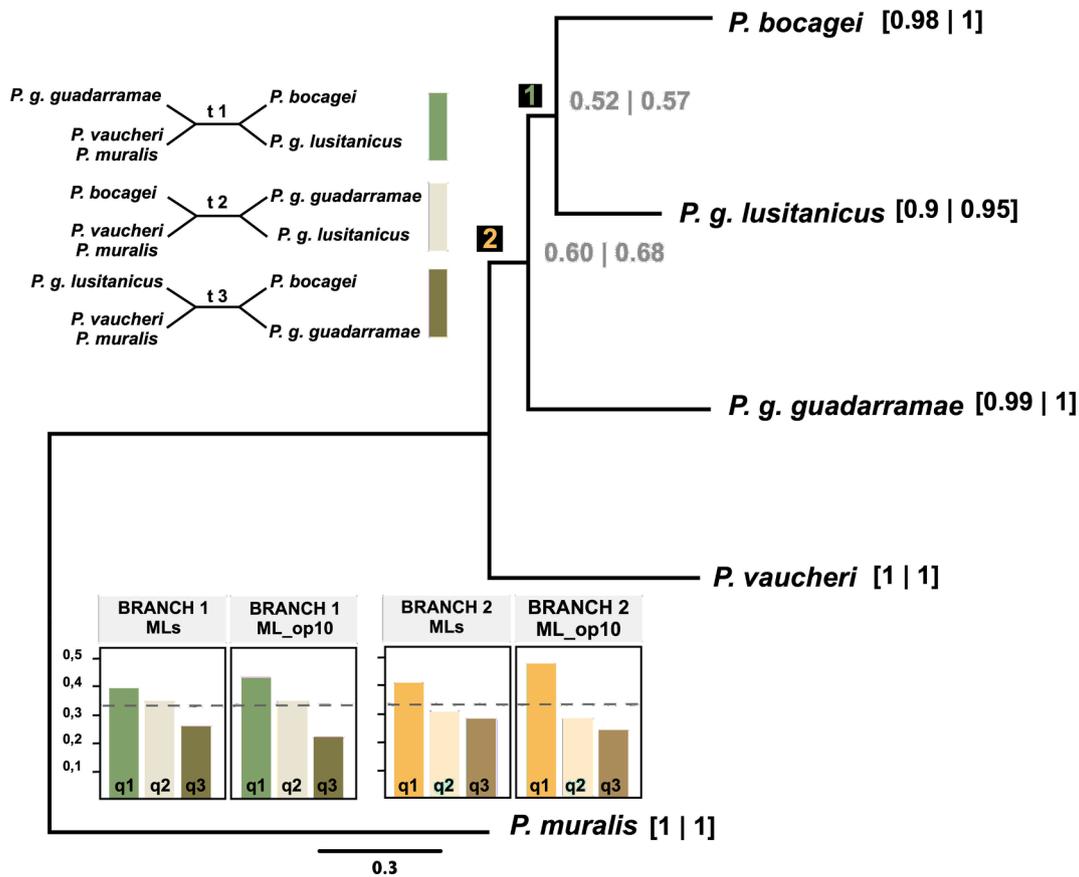


Fig. 3. MSC tree inferred with ASTRAL-III. Local posterior probabilities are shown for respective branches (1 and 2) and on tips for the terminal branches. For each internal edge, quartet support values are given for inferences using ML trees (left; "MLs") and using GTs obtained by collapsing nodes with support less than 10% (right; "ML_op10"). Quartet scores are presented for edges 1 and 2 (green and yellow bar-plot insets) and respective topologies t1, t2 and t3 are shown for edge 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of estimates, posteriors were higher for higher migration rate priors (see Appendix A, Table A4). ESS's on migration rates were always very high (thousands). Notably, in all cases, gene-flow was estimated to occur between exactly the same taxa and with similar posterior means: between *P. bocagei* and *P. g. lusitanicus* (in both ways, and these being the higher estimates), between *P. g. lusitanicus* and *P. g. gadarramae* (asymmetrical and much lower than the previous one), and also between *P. muralis* and both *P. g. gadarramae* and *P. vaucheri* (these possibly reflecting gene flow from other unsampled taxa or methodological artefacts; see Appendix A, Table A5 and Fig. 4).

The species-tree distribution always highlighted three possible topologies that occurred at roughly similar frequencies across runs (Fig. 4), with the most frequent one (by a small margin) making *P. g. gadarramae* sister to *P. bocagei*, followed by a topology where *P. gadarramae* is monophyletic, with the topology previously recovered by the MSC methods (*P. bocagei* sister to *P. g. lusitanicus*) occurring always at a slightly lower proportion than the previous two (Appendix A, Table A4). Notably, none of these topologies received good support (see Table A4). The species trees distribution, summary tree, and posterior probabilities for topology and migration are represented in Fig. 4 for the run with the highest posterior mean. In summary, AIM did not support a single particular topology with respect to *P. bocagei*, *P. g. gadarramae* and *P. g. lusitanicus*, suggesting instead near-simultaneous divergence between the three taxa.

The topology of the species-tree inferred under the MSC using the mtDNA locus alone is identical to the mtDNA gene-tree topology, with *P. gadarramae* monophyletic and *P. bocagei* sister to these, yet, support for the sister relationship between *P. g. lusitanicus* and *P. g. gadarramae*, when accounting for ILS (0.86), is slightly lower than the one from the

gene-tree (0.98 in this study; 0.95 in Kaliontzopoulou et al. (2011b)). Inferences under Yule and birth-death prior produced identical results and all runs converged easily with good ESS for all parameters (Appendix A, Fig. A1).

