Comparative assessment of range-wide patterns of genetic diversity and structure with SNPs and microsatellites: a case study with Iberian amphibians

Miguel Camacho-Sanchez[1], Guillermo Velo-Antón[1], Jeffrey Hanson[1], Ana Verissimo[1], Inigo Martinez-Solano[2], Adam Marques[1], Craig Moritz[3], and Silvia Carvalho[1]

[1]CIBIO
[2]Museo Nacional de Ciencias Naturales
[3]Australian National University

July 22, 2020

Abstract

Reduced representation genome sequencing has popularized the application of single nucleotide polymorphisms (SNPs) to address evolutionary and conservation questions in non-model organisms. Patterns of genetic structure and diversity based on SNPs often diverge from those obtained with microsatellites to different degrees, but few studies have explicitly compared their performance under similar sampling regimes in a shared analytical framework. We compared range-wide patterns of genetic structure and diversity in two amphibians endemic to the Iberian Peninsula: *Hyla molleri* and *Pelobates cultripes*, based on microsatellite (18 and 14 loci) and SNP (15,412 and 33,140 loci) datasets of comparable sample size and spatial extent. Model-based clustering analyses with STRUCTURE revealed minor differences in genetic structure between marker types, but inconsistent values of the optimal number of populations (K) inferred. SNPs yielded more repeatable and less admixed ancestries with increasing K compared to microsatellites. Genetic diversity was weakly correlated between marker types, with SNPs providing a better representation of southern refugia and of gradients of genetic diversity congruent with the demographic history of both species. Our results suggest that the larger number of loci in a SNP dataset can provide more reliable inferences of patterns of genetic structure and diversity than a typical microsatellite dataset, at least at the spatial and temporal scales investigated.

Introduction

Nuclear microsatellites became popular during the 1990's as a powerful tool to assess patterns of genetic variation in populations (Allendorf, 2017; Ellegren, 2004). While they are still widely used, the development of Genotyping-by-Sequencing techniques, like RADseq (Baird et al., 2008; Miller, Dunham, Amores, Cresko, & Johnson, 2007) and similar techniques of genome complexity reduction (e.g. ddRAD, bestRAD), coupled with the decreasing costs of massive parallel sequencing, have extended the reach of massive single nucleotide polymorphism (SNP) genotyping to the study of non-model organisms (Baird et al., 2008; Davey et al., 2011; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012; Allendorf, 2017; Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Putman & Carbone, 2014). This has led to a discussion about the relative benefits of using each type of marker in conservation and evolutionary biology (Allendorf, 2017; Hodel et al., 2017; Morin, Luikart, & Wayne, 2004; Puckett, 2017).

Mutation rates in microsatellites are several orders of magnitude higher than those estimated for SNPs (Dallas, 1992; Ellegren, 2004; Lynch, 2010; Weber & Wong, 1993; Zhang & Hewitt, 2003). Combined with the

1

larger number of possible alleles for a single locus, microsatellites provide immense levels of polymorphism, yielding high statistical power in population genetic inference (Allendorf, 2017; Avise, 2004). Microsatellites are very sensitive to sudden, or recent, demographical processes, and are well suited to detect subtle population structure or recent bottlenecks (Haasl & Payseur, 2011; Luikart & Cornuet, 1998; Pereira, Teixeira, & Velo-Antón, 2018; Putman & Carbone, 2014). However, high polymorphism is usually associated with homoplasy (Garza & Freimer, 1996; Hedrick, 1999; Queney, Ferrand, Weiss, Mougel, & Monnerot, 2001), and poses difficulties in fitting adequate evolutionary models to heterogeneous mutation processes (Di Rienzo et al., 1994; Ellegren, 2004; Valdes, Slatkin, & Freimer, 1993; Weber & Wong, 1993; Webster, Smith, & Ellegren, 2002). This can lead to unreliable estimates of divergence times (Kalinowski, 2002; Queney et al., 2001) and underestimation of genetic differentiation between populations caused by high intra-populational heterozygosity (Hedrick, 1999). Furthermore, microsatellites are not well suited to reconstruct the evolutionary history of lineages or species under certain demographic scenarios, for instance during range expansions, when consecutive founder events and allele surfing processes in newly formed populations inflate genetic differentiation (Pereira et al., 2018). A microsatellite locus contains from four to twelve times more information than a SNP (Liu, Chen, Wang, Oh, & Zhao, 2005). However, current genotyping costs for SNPs are relatively low, so the lower per-locus information of SNPs is largely compensated by the sequencing of thousands of them at a similar cost than the genotyping of a few microsatellites (Hodel et al., 2016; Puckett, 2017). A large number of SNPs and their genome-wide distribution secure a range of mutation rates that can, in principle, provide sufficient information at different evolutionary scales, from recent demographic processes within-species to interspecies phylogenies (DeFaveri, Viitaniemi, Leder, & Merilä, 2013; Petersen et al., 2013).

The different molecular nature of SNPs and microsatellites is expected to impact their resolution power at different evolutionary scales, with microsatellites better reflecting recent demographic processes but rapidly losing resolution above the species level, and SNPs providing less information per locus but securing resolution of demographic processes over a wider evolutionary window (DeFaveri et al., 2013; Estoup, Jarne, & Cornuet, 2002; Haasl & Payseur, 2011). A review of the recent literature shows that thousands of SNPs are generally more powerful in detecting genetic structure than typical microsatellite datasets (Elbers, Clostio, & Taylor, 2017; Hodel et al., 2017; Jeffries et al., 2016; Malenfant, Coltman, & Davis, 2015; McCartney-Melstad, Vu, & Shaffer, 2018; Puckett, 2017; Puckett & Eggert, 2016; Rašić, Filipović, Weeks, & Hoffmann, 2014). The choice of marker (SNPs*versus* microsatellites) also seems to affect estimates of the proportions of individual ancestries and the inferred optimal number of clusters (Bradbury et al., 2015; Malenfant et al., 2015; Elbers et al., 2017; Bohling et al., 2019). These studies have made important contributions to our understanding of differences in patterns of genetic diversity and structure using both types of markers. However, the lack of comparable datasets, differences in the clustering methods used, and the absence of metrics allowing direct comparisons across marker types limit generalization of these results.

