Received: 8 December 2021 | Revised: 26 May 2022 | Accepted: 30 May 2022

DOI: 10.1111/eva.13431

ORIGINAL ARTICLE

Evolutionary Applications Open Access WILEY

From STRs to SNPs via ddRAD-seq: Geographic assignment of confiscated tortoises at reduced costs

Roberto Biello[1,2] | Mauro Zampiglia[3,4] | Silvia Fuselli[1] | Giulia Fabbri[1,5] | Roberta Bisconti[3] | Andrea Chiocchio[3] | Stefano Mazzotti[6] | Emiliano Trucchi[1,7] | Daniele Canestrelli[3] | Giorgio Bertorelle[1]

[1]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy

[2]Department of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, UK

[3]Department of Ecological and Biological Science, Tuscia University, Viterbo, Italy

[4]Central Laboratory for the National DNA Database, Prison Administration Department, Ministry of Justice, Rome, Italy

[5]Department of Veterinary Medicine, University of Sassari, Sassari, Italy

[6]Natural History Museum, Ferrara, Italy

[7]Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy

Correspondence

Roberto Biello and Giorgio Bertorelle, Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy.
Email: r.s.biello@gmail.com (R.B.) and ggb@unife.it (G.B.)

Funding information

the Carabinieri Forestali; Tuscia University (Viterbo); University of Ferrara

Abstract

Assigning individuals to their source populations is crucial for conservation research, especially for endangered species threatened by illegal trade and translocations. Genetic assignment can be achieved with different types of molecular markers, but technical advantages and cost saving are recently promoting the shift from short tandem repeats (STRs) to single nucleotide polymorphisms (SNPs). Here, we designed, developed, and tested a small panel of SNPs for cost-effective geographic assignment of individuals with unknown origin of the endangered Mediterranean tortoise *Testudo hermanni*. We started by performing a ddRAD-seq experiment on 70 wild individuals of *T. hermanni* from 38 locations. Results obtained using 3182 SNPs are comparable to those previously obtained using STR markers in terms of genetic structure and power to identify the macro-area of origin. However, our SNPs revealed further insights into the substructure in Western populations, especially in Southern Italy. A small panel of highly informative SNPs was then selected and tested by genotyping 190 individuals using the KASP genotyping chemistry. All the samples from wild populations of known geographic origin were genetically re-assigned with high accuracy to the original population. This reduced SNPs panel represents an efficient molecular tool that enables individuals to be genotyped at low cost (less than €15 per sample) for geographical assignment and identification of hybrids. This information is crucial for the management in-situ of confiscated animals and their possible re-allocation in the wild. Our methodological pipeline can easily be extended to other species.

KEYWORDS

assignment, informative SNPs, Mediterranean tortoises, RAD sequencing, reduced SNP panel, *Testudo hermanni*

1 | INTRODUCTION

In the last twenty years, next-generation sequencing (NGS) favored a shift from short tandem repeats (STRs) markers to single nucleotide polymorphisms (SNPs) in molecular studies of many organisms (Barbosa et al., 2020; Rajora, 2019; Seeb et al., 2011). This transition improved our understanding in several areas relevant to conservation, such as the study of inbreeding (Grossen et al., 2020; Robinson et al., 2019), hybridization (Sinding et al., 2018; VonHoldt et al., 2016), population structure and admixture (Jeffries et al., 2016; Zimmerman et al., 2020), and population, parentage and kinship assignment (Kleinman-Ruiz et al., 2017; Roques et al., 2019).

Some of the most common and widespread techniques used for genotyping SNPs in nonmodel organisms are based on reduced libraries as the restriction site-associated DNA sequencing (RAD-seq; Davey et al., 2011; Hohenlohe et al., 2011; Miller et al., 2007), its variant double-digest RAD-seq (ddRAD-seq; Peterson et al., 2012) and other derived methods (Campbell et al., 2018). Currently, sequencing of nuclear genomic regions with reduced library approaches offers the possibility to obtain, at low price, several thousands of SNPs. Moreover, selecting a posteriori only the informative SNPs that maintain as much as possible the original dataset information (e.g., SNP-based panel) required for the purpose of interest can considerably reduce the subsequent genotyping costs and technical efforts necessary to analyze a large number of individuals. This aspect is crucial in conservation genetics where practitioners need accuracy but also cheap and easy-to-implement guidance (Holderegger et al., 2019). SNP-based panels, compared to STRs, address this need by simplifying the allele scoring procedure, allowing data transferability across studies, with the potential for high-throughput screening (e.g., Garvin et al., 2010). SNP arrays are also flexible, with low-density arrays allowing tens to thousands of SNPs to be genotyped and high-density arrays or "SNP chips" supporting tens of thousands to millions of SNPs (Humble et al., 2020; von Thaden et al., 2020).