3.2. Assessment of reproductive isolation and currents levels of gene flow between *P. g. lusitanicus*, *P. g. gadarramae* and *P. bocagei*

The final SNP datasets included 5024 loci across 165 individuals for the complete dataset, 8233 loci across 53 individuals for the CZ1 dataset and 4405 loci across 114 individuals for the CZ2 dataset. The average coverage across individuals was 22.7 (ranging from 9.1 to 66.5) for the complete dataset, 24.0 (13.6 – 37.8) for CZ1 dataset and 22.3 (9.3 – 67.3) for CZ2. Across loci the coverage was 22.9 (12.1 – 91.6) for the complete dataset, 24.2 (11.1 – 151.2) for CZ1 and 22.5 (12.0 – 95.8) for CZ2. The maximum missing data per individual was 46% (complete dataset), 38% (CZ1) and 48% (CZ2).

The PCA performed on SNP data with the complete dataset (i.e., including samples from the three species) showed that *P. g. gadarramae* segregates from *P. g. lusitanicus* and *P. bocagei* along PC1, which explained 34.1% of the variance, while *P. g. lusitanicus* and *P. bocagei* are separated along PC2, that explains 18.2% of the variance (Fig. 5a), indicating that *P. g. lusitanicus* shares more alleles with *P. bocagei* than with *P. g. gadarramae*. Even though some individuals from each contact zone were slightly displaced from the main distribution of their taxon towards the other taxon, which indicates limited admixture, the three taxa remain clearly separated.

Both PCAs performed with CZ1 (Fig. 5b) and CZ2 (Fig. 5c) segregated samples from both species along PC1, which explained 45.2% (CZ1) or

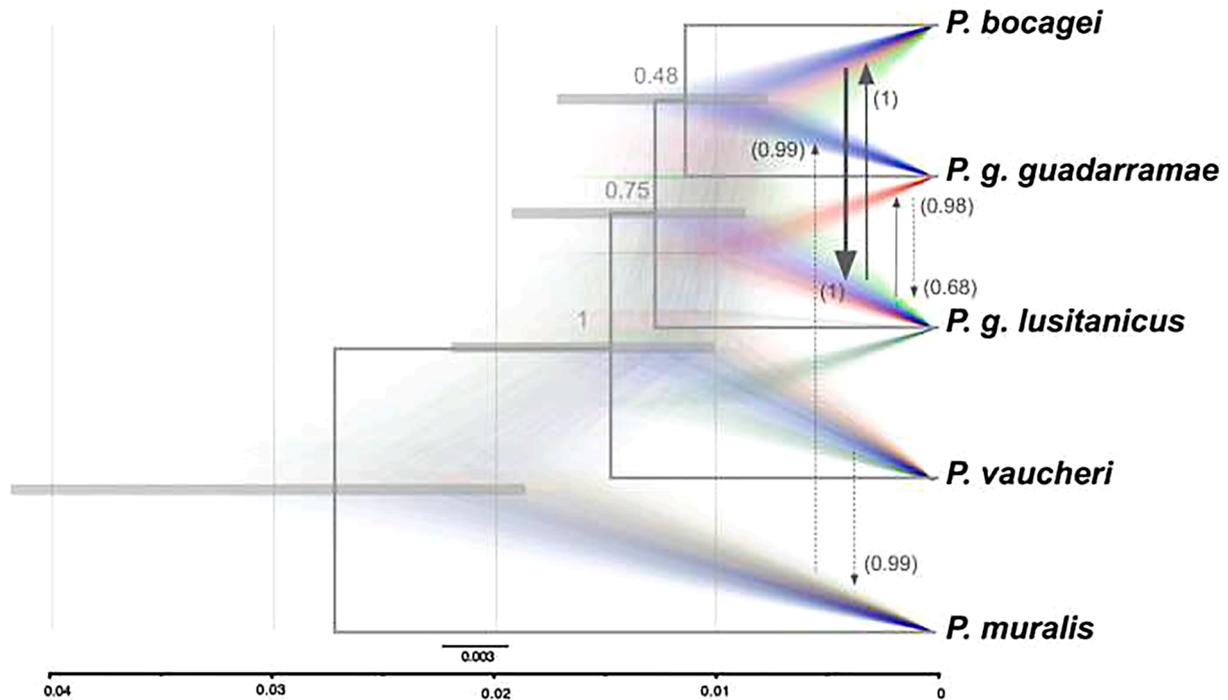


Fig. 4. Species tree distribution for the AIM-model run with higher posterior (see Table A4). Blue, red and green tree sets represent the three most common topologies (respectively, 31.02, 25.93 and 23.43% of the total trees). The green topology (less visible) corresponds to a sister relationship between *P. bocagei* and *P. g. lusitanicus*, with *P. g. gadarramae* sister to both. A summary tree (mean heights of all the trees with identical topology to the one with the maximum posterior probability) is shown in grey, which is also the topology of the MCC tree. Node bars represent node heights 95%HPD intervals and evidence the nearly simultaneous differentiation of the taxa in question. Arrows between branches of the ST distribution represent migration estimates with Bayes factors > 20 (see main text). Their posterior probabilities are given close to each arrow tip – thickness proportional to amount of gene-flow (see details in Appendix A, Table A5). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

31.1% (CZ2) of the variance (Fig. 5c and 5e, respectively). PC2 explained only 2.6% of the variance for CZ1 and 6% for CZ2 and captured in more detail (when compared to the PCA on the complete dataset) intraspecific variability within *P. g. gadarramae* and within *P. g. lusitanicus*, respectively. As shown before, in both contact zones there are a few individuals slightly displaced from the main distribution, although this was more evident in CZ1.