We present an explicit comparison of patterns of genetic structure and diversity based on comparable datasets of microsatellites and SNPs in two amphibian species: the Iberian tree frog, *Hyla molleri*Bedriaga, 1889, and the Western Spadefoot, *Pelobates cultripes*(Cuvier, 1829). Both are nearly endemic to the Iberian Peninsula (with some populations reaching southern France), and their range-wide phylogeography has been previously investigated based on mitochondrial and microsatellite datasets (Gutiérrez-Rodríguez, Barbosa, & Martínez-Solano, 2017; Sánchez-Montes, Recuero, Barbosa, & Martinez-Solano, 2019). These studies linked their contrasting phylogeographic patterns with different demographic histories during the Late Quaternary. *Hyla molleri* is present in Continental and Atlantic Iberia, and its higher tolerance to colder conditions was hypothesized to account for their inferred demographic stability since the Last-Glacial Maximum (~ 21,000 years ago) (Sanchez-Montes et al., 2019). In contrast, *P. cultripes* is a more thermophilous species present in southern and central Iberia, in areas with a Mediterranean influence. This species seems to have experienced important range contractions to southern glacial refugia during colder times in the Pleistocene, resulting in a south-to-north gradient of decreasing genetic diversity (Gutierrez-Rodriguez et al., 2017). The availability of comprehensive microsatellite datasets and the contrasting demographic histories in a shared geographical area make these two species good study systems for a robust comparative assessment of patterns of genetic diversity and structure obtained with microsatellites and SNPs.

Materials and Methods

We used published microsatellite datasets for *H. molleri* and *P. cultripes* (Gutierrez-Rodriguez et al., 2017; Sanchez-Montes et al., 2019) and generated SNP datasets for both species. Patterns of genetic structure between markers were compared based on model-based clustering analyses and those of genetic diversity were assessed with individual heterozygosity estimates.

Data collection

Samples from *H. molleri* and *P. cultripes* covered most of their current ranges (Figure S1; Table S1). They were evenly distributed across the main genetic clusters determined in previous works with microsatellites (Gutierrez-Rodriguez et al., 2017; Sanchez-Montes et al., 2019), securing the representation of more than 20 samples per north/south clusters. Microsatellite genotypes from *H. molleri* included 84 individuals from 25 localities genotyped at 18 loci (10% missing data) from Sanchez-Montes et al. (2019). Microsatellite genotypes from *P. cultripes* included 83 individuals from 43 localities genotyped at 14 loci (0 % missing data) from Gutierrez-Rodriguez et al. (2017). To facilitate comparisons between marker datasets, we selected the same 83 individuals of *P. cultripes* for SNP genotyping. However, in *H. molleri* only 39 individuals from the microsatellite dataset were amenable for SNP genotyping. In this case, we sampled additional individuals from the same or nearby locations as represented in the original microsatellite study to complete a dataset of 90 individuals from 25 localities (Table S1; Figure S1).

Genomic DNA was extracted with ExtractMe Genomic DNA 96-Well kits (DNA GDAŃSK), and concentrated with QIAamp DNA Micro (QIAGEN GmbtH) kits, when necessary. DNA extracts from *H. molleri* and *P. cultripes* were standardized to 500 ng of DNA (with exceptions as low as 390 ng) and sent for sequencing at Diversity Arrays Technology (Australia), which uses a proprietary protocol to sequence reduced representation of the genome from double-digested restriction fragments. We chose DArTseq because it has been reported to work well with large and complex genomes, like those of amphibians (Lambert, Skelly, & Ezaz, 2016). The restriction fragments generated were sequenced in an Illumina HiSeq 2500 as single-end reads of 77 nucleotides (nt). The sequencing depths for *H. molleri* and *P. cultripes* were 7.7 and 5 million reads per sample, respectively. Diversity Arrays Technology provides genotypes from the proprietary DArT-Soft14 pipeline in a text file along with several quality parameters on each SNP. Around 30% of the samples in the run are included as internal replicates to provide confidence levels on the genotype calls.

Data filtering

We applied several filtering steps to the SNP genotype matrices using R 3.6.0 (R Core Team 2019) functions from the *dartR* 1.1.11 package (Gruber, Unmack, Berry, & Georges, 2018) and custom code. The filters were applied as follows. First, we retained samples with a proportion of loci with calls (call rate per individual) greater than 0.35 and loci with high confidence on their genotype calls (*RepAvg* parameter from DArTseq greater than 0.95). We kept loci with balanced alleles (proportion of reads for each allele across samples between 0.15 and 0.85) and removed loci whose coverage was 3.5 times higher than the median coverage across loci to remove potential paralogs (O'Leary, Puritz, Willis, Hollenbeck, & Portnoy, 2018). Then, we removed loci with a call rate (proportion of samples with a call) lower than 0.8, retained only one SNP per contig (the one with greatest repeatability) and removed alleles with a frequency less than 0.02 (O'Leary et al., 2018) (Figure S2).