Until recently, reduced SNP assays were used mostly for domestic species (Johnston et al., 2017; Ogden et al., 2012; Pertoldi et al., 2009). Nowadays, custom species-specific SNP panels are being developed for several wild species providing insights into diverse topics including conservation genomics (Eriksson et al., 2020; Feutry et al., 2020; Henriques et al., 2018; Meek et al., 2016). Small panels of SNPs (<100) have already been used for the identification of distinct genetic populations to delineate biologically accurate management units and for the identification of individuals, including the assignment of individuals to their source population (Förster et al., 2018; Henriques et al., 2018; Kleinman-Ruiz et al., 2017).

This approach can be described in four steps. First, a large number of SNPs is isolated and typed in a small number of individuals; second, the suitability of these markers to answer the question of interest is tested; third, a much smaller panel of SNPs is identified, given it can answer the same question of interest but at lower costs; and fourth, the subset is tested using a larger sample of individuals.

Here, we applied these steps to isolate informative SNPs for the assignment of samples to their population or area of origin. Inferring the geographic origin of living organisms from their genotypes is of great interest in wildlife management, conservation, and forensic applications. It can provide information about gene flow, migration patterns and connectivity in natural populations (Kremer et al., 2012) but can also help inform wildlife managers about illegal animal translocations and poaching hotspots (Biello et al., 2021; Ogden & Linacre, 2015). Our study produced a tool specific for the Herman's tortoise, but it can be considered also as a useful example to facilitate the production of reduced SNP panels in other species.

The Hermann's tortoise, *Testudo hermanni* Gmelin (1789), is one of the most endangered reptiles in Europe. The intense harvesting for pet trade, especially before the 1980s when it was not banned yet (Ljubisavljević et al., 2011), the releases of non-native individuals, and the habitat reduction and degradation (Stubbs et al., 1985) are the major threats for this species (Bertolero et al., 2011). *T. hermanni* is distributed in disjoint populations across Mediterranean Europe, from Spain to the Balkans, including various Mediterranean islands. Based on mtDNA and STR markers, two subspecies are recognized (Biello et al., 2021; Perez et al., 2014): the eastern *T. h. boettgeri* and the western *T. h. hermanni*. The species is included in the list of strictly protected fauna species by the Bern Convention on the Conservation of European Wildlife and Natural Habitat. The western subspecies, *T. h. hermanni*, is classified as "Endangered" by the IUCN Red List (1996). Previous genetic studies suggest that individuals can be correctly assigned in most cases to their macro-area of origin with a small panel of STR markers (Biello et al., 2021; Perez et al., 2014). However, given the large number of individuals of unknown origin hosted in rescue centers and potentially suitable for reintroductions plans, the possibility to use modern NGS technologies to decrease the genotyping costs should be carefully considered.

The main goal of this study is to introduce and validate a small panel of SNPs extracted from genome-wide distributed markers for the cost-effective geographic assignment of large numbers of confiscated *T. hermanni* individuals, necessary for their management and their possible re-allocation in the wild. We accomplished this goal following the four steps defined above. In particular, 1. we performed a ddRAD-seq experiment on 84 wild individuals of which 70 provided informative genotypes for further analyses; 2. we verified that the results provided by the ddRAD-seq loci, in terms of genetic structure and power to identify the macro-area of origin, are adequate (in our case, considering the results obtained with a sample of 292 wild individuals already typed at 7 STRs); 3. we designed a cost-effective reduced panel of SNPs with retained informativeness; and 4. we tested the reduced SNP panel in 190 individuals using the KASP-by-Design Fluidigm Assays (LGC Genomics; He et al., 2014). Our study has practical consequences for the conservation, management, and reintroduction of *T. hermanni*, and it is also demonstrative of a technical pipeline applicable to other species.

2 | MATERIALS & METHODS

2.1 | Samples, DNA extraction, and ddRAD-seq library preparation

Eighty-four samples already available in our laboratories were selected to perform the ddRAD-seq genotyping experiment. These samples belonged to individuals collected across most of the contemporary geographical range of the species and covering all the previously known genetic clusters. Of these, 70 were retained for further analyses after quality control and filtering (see next paragraph and Figures S1–S12).

DNeasy Blood and Tissue Kit (Qiagen) and Quick-DNA Universal Kit (Zymo Research) were used to extract genomic DNA from FTA cards and whole blood stored at −20°C, respectively, following manufacturer instructions. DNA was quantified with Qubit using the dsDNA BR kit (Invitrogen). ddRAD-seq libraries were prepared using *Sbf*I and *Nco*I for restriction digestion. Sequencing was performed on the Ion Torrent PGM with the Ion PGM Hi-Q View Sequencing kit (Life Technologies) following manufacturer's instructions. See Supplementary Methods for more details about library preparation and sequencing.