Structure results showed a very high proportion of assignment of each individual to one of the three species (Fig. 5b–d), most of them with the Q_L 90% CIs overlapping one or zero (Fig. 5c and d) which we consider as “pure” genotypes. In the transect across the *P. g. lusitanicus* and *P. g. gadarramae* distribution limits (CZ1) 46 individuals had a Q_L value of zero or one (Fig. 5b), with 29 individuals assigned to *P. g. gadarramae* and 19 to *P. g. lusitanicus* (Q_L 90% CIs overlapping zero or one, respectively). Five individuals had a genomic composition significantly different from the parental genotypes (i.e., 90% CIs did not overlap with either zero or one) but with Q_L close to zero or one. These five samples came from the locations NAV, ALF, REB and NFR, which are the populations that are closest to the centre of contact zone; this suggests that these non-parental genotypes represent a residual degree of admixture between both taxa. The closest locations where we sampled these five individuals were NAV – ALF (5 km apart) and NAV – REB (6 km apart). The furthest apart were NAV – NFR (17 km). No recently admixed individuals (F1, F2 or first-generation backcross) were identified (none of the individuals had Q_L between 0.25 and 0.75), which is not surprising given the lack of close geographic contact between the two species in our sampling.

The best-fitted model for the *HI* cline was the none/none model: no tails fitted and p_{\min}/p_{\max} not estimated ($AIC = 5.12$; Appendix A, Table A6). The genomic cline analysis based on the *HI* (Fig. 6) revealed a steep and narrow cline ($w = 3.93$ km; 1.23 – 11.25 km 95% CI) centred between populations NAV and ALF/REB ($c = 15.03$ km northeast of SAB;

12.06 – 17.42 km 95% CI).

4. Discussion

4.1. Nuclear phylogenies and analysis of contact zone support the species rank of *P. lusitanicus*

Previous studies have demonstrated a high degree of genetic differentiation between the three taxa that are the focus of this study (*P. bocagei*, *P. g. gadarramae* and *P. g. lusitanicus*; Pinho et al. 2007, 2008; Kaliontzopoulou et al., 2011b). However, an open question was whether the distinctiveness and degree of reproductive isolation between the two subspecies of *P. gadarramae* were large enough to warrant them species status, given their presumed sister taxa relationship and lack of obvious morphological differentiation (Geniez et al., 2014). On the contrary, *P. bocagei* and *P. g. lusitanicus* have long been treated as distinct species because they maintain their easily observable morphological differences in sympatry (e.g., Arnold et al., 1978; Galán, 1986) even if their reproductive isolation had not been formally assessed (but see Arntzen and Sá-Sousa (2007) and Pinho et al. (2007) for evidence of occasional admixture). Our results support the distinctiveness of the three forms and further suggest that *P. g. lusitanicus*, *P. g. gadarramae* and *P. bocagei* are best treated with the same status; we will thus refer to the two former subspecies of *P. gadarramae* as *P. lusitanicus* and *P. gadarramae* from here on.

Despite the phylogenetic relationships recovered by the mtDNA data Kaliontzopoulou et al., (2011b) and of their morphological similarity, nuclear phylogenies do not provide robust support for the monophyly of *P. gadarramae* as traditionally considered (although monophyly cannot be conclusively rejected either). Analyses based on the MSC and methods allowing for limited amounts of gene flow consistently recover a paraphyletic *P. gadarramae* (if *P. lusitanicus* is considered as a