Genetic structure

We conducted model-based genetic structure analyses in STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000). For each dataset, we performed 10 replicate runs with values assuming a number of clusters (K) between 1 and 8 (K = 1 to K = 8), to encompass the optimal number of clusters (K = 2, K = 4 and K = 6) found in previous studies with microsatellites for the study species (Gutiérrez-Rodríguez et al., 2017;

Sánchez-Montes et al., 2019), and explore any potential finer substructure. We used an admixture model with correlated allele frequencies (Falush, Stephens, & Pritchard, 2003) with no prior information on sample origin. For the microsatellite data, we used the same run lengths as in the original publications: 500,000 burnin steps followed by 1,000,000 iterations. For the SNP datasets, run lengths were shorter as data chains often converge faster: 30,000 burnin steps followed by 10,000 iterations (Table S2; Figures S3.1 and S3.2). For the SNP runs, we estimated lambda with K = 1 by averaging lambda estimates across three replicate runs. These values of lambda (0.67 for *H. molleri* and 0.69 for *P. cultripes*) were then used across all runs of the SNP data, whereas for microsatellites lambda was fixed to 1. Lower values of lambda can improve the modeling of correlated allele frequencies when using SNPs, where often the data is skewed towards rare alleles (Falush et al., 2003). We ran STRUCTURE in parallel in 8 cores using *Structure_threader* (Pina-Martins, Silva, Fino, & Paulo, 2017), recording steps to log files every 50 and 5000 iterations for the SNP and microsatellite data, respectively.

Convergence between the 10 replicate runs for each K was evaluated using Gelman and Rubin's convergence diagnostic, GR (Gelman & Rubin, 1992), with function *coda* ::*gelman.diag* (Plummer, Best, Cowles, & Vines, 2006). Values below 1.05 indicate good convergence (Vats & Knudson, 2018). We used KFinder (Wang, 2019) to compare the best number of clusters for each dataset through three approaches: (1) Pr[X|K], the probability of data X given K clusters (Pritchard et al., 2000), (2) Evanno's ΔK, which considers the rate of change in the logarithm of the probability of data between successive K values (Evanno, Regnaut, & Goudet, 2005), and (3) PI, parsimony index, a newly proposed metric that favors K values yielding clusters with the most consistent and with minimal average individual admixture. The latter is assumed to be a more consistent metric across a wider range of demographic scenarios (Wang, 2019). We ran CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) on STRUCTURE outputs. CLUMPAK feeds the software CLUMPP with results of replicate runs for each K value to generate consensus solutions for the distinct modes. It also computes the similarity between Q-matrices (ancestry matrices) from each run and matches clusters across successive values of K.

STRUCTURE results were contrasted with a model-free hierarchical clustering method using the Neighbor-Joining algorithm on pairwise genetic distances (Supplementary File S1).

Congruence in ancestries between microsatellite and SNP datasets

We assessed the congruence of the Q-matrices from STRUCTURE results between SNP and microsatellite datasets using the Symmetric Similarity Coefficient (SSC) (Jakobsson & Rosenberg, 2007). For *P. cultripes* , since all individuals were identical among datasets, we ran CLUMPAK over the combined STRUCTURE results from both markers (n = 20 runs per K). The CLUMPP algorithm in CLUMPAK computes a pairwise distance matrix for all runs in each K based on the SSC. For *H. molleri* , since we sampled different individuals from the same localities for microsatellites and SNPs, we averaged individual ancestries per locality, and used R package *starmie* 0.1.2 (Tonkin-Hill & Lee, 2016) to run CLUMPP and compute the similarity coefficients. SSC ranges from negative values to a maximum of 1 when Q-matrices are identical. Pairwise SSCs were computed between runs from the same marker (SNPs-SNPs, microsatellites-microsatellites), in addition to cross-comparisons between markers (SNPs-microsatellites). To aid visualization of spatial patterns of genetic structure, we computed mean ancestries per locality for each species and marker from major clusters after CLUMPAK results. Then, for each species and K value, we aligned the microsatellite and SNP matrices using the CLUMPP algorithm from *starmie* 0.1.2 (Tonkin-Hill & Lee, 2016).

We evaluated admixture in individual ancestries of *P. cultripes* for each K in STRUCTURE using a newly developed index: the Coefficient of Admixture, CA. CA_{Ki} for individual *i* across clusters of a Q-matrix from a given K in STRUCTURE represent individual levels of genetic admixture, 0 indicating all ancestry belonging to a single cluster, and 1, equal proportions across clusters (details in Supplementary File S2).

4

Genetic diversity

Individual heterozygosity with each marker type was computed as the proportion of heterozygous loci standardized by the heterozygosity of loci across the dataset (standardized multilocus heterozygosity, sMLH) (Coltman, Pilkington, Smith, & Pemberton, 1999), using *inbreedR::sMLH* (Stoffel et al., 2016) in R. We then represented the median sMLH per locality for each dataset in a map to describe the spatial distribution of genetic diversity. Pearson correlations were computed between sMLH from microsatellite and SNP data, for individuals of *P. cultripes* . We also explored patterns of genetic diversity along the axes of demographic expansions from inferred glacial refugia in both species. For that purpose, the putative effects of latitudinal and longitudinal gradients on patterns of genetic diversity were assessed with linear models in R, using sMLH as a dependent variable and latitude and longitude as fixed effects.

Results

We produced panels of 15,412 SNPs (7.6 % missing data) for 90 individuals of *H. molleri* and 33,140 SNPs (5.2 % missing data) for 83 individuals of *P. cultripes* .

Genetic structure

STRUCTURE runs converged well for low K values but not for larger K values (Table S2; Figures S3.1 and S3.2). The best-supported number of genetic clusters (K) identified using STRUCTURE varied according to the metric used (PI or ΔK) and marker type. In most cases, we found the best support for two genetic lineages (K = 2), but some metrics identified further substructure, with up to six genetic clusters (K = 6) when using PI (Table S3; Figures S4.1 and S4.2).

Ancestries derived from both markers were spatially coherent at different K values. That is, individuals from the same or nearby localities shared similar ancestries and more admixed individuals coincided with geographical shifts in cluster assignment (Figure 1). For K = 2, both marker types were congruent in identifiying major subdivisions in each species: a northern and a southern lineage for *H. molleri* , and a central-western and a northeastern lineage for *P. cultripes* . From K = 3 to K = 8, the spatial patterns of genetic structure for both species were largely congruent between marker types in terms of admixture levels and ancestry group assignment (Figures 1 and S5). Both markers generally agreed on the genetic ancestry of localities or group of localities as sharing a singular genetic ancestry, although the K value at which for a given assignment to a cluster could differ between markers. For instance, for *H. molleri* , the western-coastal populations from Portugal (dark purple, Figure 1) formed a well-differentiated cluster at K = 3 with SNPs and at K = 4 with microsatellites. Another example is the locality Ojos de Villaverde, at the southeastern-most corner of the distribution of *H. molleri.* This locality appeared well differentiated at K = 4 for SNPs (green), but at K = 5 in microsatellites (magenta) (Figure 1). In *P. cultripes* , we observed the same phenomenon. For instance, the localities from northwestern Portugal were very differentiated at K = 4 with SNPs (green), but at K = 5 with microsatellites (green, Figure 1). Both markers agreed in localities within the northern half of the Iberian Peninsula with nearly "pure" ancestries and no further clustering after K = 4, and yielded very admixed localities in the southern half of Iberia from K = 4 to K = 8, although the levels of admixture and the ancestry assignments differed notably between markers. In *P. cultripes* , for K = 7 and K = 8, microsatellites yielded more admixed individual ancestries compared to SNPs (Figure S5), driven by the more admixed southern localities (Figure 1). For *H. molleri* , we could not quantify reliably these differences in admixture levels between markers because the individuals analyzed for each dataset were not all the same.

Genetic structure based on STRUCTURE analyses was highly congruent with that inferred by model-free hierarchical clustering (Supplementary File S1), which yielded well-supported clades for SNPs but less so in microsatellite-based topologies.

5

Congruence in individual/locality ancestries between microsatellites and SNPs

Both species showed higher intra-marker similarity (*H. molleri* , SSCs = 0.27 - 1.00; *P. cultripes* , SSCs = 0.77 - 1.00) than inter-marker similarity (*H. molleri* , SSCs = -0.03 - 0.42;*P. cultripes* , SSCs = 0.55 − 0.89) (Figure 2). For microsatellites, ancestries were very similar (SSCs close to 1) from K = 2 to K = 8 (except K = 7) for *H. molleri* and from K = 2 to K = 4 for *P. cultripes* . For SNPs, STRUCTURE results were almost identical only from K = 2 to K = 4 for *H. molleri* , but up to K = 6 for *P. cultripes* . Larger K values were in all cases associated with less consistent results across STRUCTURE runs. For most K values, pairwise SSC values in microsatellite runs had a larger spread (i.e. a greater range of values), especially at larger K values. This spread was minimum for STRUCTURE results derived from SNPs, though at larger K values (K = 4 to K = 8 for *H. molleri* ; K = 6 to K = 8 for*P. cultripes*) they tended to converge into 2 or even 3 regions of the parameter space (Figure 2). The similarity between SNP-microsatellite runs did not follow a clear pattern along increasing K. For *H. molleri* , SSCs were homogenously lower across all K values than for *P. cultripes* , highlighting the distinct solutions obtained between datasets. For this species, SSCs were maximum at K = 2 (0.89), and minimum at K = 4 (0.55). From K = 5 to K = 8, SSCs had a small increase in the 0.58 − 0.68 range.

Microsatellites yielded more admixed ancestries at larger values of K (i.e. K = 7 and K = 8; Figure S5) which seem to be driven by the more complex patterns of genetic structure in the southern localities (Figure 1).

Genetic diversity

Correlation of genetic diversity between microsatellites and SNPs-based measures was weak in both species (*P. cultripes* , Pearson's $r = 0.39$, $P < 0.001$). Genetic diversity (sMLH) from SNPs in *H. molleri* was highest in southwest Iberia and decreased towards northern ($\beta = -0.08$; $P < 0.001$) and eastern localities ($\beta = -0.04$; $P = 0.02$) (Figure 3; Table S4). We did not detect a significant correlation of microsatellite diversity with latitude ($P = 0.63$) or longitude ($P = 0.10$).

For *P. cultripes* , genetic diversity decreased with latitude for SNPs ($\beta = -0.07$; $P < 0.001$) and microsatellites ($\beta = -0.09$; $P < 0.001$). Longitude had a marginal effect on diversity from SNPs ($\beta = -0.02$; $P = 0.06$) but not from microsatellites ($P = 0.93$). Both markers agreed in diversity being (1) extremely low in the north-eastern localities, in costal France, both on the Atlantic and Mediterranean sides, (2) moderately low in the Northern Plateau and along the Mediterranean coast and interior, and (3) greatest in the central south-western localities (Figure 3; Table S4). These south-western localities also showed the largest complexity in genetic structure and patterns of admixture across K (Figure 1).

Discussion

Our comparative assessment revealed that a typical microsatellite dataset (18 loci in *H. molleri* and 14 in *P. cultripes*) can yield similar range-wide patterns of genetic structure than those inferred with a few thousand SNPs (15,412 and 33,140, respectively). Differences across marker types involved mainly inference of the optimal number of clusters (K), and assessment of individual and population admixture levels.