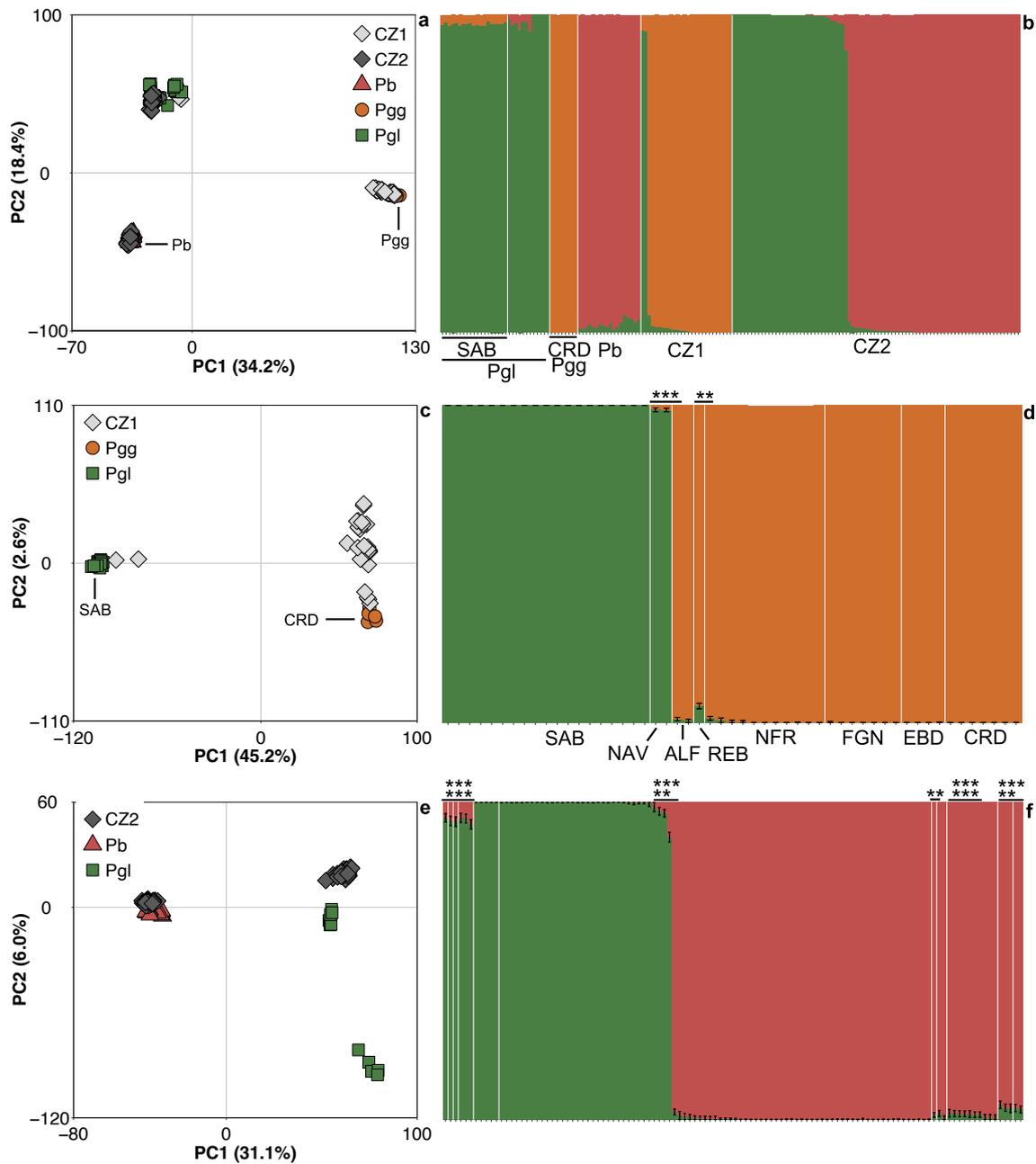


Fig. 5. Analysis of genetic variability and admixture based on the complete (4231 SNPs), CZ1 (8233 SNPs) and CZ2 (4828) datasets. **a, c, e**) Principal Component Analysis (PCA) showing the variation explained by each axis (PC), respectively for the complete, CZ1 and CZ2 datasets; each individual is represented by one symbol; symbols and colours as in Fig. 1. **b, d, f**) Results from individual multilocus genotype clustering analyses performed with Structure, respectively for the complete, CZ1 and CZ2 datasets; each individual is represented as a horizontal bar partitioned into the three coloured segments ($K = 3$ in panel b) and two coloured segments ($K = 2$ in panels d and f); segment length is proportional to the assignment probability to each species (orange: *P. g. guadarramae* ancestry; green *P. g. lusitanicus* ancestry; red *P. bocagei* ancestry). Vertical white lines in panel b delimit reference samples from each contact zone (CZ1 and CZ2) and in panel d and f delimit each sampling location. Black vertical lines in each individual of panel d and f represents the Q_L 90% CI. Individuals with Q_L 90% CIs not overlapping zero or one are highlighted with an asterisk; locations acronyms as in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

subspecies), with *P. bocagei* sister to *P. lusitanicus*. Similar results were achieved by a another study, using high-throughput sequencing methods, that was carried out simultaneously to our work (Yang et al., 2021). Yet when we incorporate (increasing levels of) gene flow in species tree analyses (Fig. 4), the three possible diversification histories involving these taxa appear at similar frequencies in the posterior tree distribution. A consensual view could be, therefore, an almost simultaneous diversification of these three taxa, accompanied by post-divergence gene flow, in a scenario where the real phylogenetic relationships become extremely hard – if not truly impossible – to recover.

Interestingly, migration rate estimates are higher between *P. bocagei* and *P. lusitanicus* than between *P. guadarramae* and *P. lusitanicus*, again highlighting the evolutionary distinctiveness of the latter species pair. We may therefore speculate that the strongly supported sister taxa relationship between *P. bocagei* and *P. lusitanicus* inferred by MSC methods may be an artifact caused by post-divergence gene flow. Regarding mtDNA, in a scenario of nearly simultaneous divergence, ILS can cause discordance between the species tree and the mtDNA gene tree (McCracken and Sorenson, 2005; Nolen et al., 2020; e.g., Wang et al., 2018). We thus additionally inferred relationships based on a mtDNA

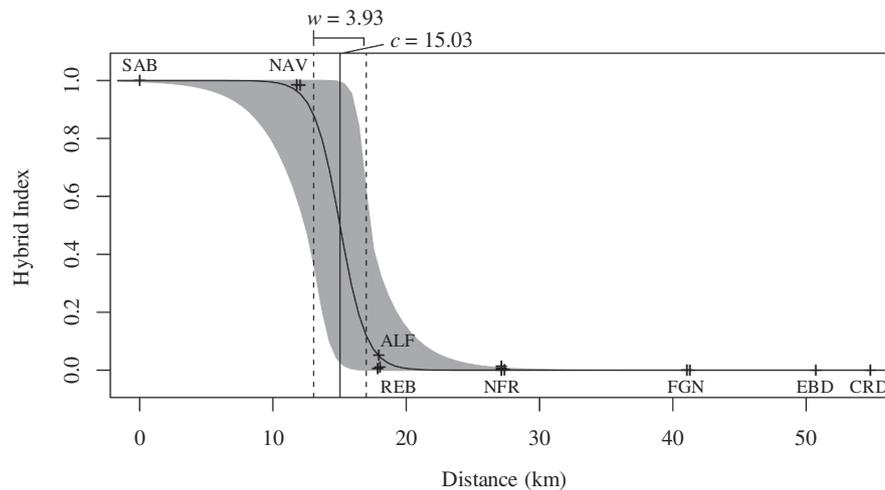


Fig. 6. Geographic cline analysis based on the hybrid index across the transition from *P. g. lusitanicus* ($HI = 1$) to *P. g. gadarramae* ($HI = 0$). The curve represents the estimated cline for the none/none model, and the shaded area is the 95% CI; the vertical solid line represents the centre (c) of the cline, the dashed lines the width (w) and crosses represent individuals from each locality (acronyms as in Fig. 1b).

locus under the MSC framework, to have a more realistic assessment of posterior clade supports. As expected, support for *P. gadarramae* and *P. lusitanicus* monophyly accordingly decreases when measured through the inferred species-tree distribution. We thus suggest that ILS could also explain the close relationship between *P. gadarramae* and *P. lusitanicus* recovered in mtDNA phylogenies. Further analyses will be needed to understand the role of ILS in mtDNA and/or nuclear gene flow in generating the phylogenetic discordance we recovered between these two types of markers.