Effect of marker type in model-based clustering and genetic diversity

We found overall concordance between markers in recovering the same major genetic clusters in STRUC-TURE analyses (Figure 1), although the model-free clustering approach based on NJ yielded poorly supported clustering for microsatellites compared to SNPs (Supplementary File S1). This shows that population genetic models implemented in STRUCTURE are efficient to infer genetic structure from a few highly poly-morphic microsatellite loci, with comparable performance to analyses using thousands of bialellic SNPs. Many population genetics studies often rely on a few microsatellites compared to the hundreds or thousands

of SNPs needed to address similar questions regarding population structure (Haasl & Payseur, 2011; Puckett, 2017). Previous research comparing both marker types claimed that SNPs offered a "better" resolution to address biological questions when compared to microsatellites, usually referring to SNPs being able to identify more differentiated genetic clusters (Rašić et al., 2014; Malenfant et al., 2015; Jeffries et al., 2016; Puckett & Eggert, 2016; Elbers et al., 2017; Hodel et al., 2017; McCartney-Melstad et al., 2018). These assertions in favor of SNPs over microsatellites could potentially be exaggerated, because they mostly derive from non-parametric (e.g. PCA or DAPC) instead of model-based methods.

There were, however, discordances between markers. The inferred optimal number of clusters was not consistent across marker types and method of estimation (Figure S4; Table S3). The clear peak of ΔK at K = 2 in the SNP dataset in *H. molleri* contrasted with the peak of ΔK at K = 4 and K = 6 in microsatellites in our results and in Sánchez-Montes et al. (2019), respectively. Also, the PI pointed to higher larger optimal values of K than those selected by Evanno's ΔK (Figure S4). The clear peaks of ΔK at K = 2 in the SNP datasets describe the top level of hierarchical population structure and must be interpreted cautiously, since K = 2 is the optimal K value most often reported across studies even when further genetic substructure is present (Janes et al., 2017). The number of samples per population can have a strong effect on the optimal K value inferred (Puechmaille, 2016). Furthermore, the history of populations is often more complex than the "top-level" clustering approach in STRUCTURE, and as K increases, violations in the assumptions of STRUCTURE may hamper the inference of the correct population structure (Lawson, van Dorp, & Falush, 2018).

Inferred ancestral groups (clusters) and their proportions of ancestry were not fully congruent between marker types. The similarity in the Q-matrices between markers varied for both species across K. This was evidenced by ancestral groups arising at different K depending on the marker type and the different characterization of ancestral groups reflected in the amount of admixture and spatial extent of the clusters (Figure 1). Also, genetic admixture was higher in microsatellites than in SNPs only at larger K values for *P. cultripes* , driven by the localities with higher genetic diversity (central and southern Iberia) (Figures 1 and S5). Greater genetic admixture detected by microsatellites, together with their greater variance in STRUCTURE solutions at large values of K (Figure 2), suggest microsatellites have reduced power to detect weaker or more complex signals of genetic structure, as those reflected at larger values of K. For SNPs, even at larger values of K (K > 6), the SSC fell into alternative discrete solutions (Figure 2). These alternative solutions to the optimal K problem deserve independent biological interpretations (Kopelman et al., 2015; Pritchard et al., 2000; Wang et al., 2007), but should be considered with caution to avoid over-interpretation (Lawson et al., 2018). Previous studies comparing STRUCTURE results between SNPs and microsatellites used datasets or approaches that were not fully comparable between the two marker types, limiting the scope of their conclusions. For instance, Bradbury et al. (2015) described different levels of admixture between markers but used different biological samples for each marker type, while Bohling et al., (2019) relied on different clustering approaches, NGSadmix (for SNP data) and STRUCTURE (for microsatellites), to conclude that microsatellites yielded less precise and less consistent results. Of the few studies that used exactly the same individuals and clustering approach across different marker types, Lemopoulos et al. (2018) found nearly identical ancestry memberships, whereas Malenfant et al. (2015) reported more admixed ancestries for microsatellites, in agreement with our results.

Genetic diversity increased with latitude in SNPs for both species but only in *P. cultripes* for microsatellites. Genetic diversity as estimated from SNPs was spatially more coherent with genetic structure, showing less variance between localities from the same cluster (e.g. the southern group of *P. cultripes*) (Figure 3). The differences in genetic diversity between marker types resulted in a weak correlation of the corresponding sMLH values (Supplementary Figure S6). This suggests that microsatellites might be offering sub-optimal representative measures of genomic diversity (Fischer et al., 2017; Väli, Einarsson, Waits, & Ellegren, 2008) compared to SNPs.

The lack of full agreement between spatial patterns of genetic diversity between marker types, even when comparable datasets and analytical methods are used could be explained by the larger number of SNPs

7

compared to microsatellite loci. In this sense, the high number of SNPs may overcome some of the limitations of using few loci as surrogates of genome-wide variation, like stochasticity related to loci selection and the associated ascertainment bias (Fischer et al., 2017; Guillot & Foll, 2009; Lemopoulos et al., 2019; Morin et al., 2004). Different marker discovery approaches (e.g. representation of functional genomic regions) could be related to some of the differences between markers (Clark, Hubisz, Bustamante, Williamson, & Nielsen, 2005; Dufresnes, Brelsford, Béziers, & Perrin, 2014; Lachance & Tishkoff, 2014). Additionally, differences in type and rate mutation could also account for the differences in patterns between markers (Ellegren, 2004; Morin et al., 2004). The better representation of loci covering a wider evolutionary scale in SNPs compared to microsatellites (Haasl & Payseur, 2011; Linck & Battey, 2019; Morin et al., 2004) could be responsible to some loss of resolution with microsatellites when representing older demographic processes.

Contribution of SNPs to the evolutionary history of H. molleri and P. cultripes

Our SNPs results on *H. molleri* are consistent with those of Sánchez-Montes et al. (2019) in recovering two major clusters, southern and northern, coinciding with the two major mitochondrial lineages and the north/south microsatellite clusters. Patterns of genetic diversity as measured with SNPs increase with latitude and decrease from coastal localities of central Portugal towards the east (Figure 3). Sánchez-Montes et al. (2019) also found greater mitochondrial and microsatellite diversity in western localities, but no clear association with latitude. Our findings support the existence of southwestern refugia for *H. molleri* in Iberia, where it would have persisted through glacial cycles in Atlantic central-south Portugal and Sierra Morena, followed by two major historical dispersal axes, towards the north and east.

For *P. cultripes* , analysis of the SNP dataset provided results congruent with microsatellites and mitochondrial DNA from Gutiérrez-Rodríguez et al. (2017), identifying three main lineages: a southern one with high genetic diversity and complex genetic structure, a second lineage in the Northern Plateau with low genetic diversity, and a third lineage in the northeast, with very low genetic diversity (Figures 1 and 3). The two latter groups were not further sub-structured, but we found signs of finer substructure in the southern lineage. Our results support northern and eastern colonization routes from southern refugia, with the Northern-Plateau lineage probably resulting from a relatively recent colonization event, contrasting with the interpretation of Gutiérrez-Rodríguez et al. (2017), who suggested the existence of a Northern-Plateau refugium. This study thus adds to the growing body of evidence showing the importance of southern refugia for a broad range of taxa in the Iberian Peninsula across glaciations in the Pleistocene (Gómez & Lunt, 2007). Inferred trends of decreasing genomic diversity towards northern latitudes provide valuable information for the management of the genetically diverse populations from southern refugia and their less diverse northern counterparts, both of which face increased risk of extinction under future climatic scenarios (Araújo, Guilhaumon, Neto, Ortego, & Calmaestra, 2011).

Data Accessibility

Raw data, genotypes, code for data analysis and the generation of figures are available in a GitHub repository (*github.com/csmiguel/usat_snp*), and a permanent release has been deposited in ZENODO (DOI:10.5281/zenodo.3953536).

Author Contributions

SBC, GVA, MC and AV conceived the study. GVA, SBC, ÍMS and AV contributed to the samples and genetic data. AM conducted lab work. All authors contributed with critical comments and data analysis. MCS led data analysis and writing of the manuscript. All authors contributed critically to writing the manuscript and gave final approval for publication.

References

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* ,*26* (2), 420–430. doi: 10.1111/mec.13948

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* , *17* (2), 81–92. doi: 10.1038/nrg.2015.28

Araújo, M. B., Guilhaumon, F., Neto, D. R., Ortego, I. P., & Calmaestra, R. (2011). *Impactos, Vulnerabilidad y Adaptación al Cambio Climático de la Biodiversidad Española. 2 Fauna de Vertebrados* . doi: 10.13140/RG.2.1.3766.3200

Avise, J. C. (2004). *Molecular Markers, Natural History and Evolution* (2nd Editio). Sunderland, Massachusetts: Sinauer Associates, Inc.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., . . . Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* ,*3* (10), e3376. doi: 10.1371/journal.pone.0003376

Bohling, J., Small, M., Von Bargen, J., Louden, A., & DeHaan, P. (2019). Comparing inferences derived from microsatellite and RADseq datasets: a case study involving threatened bull trout.*Conservation Genetics* , *20* (2), 329–342. doi: 10.1007/s10592-018-1134-z

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (2015). Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure.*Molecular Ecology* , *24* (20), 5130–5144. doi: 10.1111/mec.13395

Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H., & Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research* , *15* (11), 1496–1502. doi: 10.1101/gr.4107905

Coltman, D. W., Pilkington, J. G., Smith, J. A., & Pemberton, J. M. (1999). Parasite-mediated selection against inbred soay sheep in a free-living, island population. *Evolution* , *53* (4), 1259–1267. doi: 10.1111/j.1558-5646.1999.tb04538.x

Cuvier, G. (1829). *Le règne animal distribué d'après son organisation pour servir de base à l'histoire naturelle des animaux, et d'introduction à l'anatomie comparée* (2ème éditi). Paris: Déterville & Crochard.

Dallas, J. F. (1992). Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome* ,*3* (8), 452–456. doi: 10.1007/BF00356155

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* , *12* (7), 499–510. doi: 10.1038/nrg3012

DeFaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources* ,*13* (3), 377–392. doi: 10.1111/1755-0998.12071

Di Rienzo, A., Peterson, A. C., Garzat, J. C., Valdes, A. M., Slatkint, M., & Freimer, N. B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America* , *91* (8), 3166–3170. Retrieved from https://www.pnas.org/content/pnas/91/8/3166.full.pdf

Dufresnes, C., Brelsford, A., Béziers, P., & Perrin, N. (2014). Stronger transferability but lower variability in transcriptomic- than in anonymous microsatellites: Evidence from Hylid frogs. *Molecular Ecology Resources* , *14* (4), 716–725. doi: 10.1111/1755-0998.12215

Elbers, J. P., Clostio, R. W., & Taylor, S. S. (2017). Population genetic inferences using immune gene SNPs mirror patterns inferred by microsatellites. *Molecular Ecology Resources* , *17* (3), 481–491. doi: 10.1111/1755-0998.12591

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* , *5* (6), 435–445. doi: 10.1038/nrg1348

Estoup, A., Jarne, P., & Cornuet, J.-M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* , *11* (9), 1591–1604. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12207711

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* , *14* (8), 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* , *164* (4), 1567–1587. doi: 10.1001/jama.1987.03400040069013

Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., . . . Widmer, A. (2017). Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri* .*BMC Genomics* , *18* (1), 1–15. doi: 10.1186/s12864-016-3459-7

Garza, J. C., & Freimer, N. B. (1996). Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Research* ,*6* (3), 211–217. doi: 10.1101/gr.6.3.211

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* , *7* (4), 457–472.