In addition, analyses of genetic structure revealed that *P. lusitanicus*, *P. gadarramae* and *P. bocagei* have clearly distinct genomes. In the PCA analysis of the complete dataset, *P. gadarramae* is separated from both *P. lusitanicus* and *P. bocagei* along PC1 (Fig. 5a) denoting higher genomic similarity between *P. lusitanicus* and *P. bocagei* than between *P. lusitanicus* and *P. gadarramae*. This could be a consequence of a larger influence of gene flow in the genomic composition of the two sympatric species, as our analyses of nuclear genealogies seem to suggest. Also, the patterns of genomic distinctiveness in the PCAs of the CZ1 and CZ2 datasets are higher than population-level differentiation and similar to other analyses of pairs of indubitably distinct *Podarcis* species in Iberian Peninsula (Caeiro-Dias et al., 2020). Structure analyses of the three datasets detected only residual levels of introgression between species pairs. In the syntopic populations of *P. lusitanicus* and *P. bocagei* (CZ2) most genotypes are pure parental genotypes, with only 6% showing signs of residual admixture (i.e., the 90% Q_L CIs do not overlap 1 or 0).

The geographically limited occurrence of admixed individuals and narrow geographic clines (clines width less than 4 km, see results) imply either relatively strong reproductive isolation or a very recent zone of secondary contact. The last hypothesis seems unlikely since reported levels of admixture can be only achieved after several generations of recurrent backcrosses and/or reproduction between admixed genotypes. In addition, there is no evidence for recent changes in distribution of these species. Even if hybridization and admixture are common in the centre of the hybrid zone (which we did not locate nor sample yet), steep clines without introgression away from the contact zone demonstrates partial but strong pre- and/or post-zygotic isolation (Dufresnes et al., 2020; Jiggins and Mallet, 2000; Mallet et al., 1990). Further details on the evolutionary forces involved in the origin and maintenance of reproductive isolation between these two species would benefit from sampling the centre of the hybrid zone between populations NAV and REB/ALF (but also more populations further away from the contact zone and employing a combination of empirical assessment of pre-mating barriers and genomic and geographic cline analysis to assess the

strength and variation of post-zygotic barriers.

In terms of taxonomy, narrow stable hybrid zones are now typically interpreted as supporting species status (see e.g., Dufresnes et al., 2020; Speybroeck et al., 2020 and examples therein). Other examples in Iberian Lacertidae include the related (see Kaliontzopoulou et al., 2011b) *P. bocagei* and *P. carbonelli* which exhibit strong reproductive isolation but show signs of introgression across populations about 18 km apart (Pinho et al., 2009) or *Timon lepidus* and *T. nevadensis* which meet in a steep and narrow hybrid zone of an estimated width of 10 km with very limited signs of admixture in populations 25 km apart (Miraldo et al., 2013).

4.2. *Podarcis lusitanicus* and *P. gadarramae* are cryptic species

Despite their ecological divergence and long independent evolution, the morphological data available so far do not allow unambiguous separation of *P. lusitanicus* and *P. gadarramae* so these species are (based on present knowledge) real cryptic species. They differ on average in several characters, but these differences are not consistent, partly due to individual variation and partly due to extensive geographical variation in both species (Geniez et al., 2014). This is not unique, as the *Podarcis hispanicus* complex has long been used as an example of cryptic diversity (Kaliontzopoulou et al., 2011b). In this group, morphological variation does not always allow to unambiguously identify individuals to their respective species, and interspecific variation is often overwhelmed by local variation, suggesting that local adaptation and/or phenotypic plasticity are major drivers of morphological evolution (Kaliontzopoulou et al., 2011a). However, *P. lusitanicus* and *P. gadarramae* are the only species of the *Podarcis hispanicus* complex that even experienced observers cannot (yet) reliably distinguish based on morphology (pers. obs.). This is even more surprising as *P. bocagei* differs from both *P. lusitanicus* and *P. gadarramae* in coloration, morphology and pholidosis, lending support to the hypothesis that the similarity of *P. lusitanicus* and *P. gadarramae* may be due to evolutionary stasis maintaining the morphology of their common ancestor rather than to convergence and that *P. bocagei*, which is widely sympatric with *P. lusitanicus*, evolved into a distinct morphotype.

4.3. Conclusion

As anticipated by Geniez et al. (2014), the new data presented here support the view that the two former subspecies of *Podarcis gadarramae* constitute distinct species. Different phylogenetic approaches using

multilocus sequence data do not unequivocally support that of *P. lusitanicus* and *P. guadarramae* are monophyletic. Population structure and geographic clines analyses across their distribution boundaries suggests that contemporary gene flow between the two species is geographically restricted, suggesting the existence of strong reproductive barriers. Moreover, the levels of divergence and reproductive isolation between these two species are similar to those between *P. lusitanicus* and *P. bocagei*. We thus recommend that the two former subspecies of *P. guadarramae* should be viewed as distinct species and referred to as *Podarcis lusitanicus* and *Podarcis guadarramae*.

CRedit authorship contribution statement

Guilherme Caeiro-Dias: Investigation, Formal analysis, Writing – original draft. **Sara Rocha:** Formal analysis, Writing – original draft, Writing - review & editing. **Alvarina Couto:** Investigation, Formal analysis. **Carolina Pereira:** Investigation. **Alan Brelford:** Investigation. **Pierre-André Crochet:** Conceptualization, Data curation, Writing - review & editing, Supervision, Funding acquisition. **Catarina Pinho:** Conceptualization, Formal analysis, Data curation, Writing – original draft, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgements

We thank Antigoni Kaliontzopoulou and D. James Harris for providing samples for the phylogenetic analyses and Antigoni Kaliontzopoulou, Ana Perera, Isabel Damas, Beatriz Tomé, Catarina Rato, Isabel Tavares, Daniela Rosado, Miguel Carretero, David Cunha, Philippe Geniez and Olivier Chaline, who helped collecting the samples used for CZ2 dataset. We also thank Graham Jones and Nicola Müller for assistance in running and interpreting the results of DENIM and AIM, respectively. This study benefited from the Montpellier Bioinformatics Biodiversity platform supported by the LabEx CeMEB, an ANR “Investissements d’avenir” program (ANR-10-LABX-04-01). GCD was supported by a PhD grant (SFRH/BD/89750/2012), CP by IF contract (IF/01597/2014/CP1256/CT0009) under the Programa Operacional Potencial Humano—Quadro de Referência Estratégico Nacional funds from the European Social Fund and Portuguese Ministério da Educação e Ciência and PAC by the ANR grant ANR-19-CE02-0011 – IntroSpec. Support was also provided by national funds through FCT projects: PTDC/BIA-BEC/102179/2008 – FCOMP-01-0124-FEDER-007062, under FEDER COMPETE funds and PTDC/BIA-EVL/30288/2017 – NORTE-01-0145-FEDER-30288 co-funded by NORTE2020 through Portugal 2020 and FEDER Funds. Most specimens for this study were captured and handled under permit numbers EP/CYL/122/2011 by Dirección General del Medio Natural de la Consejería de Medio Ambiente (Junta de Castilla y León, Spain); 171-180/2012/CAPT, 114/2013/CAPT and 486-490/2014/CAPT by Instituto da Conservação da Natureza e das Florestas (ICNF, Portugal).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2021.107270>.

References

Abellán, P., Svenning, J.-C., 2014. Refugia within refugia—patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biol. J. Linn. Soc.* 113, 13–28.

- Andrade, P., Pinho, C., de Lanuza, G.P., Afonso, S., Brejcha, J., Rubin, C.-J., Wallerman, O., Pereira, P., Sabatino, S.J., Bellati, A., 2019. Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proc. Natl. Acad. Sci.* 116, 5633–5642.
- Arnold, E.N., Burton, J.A., Oviden, D., 1978. *Field guide to the reptiles and amphibians of Britain and Europe*. Collins.
- Arntzen, J.W., Sá-Sousa, P., 2007. Morphological and genetical differentiation of lizards (*Podarcis bocagei* and *P. hispanica*) in the Ria de Arosa archipelago (Galicia, Spain) resulting from vicariance and occasional dispersal. In: *Biogeography, Time, and Place: Distributions, Barriers, and Islands*. Springer, pp. 365–401.
- Barido-Sottani, J., Bošková, V., Plessis, L.Du., Kühnert, D., Magnus, C., Mitov, V., Müller, N.F., Pečerska, J., Rasmussen, D.A., Zhang, C., 2018. Taming the BEAST—A community teaching material resource for BEAST 2. *Syst. Biol.* 67, 170–174.
- Barton, N.H., Gale, K.S., 1993. Genetic analysis of hybrid zones. *Hybrid Zo. Evol. Process* 13–45.
- Bassitta, M., Buades, J.M., Pérez-Cembranos, A., Pérez-Mellado, V., Terrasa, B., Brown, R.P., Navarro, P., Lluch, J., Ortega, J., Castro, J.A., 2020. Multilocus and morphological analysis of south-eastern Iberian Wall lizards (Squamata, *Podarcis*). *Zool. Scr.* <https://doi.org/10.1111/zsc.12450>.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brelford, A., Dufresnes, C., Perrin, N., 2016. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity (Edinb.)* 116, 177.
- Busack, S., Lawson, R., Arjo, W., 2005. Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. *Amphibia-Reptilia* 26, 239–256. <https://doi.org/10.1163/1568538054253438>.
- Caeiro-Dias, G., Brelford, A., Kaliontzopoulou, A., Meneses-Ribeiro, M., Crochet, P.-A., Pinho, C., 2020. Variable levels of introgression between the endangered *Podarcis carbonelli* and highly divergent congeneric species. *Heredity* 1–14.
- Caeiro-Dias, G., Luís, C., Pinho, C., Crochet, P.A., Sillero, N., Kaliontzopoulou, A., 2018. Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex. *J. Zool. Syst. Evol. Res.* <https://doi.org/10.1111/jzs.12219>.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.
- Catchen, J.M., 2013. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. <https://doi.org/10.1111/mec.12354>.
- Collinson, J.M., Dufour, P., Hamza, A.A., Lawrie, Y., Elliott, M., Barlow, C., Crochet, P.-A., 2017. When morphology is not reflected by molecular phylogeny: the case of three ‘orange-billed terns’ *Thalasseus maximus*, *Thalasseus bergii* and *Thalasseus bengalensis* (Charadriiformes: Laridae). *Biol. J. Linn. Soc.* 121, 439–445.
- Crespo, E., Vicente, L., Sá-Sousa, P., 2002. Morphological variability of *Podarcis hispanica* (Sauria: lacertidae) in Portugal. *Amphibia-Reptilia* 23, 55–69.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772.
- Derryberry, E.P., Derryberry, G.E., Maley, J.M., Brumfield, R.T., 2014. HZAR: hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* 14, 652–663.
- Drummond, A.J., Bouckaert, R.R., 2015. *Bayesian evolutionary analysis with BEAST*. Cambridge University Press, p. 122.
- Dufresnes, C., Pribille, M., Alard, B., Gonçalves, H., Amat, F., Crochet, P.-A., Dubey, S., Perrin, N., Fumagalli, L., Vences, M., 2020. Integrating hybrid zone analyses in species delimitation: lessons from two anuran radiations of the Western Mediterranean. *Heredity* 124, 423–438.
- Earl, D.A., VonHoldt, B.M., 2012. Structure Harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361.
- Eto, K., Matsui, M., 2014. Cytonuclear discordance and historical demography of two brown frogs, *Rana tagoi* and *R. sakuraii* (Amphibia: Ranidae). *Mol. Phylogenet. Evol.* 79, 231–239.
- Eto, K., Matsui, M., Sugahara, T., Tanaka-Ueno, T., 2012. Highly complex mitochondrial DNA genealogy in an endemic Japanese subterranean breeding brown frog *Rana tagoi* (Amphibia, Anura, Ranidae). *Zool. Sci.* 29, 662–671.
- Flot, J., Tillier, A., Samadi, S., Tillier, S., 2006. Phase determination from direct sequencing of length-variable DNA regions. *Mol. Ecol. Notes* 6, 627–630.
- Galán, P., 1986. Morfología y distribución del género *Podarcis* Wagler, 1830 (Sauria, Lacertidae) en el noroeste de la Península Ibérica. *Rev. Esp. Herp* 1, 85–142.
- Gay, L., Crochet, P.A., Bell, D.A., Lenormand, T., 2008. Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution (N. Y.)* 62, 2789–2806. <https://doi.org/10.1111/j.1558-5646.2008.00491.x>.
- Geniez, P., Cluchier, A., Sá-Sousa, P., Guillaume, C.P., Crochet, P.A., 2007. Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) I: Redefinition, morphology and distribution of the nominotypical taxon. *Herpetol. J.* 17, 69–80.
- Geniez, P., Sá-Sousa, P., Guillaume, C.P., Cluchier, A., Crochet, P.A., 2014. Systematics of the *Podarcis hispanicus* complex (Sauria, Lacertidae) III: Valid nomenclature of the western and central Iberian forms. *Zootaxa* 3794, 1–51. <https://doi.org/10.11646/zootaxa.3794.1.1>.