Gómez, A., & Lunt, D. H. (2007). Refugia within Refugia: Patterns of Phylogeographic Concordance in the Iberian Peninsula. In S. Weiss & N. Ferrand (Eds.), *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity* (pp. 155–188). doi: 10.1007/1-4020-4904-8_5

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). DARTR : An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources* ,*18* (3), 691–699. doi: 10.1111/1755-0998.12745

Guillot, G., & Foll, M. (2009). Correcting for ascertainment bias in the inference of population structure. *Bioinformatics* ,*25* (4), 552–554. doi: 10.1093/bioinformatics/btn665

Gutierrez-Rodriguez, J., Barbosa, A. M., & Martinez-Solano, I. (2017). Present and past climatic effects on the current distribution and genetic diversity of the Iberian spadefoot toad (*Pelobates cultripes*): an integrative approach. *Journal of Biogeography* ,*44* (2), 245–258. doi: 10.1111/jbi.12791

10

Haasl, R. J., & Payseur, B. A. (2011). Multi-locus inference of population structure: A comparison between single nucleotide polymorphisms and microsatellites. *Heredity* , *106* (1), 158–171. doi: 10.1038/hdy.2010.21

Hedrick, P. W. (1999). Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. *Evolution* ,*53* (2), 313. doi: 10.2307/2640768

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

Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., . . . Soltis, P. S. (2016). The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. *Applications in Plant Sciences* , *4* (6), 1600025. doi: 10.3732/apps.1600025

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure.*Bioinformatics* , *23* (14), 1801–1806. doi: 10.1093/bioinformatics/btm233

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum.*Molecular Ecology* , *26* (14), 3594–3602. doi: 10.1111/mec.14187

Jeffries, D. L., Copp, G. H., Handley, L. L., Hakan Olsen, K., Sayer, C. D., & Hanfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius* , L. *Molecular Ecology* ,*25* (13), 2997–3018. doi: 10.1111/mec.13613

Kalinowski, S. T. (2002). How many alleles per locus should be used to estimate genetic distances? *Heredity* , *88* , 62–65. doi: 10.1038/sj/hdy/6800009

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across *K* .*Molecular Ecology Resources* , *15* (5), 1179–1191. doi: 10.1111/1755-0998.12387

Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it.*BioEssays* , *35* (9), 780–786. doi: 10.1002/bies.201300014

Lambert, M. R., Skelly, D. K., & Ezaz, T. (2016). Sex-linked markers in the North American green frog (*Rana clamitans*) developed using DArTseq provide early insight into sex chromosome evolution. *BMC Genomics* , *17* (1), 844. doi: 10.1186/s12864-016-3209-x

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications* , *9* (1), 1–11. doi: 10.1038/s41467-018-05257-7

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkila, S., Vasemagi, A., Huusko, A., Hyvarinen, P., . . . Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. *Ecology and Evolution* , *9* (4), 2106–2120. doi: 10.1002/ece3.4905

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets.*Molecular Ecology Resources* , *19* (3), 639–647. doi: 10.1111/1755-0998.12995

Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* , *6* (1), S26. doi: 10.1186/1471-2156-6-S1-S26

Luikart, G., & Cornuet, J.-M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* , *12* (1), 228–237.

Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics* , *26* (8), 345–352. doi: 10.1016/j.tig.2010.05.003

Malenfant, R. M., Coltman, D. W., & Davis, C. S. (2015). Design of a 9K illumina BeadChip for polar bears (Ursus maritimus) from RAD and transcriptome sequencing. *Molecular Ecology Resources* ,*15* (3), 587–600. doi: 10.1111/1755-0998.12327

McCartney-Melstad, E., Vu, J. K., & Shaffer, H. B. (2018). Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. *Molecular Ecology* , *27* (22), 4430–4443. doi: 10.1111/mec.14892

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers.*Genome Research* , *17* (2), 240–248. doi: 10.1101/gr.5681207

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* ,*19* (4), 208–216. doi: 10.1016/j.tree.2004.01.009

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists.*Molecular Ecology* , (June), 3193–3206. doi: 10.1111/mec.14792

Pereira, P., Teixeira, J., & Velo-Anton, G. (2018). Allele surfing shaped the genetic structure of the European pond turtle via colonization and population expansion across the Iberian Peninsula from Africa. *Journal of Biogeography* , *45* (9), 2202–2215. doi: 10.1111/jbi.13412

Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., . . . McCue, M. E. (2013). Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data.*PLoS ONE* , *8* (1). doi: 10.1371/journal.pone.0054997

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* (5), e37135. doi: 10.1371/journal.pone.0037135

Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. *Molecular Ecology Resources* , *17* (6), e268–e274. doi: 10.1111/1755-0998.12702

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). {CODA}: Convergence Diagnosis and Output Analysis for {MCMC}. *R News* . Retrieved from https://www.r-project.org/doc/Rnews/Rnews_2006-1.pdf

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* ,*155* (2), 945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Puckett, E. E. (2017). Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions.*Conservation Genetics Resources* , *9* (2), 289–304. doi: 10.1007/s12686-016-0643-7

Puckett, E. E., & Eggert, L. S. (2016). Comparison of SNP and microsatellite genotyping panels for spatial assignment of individuals to natal range: A case study using the American black bear (*Ursus americanus*). *Biological Conservation* , *193* , 86–93. doi: 10.1016/j.biocon.2015.11.020

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* , *16* (3), 608–627. doi: 10.1111/1755-0998.12512

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies.*Ecology and Evolution* , *4* (22), 4399–4428. doi: 10.1002/ece3.1305