- Gomes, V., Carretero, M.A., Kaliontzopoulou, A., 2016. The relevance of morphology for habitat use and locomotion in two species of wall lizards. *Acta Oecologica* 70, 87–95.
- Gómez, A., Lunt, D.H., 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia*. Springer, pp. 155–188.
- Guillaume, C.P., Geniez, P., 1986. Contribución a la biogeografía ya la sistemática de las lagartijas del género *Podarcis* en Península Ibérica y África del Norte, in: First National Congress of Herpetology, Benicàssim, 1–3 November 1986.
- Harris, D.J., Carranza, S., Arnold, E.N., Pinho, C., Ferrand, N., 2002. Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *J. Biogeogr.* 29, 1257–1262. <https://doi.org/10.1046/j.1365-2699.2002.00744.x>.
- Harris, D.J., Sá-Sousa, P., 2002. Molecular phylogenetics of Iberian wall lizards (*Podarcis*): is *Podarcis hispanica* a species complex? *Mol. Phylogenet. Evol.* 23, 75–81. <https://doi.org/10.1006/mpev.2001.1079>.
- Harris, D.J., Sá-Sousa, P., 2001. Species distinction and relationships of the western Iberian *Podarcis* lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. *Herpetol. J.* 11, 129–136.
- Jiao, X., Flouri, T., Rannala, B., Yang, Z., 2020. The Impact of Cross-Species Gene Flow on Species Tree Estimation. *Syst. Biol.* <https://doi.org/10.1093/sysbio/syaa001>.
- Jiggins, C.D., Mallet, J., 2000. Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15, 250–255.
- Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.
- Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071.
- Jones, G.R., 2019. Divergence estimation in the presence of incomplete lineage sorting and migration. *Syst. Biol.* 68, 19–31.
- Kaliontzopoulou, A., Adams, D.C., van der Meijden, A., Perera, A., Carretero, M.A., 2012. Relationships between head morphology, bite performance and ecology in two species of *Podarcis* wall lizards. *Evol. Ecol.* 26, 825–845.
- Kaliontzopoulou, A., Carretero, M.A., Llorente, G.A., 2011a. Morphology of the *Podarcis* wall lizards (Squamata: Lacertidae) from the Iberian Peninsula and North Africa: patterns of variation in a putative cryptic species complex. *Zool. J. Linn. Soc.* 164, 173–193. <https://doi.org/10.1111/j.1096-3642.2011.00760.x>.
- Kaliontzopoulou, A., Pinho, C., Harris, D.J., Carretero, M.A., 2011b. When cryptic diversity blurs the picture: A cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biol. J. Linn. Soc.* 103, 779–800. <https://doi.org/10.1111/j.1095-8312.2011.01703.x>.
- Kozlov, A.M., Darrriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455.
- Li, N., Stephens, M., 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165, 2213–2233.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R news* 2, 18–22.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Lima, A., Pinho, C., Larbes, S., Carretero, M., Brito, J.C., Harris, D.J., 2009. Relationships of *Podarcis* wall lizards from Algeria based on mtDNA data. *Amphibia-Reptilia* 30, 483–492. <https://doi.org/10.1163/156853809789647103>.
- Löytynoja, A., Goldman, N., 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320, 1632–1635.
- Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M., Eeley, H., 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124, 921–936.
- Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefevre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–2463.
- Maynard Smith, J., Smith, N.H., 1998. Detecting recombination from gene trees. *Mol. Biol. Evol.* 15, 590–599.
- McCracken, K.G., Sorenson, M.D., 2005. Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*)? *Syst. Biol.* 54, 35–55.
- Mellado, V.P., Villardón, M.P.G., 1986. Sistemática de “*Podarcis*” (Sauria lacertidae) ibéricas y norteafricanas mediante técnicas multidimensionales. Ediciones Universidad.
- Miraldo, A., Faria, C., Hewitt, G.M., Paulo, O.S., Emerson, B.C., 2013. Genetic analysis of a contact zone between two lineages of the ocellated lizard (*Lacerta lepida* Daudin 1802) in south-eastern Iberia reveal a steep and narrow hybrid zone. *J. Zool. Syst. Evol. Res.* 51, 45–54.
- Müller, N.F., Ogilvie, H.A., Zhang, C., Drummond, A., Stadler, T., 2018. Inference of species histories in the presence of gene flow. [bioRxiv 348391](https://doi.org/10.1101/348391).
- Müller, N.F., Rasmussen, D.A., Stadler, T., 2017. The structured coalescent and its approximations. *Mol. Biol. Evol.* 34, 2970–2981.
- Nolen, Z.J., Yildirim, B., Irisarri, I., Liu, S., Groot Crego, C., Amby, D.B., Mayer, F., Gilbert, M.T.P., Pereira, R.J., 2020. Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus* grasshoppers. *Mol. Ecol.* 29, 4985–5002.
- Oliverio, M., Bologna, M.A., Mariottini, P., 2000. Molecular biogeography of the Mediterranean lizards *Podarcis* Wagler, 1830 and *Teira* Gray, 1838 (Reptilia, Lacertidae). *J. Biogeogr.* 27, 1403–1420. <https://doi.org/10.1046/j.1365-2699.2000.00517.x>.
- Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W., Buerkle, C., 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* 21, 2991–3005.
- Pereira, C., Couto, A., Luís, C., Costa, D., Mourão, S., Pinho, C., 2013. Twenty-one new sequence markers for population genetics, species delimitation and phylogenetics in wall lizards (*Podarcis* spp.). *BMC Res. Notes* 6, 299.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7, e37135.
- Puong, M.A., Bi, K., Moritz, C., 2017. Range instability leads to cytonuclear discordance in a morphologically cryptic ground squirrel species complex. *Mol. Ecol.* 26, 4743–4755.
- Puong, M.A., Lim, M.C.W., Wait, D.R., Rowe, K.C., Moritz, C., 2014. Delimiting species in the genus *Otospermophilus* (Rodentia: Sciuridae), using genetics, ecology, and morphology. *Biol. J. Linn. Soc.* 113, 1136–1151.
- Pinho, C., Ferrand, N., Harris, D.J., 2006. Reexamination of the Iberian and North African *Podarcis* (Squamata: Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Mol. Phylogenet. Evol.* 38, 266–273. <https://doi.org/10.1016/j.ympev.2005.06.012>.
- Pinho, C., Harris, D.J., Ferrand, N., 2008. Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species. *BMC Evol. Biol.* 8, 63. <https://doi.org/10.1186/1471-2148-8-63>.
- Pinho, C., Harris, D.J., Ferrand, N., 2007. Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex the case of Iberian and North African wall lizards (*Podarcis*, Lacertidae). *Biol. J. Linn. Soc.* 91, 121–133.
- Pinho, C., Kaliontzopoulou, A., Carretero, M.A., Harris, D.J., Ferrand, N., 2009. Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: Evidence for limited natural hybridization and a bimodal hybrid zone. *J. Zool. Syst. Evol. Res.* 47, 368–377. <https://doi.org/10.1111/j.1439-0469.2009.00532.x>.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Purcell, J., Brelford, A., Wurm, Y., Perrin, N., Chapuisat, M., 2014. Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* 24, 2728–2732.
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL <https://www.R-project.org/>.
- Rabiee, M., Sayyari, E., Mirarab, S., 2019. Multi-allele species reconstruction using ASTRAL. *Mol. Phylogenet. Evol.* 130, 286–296.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901.
- Renoult, J.P., Geniez, P., Bacquet, P., Benoit, L., Crochet, P.A., 2009. Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Mol. Ecol.* 18, 4298–4315. <https://doi.org/10.1111/j.1365-294X.2009.04351.x>.
- Renoult, J.P., Geniez, P., Bacquet, P., Guillaume, C.P., Crochet, P.-A., 2010. Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) II: the valid name of the north-eastern Spanish form. *Zootaxa* 2500, 58–68.
- Rochette, N.C., Catchen, J.M., 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* 12, 2640.
- Sá-Sousa, P., 2001. Comparative chorology between *Podarcis bocagei* and *P. carbonellae* (Sauria: Lacertidae) in Portugal. *Rev. Española Herpetol.* 15, 85–97.
- Sá-Sousa, P., 2000. A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. *Herpetol. J.* 10, 1–11.
- Salvador, A., 2000. Fauna Ibérica. Vol. 10. Reptiles. Editorial CSIC Consejo Superior de Investigaciones Científicas.
- Salvador, A., 1986. *Podarcis hispanica* (Steindachner, 1870)—Iberische Mauereidechse. *Handb. der Reptil. und Amphib. Eur.* 2, 71–82.
- Salvi, D., Harris, D.J., Kaliontzopoulou, A., Carretero, M.A., Pinho, C., 2013. Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. *BMC Evol. Biol.* 13, 147.
- Sawyer, S., 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6, 526–538.
- Sayyari, E., Mirarab, S., 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes (Basel)* 9, 132.
- Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668.
- Schmitt, T., 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front. Zool.* 4, 11.
- Speybroeck, J., Beukema, W., Dufresnes, C., Fritz, U., Jablonski, D., Lymberakis, P., Martínez-Solano, I., Razzetti, E., Vamberger, M., Vences, M., 2020. Species list of the European herpetofauna—2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica. *Amphibia-Reptilia* 1, 1–51.
- Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989.
- Szymura, J.M., Barton, N.H., 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina orientalis* and *B. variegata*: comparisons between transects and between loci. *Evolution N. Y.* 45, 237–261.
- Szymura, J.M., Barton, N.H., 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina orientalis* and *B. variegata*, near Cracow in southern Poland. *Evolution* 40, 1141–1159.
- Wang, J., 2017. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Mol. Ecol. Resour.* 17, 981–990.

- Wang, K., Lenstra, J.A., Liu, L., Hu, Q., Ma, T., Qiu, Q., Liu, J., 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun. Biol.* 1, 1–9.
- Yang, W., Feiner, N., Pinho, C., While, G.M., Kaliontzopoulou, A., Harris, D.J., Salvi, D., Uller, T., 2021. Extensive introgression and mosaic genomes of Mediterranean endemic lizards. *Nat. Commun.* 12, 1–8.
- Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 153.
- Zhang, C., Sayyari, E., Mirarab, S., 2017. ASTRAL-III: increased scalability and impacts of contracting low support branches, in: *RECOMB International Workshop on Comparative Genomics*. Springer, pp. 53–75.