Queney, G., Ferrand, N., Weiss, S., Mougel, F., & Monnerot, M. (2001). Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (Oryctolagus cuniculus).*Molecular Biology and Evolution* , *18* (12), 2169–2178. doi: 10.1093/oxfordjournals.molbev.a003763

Rašić, G., Filipović, I., Weeks, A. R., & Hoffmann, A. A. (2014). Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti* . *BMC Genomics* , *15* (1), 1–12. doi: 10.1186/1471-2164-15-275

Sánchez-Montes, G., Recuero, E., Barbosa, A. M., & Martinez-Solano, I. (2019). Complementing the Pleistocene biogeography of European amphibians: Testimony from a southern Atlantic species. *Journal of Biogeography* , *46* (3), 568–583. doi: 10.1111/jbi.13515

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J. I. (2016). inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution* , *7* (11), 1331–1339. doi: 10.1111/2041-210X.12588

Tonkin-Hill, G., & Lee, S. (2016). *starmie: Population Structure Model Inference and Visualisation* . Retrieved from https://cran.r-project.org/package=starmie

Valdes, A. M., Slatkin, M., & Freimer, N. B. (1993). Allele frequencies at microsatellite loci: The stepwise mutation model revisited.*Genetics* , *133* (3), 737–749.

Vali, U., Einarsson, A., Waits, L., & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* , *17* (17), 3808–3817. doi: 10.1111/j.1365-294X.2008.03876.x

Vats, D., & Knudson, C. (2018). *Revisiting the Gelman-Rubin Diagnostic* . 1–22. Retrieved from http://arxiv.org/abs/1812.09384

Wang, J. (2019). A parsimony estimator of the number of populations from a STRUCTURE-like analysis. *Molecular Ecology Resources* ,*19* (4), 970–981. doi: 10.1111/1755-0998.13000

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., . . . Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics* ,*3* (11), 2049–2067. doi: 10.1371/journal.pgen.0030185

Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics* , *2* (8), 1123–1128. doi: 10.1093/hmg/2.8.1123

Webster, M. T., Smith, N. G. C., & Ellegren, H. (2002). Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments.*Proceedings of the National Academy of Sciences of the United States of America* , *99* (13), 8748–8753. doi: 10.1073/pnas.122067599

Zhang, D. X., & Hewitt, G. M. (2003). Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects.*Molecular Ecology* . doi: 10.1046/j.1365-294X.2003.01773.x

Legend to Figures

Figure 1: Genetic structure in *Hyla molleri* (left) and*Pelobates cultripes* (right) based on STRUCTURE analyses of the SNP and microsatellite datasets. Pies represent averaged proportion of inferred ancestries of the major mode in CLUMPAK, from K = 2 to K = 8. Shaded areas represent the species distributions. To facilitate visual comparison of spatial patterns of genetic structure between markers, Q-matrices from both markers for any given K and species were aligned using CLUMPP before plotting.

Figure 2: Comparison of STRUCTURE results in the SNP and microsatellite datasets for *H. molleri* (A) and *P. cultripes* (B). The horizontal axis shows Pairwise Symmetric Similarity Coefficients between Q-matrices from STRUCTURE runs across K values (vertical axis) using averaged ancestries per locality in *H. molleri* and individual ancestries in *P. cultripes* . Comparisons involving the same marker type (microsatellite-microsatellite: blue triangles, and SNP-SNP: green circles) show higher similarity than those involving different marker types (red squares).

Figure 3: Genetic diversity measured as multilocus heterozygosity (sMLH) for *H. molleri* (A: SNPs, B: microsatellies) and *P. cultripes* (C: SNPs, D: microsatellites). Integers represent median sMLH values for

each locality.

14

A. *Hyla molleri*   B. *Pelobates cultripes*

○ SNPs–SNPs   □ SNPs–microsatellites   △ microsatellites–microsatellites

Symmetric Similarity Coefficient

A SNPs H. molleri
B microsatellites *H. molleri*
C SNPs *P. cultripes*
D microsatellites *P. cultripes*

sMLH
Elevation (m)

1. Aldeaseca de Alba
2. Arnedo
3. Astilla del Pino
4. Banyeres de Mariola
5. Beira
6. Bemposta
7. Bienservida
8. Boceguillas
9. Boticos
10. Buenache de la Sierra
11. Burguillos del Cerro
12. Cabanillas
13. Candeleda
14. Carpio de Azaba
15. Cernégula
16. Codesal - Ferreras de Arriba
17. Cuchía
18. Don Benito
19. Doñana
20. El Pedroso
21. Embalse de Cecebre
22. Faro
23. Fuendetodos
24. Gárgoles
25. Ginzo de Limia
26. Izagre
27. La Parroquia
28. Lavariz
29. Lleida
30. Loja
31. Longueira
32. Los Escoriales
33. Los Palancares
34. Malagón
35. Malpartida de Cáceres
36. Menasalbas
37. Monforte de Lemos
38. Montblanc
39. Montuenga de Soria
40. Navalmoral de la Mata
41. Navas de Estena -
    Baños Robledillo
42. Noviercas
43. O Grove
44. Ojos de Villaverde
45. Olvega
46. Paramos
47. Plage Biscarrosse
48. Puerto de la Cubilla
49. Puerto de Urbasa
50. Redondo
51. Riudarenes
52. Ruesga
53. Santarém
54. Sedas
55. Serra da Estrela
56. Sierra de Saceruela
57. Sinarcas
58. Tivissa
59. Tocha
60. Torrefresneda
61. Valdelinjas
62. Valencia de Alcántara
63. Valgañón
64. Villamayor de los
    Montes
65. Villanueva del Duque
66. Villaviciosa