2.2 | Quality control and filtering

Raw reads generated were demultiplexed and split into individual data files using the *process_radtags* program in version 1.44 of STACKS (Catchen et al., 2011). We used TRIMMOMATIC (Bolger et al., 2014) to trim our raw reads, removing read-through adapters, low quality (light filter—modify on a case-by-case basis), and shorter than 200 bp, cropping the first two bp (GG left from restriction site). The resulting reads were 200 bp in length to ensure downstream compatibility in STACKS, which requires uniform read length when building de novo loci.

After trimming and quality-filtering, the STACKS programs *ustacks*, *cstacks*, and *sstacks* were used to build de novo catalog loci. We specified a minimum of five reads per locus to build primary catalog stacks and allowed a maximum distance of six nucleotide mismatches within these stacks. Furthermore, we allowed eight mismatches between stacks during construction of the catalog because the specimens in our dataset were represented by two subspecies of the same species.

Following de novo mapping, an initial data-filtering step was performed using the *population* component of STACKS retaining only those loci present in at least 75% of individuals at each site and with a maximum of 5 SNPs per locus (*full dataset*, $n = 3182$). In an additional filtering step, we retained only one SNP per locus (*single SNP dataset*, $n = 1179$) to minimize redundancy of genetic information due to the physical linkage between the markers. This reduced dataset was used for PCA and F_{ST}-based methods. The resulting vcf files were converted to other program-specific input files using PGDSPIDER v2.0.5.1 (Lischer & Excoffier, 2012).

2.3 | Genetic structure with ddRAD-seq in comparison with STRs

In order to compare the results provided by the ddRAD-seq loci in terms of genetic structure and power to identify the macro-area of origin of the individuals, we selected a subset of 292 wild samples genotyped previously at 7 STR loci (Biello et al., 2021; Perez et al., 2014) from the same locations of the samples included in the ddRAD-seq experiment.

The population structure was inferred from the ddRAD-seq with the package *fineRADstructure* designed to identify co-ancestry from RAD-seq data (Malinsky et al., 2018). Briefly, the algorithm compares each RAD-seq locus of each individual (recipient) with the alleles in all other individuals (potential donors) estimating the number of sequence differences (i.e., SNPs) to infer its nearest neighbor allele (donor) that is the allele (i.e., the concatenation of all the SNPs at each locus) with the least number of differences. The local co-ancestry values are then summed across all loci to obtain the co-ancestry similarity matrix for the full data. Next, an MCMC clustering algorithm infers the most likely population configuration. In summary, using information from the entire genetic variation identified with RAD-seq, the analysis estimates the number of genetic clusters, quantifies ancestry sources in each group in terms of co-ancestry value, and provides a tree-like illustration of the relationships between clusters. To perform this analysis, the output of the software *population* in STACKS was converted into a *RADpainter* input file with the python script Stacks2fineRAD.py. Samples were assigned to populations using 100,000 iterations as burn-in before sampling 100,000 iterations. A tree was built using 10,000 iterations, and the output visualized using the fineradstructureplot.r and finestructurelibrary.r R scripts (available at https://github.com/millanek/fineRADstructure). Populations were defined as clusters within the *fineRADstructure* tree and relatedness plot.

The result of the cluster analysis based on the ddRAD-seq dataset was compared with similar analyses suitable for the panel of STRs described above. Specifically, we performed a cluster analysis on the STRs dataset using the Bayesian method implemented in STRUCTURE v2.3.4 (Pritchard et al., 2000). The analysis was conducted choosing a model with admixture, uncorrelated allele frequencies, and a nonuniform ancestry prior *alpha* among clusters, as suggested by (Wang, 2017) for uneven samplings. We ran 20 replicates for each value of *K* (i.e., the number of source populations) from $K = 1$ to $K = 10$ with 750,000 MCMC after a burn-in of 500,000. Structure results were summarized and visualized with the web server CLUMPAK (Kopelman et al., 2015). We used STRUCTURE HARVESTER (Earl & vonHoldt, 2012) to infer the best value of K, based on both deltaK (Evanno et al., 2005) and the Pr[X|K] (Pritchard et al., 2000).

We further tested for population structure both the ddRAD-seq and STR datasets using a Principal Component Analysis (PCA) with the *dudi.pca* function in the R package *adegenet* v2.1.3 (Jombart, 2008). The samples were visualized following the clusters inferred by *fineRADstructure*.

Pairwise population F_{ST} values were estimated among clusters inferred from *fineRADstructure* analysis using ARLEQUIN v3.5 (Excoffier & Lischer, 2010) for ddRAD-seq and STR datasets. Furthermore, in order to investigate in detail the genetic pattern found in Calabria (Southern Italy), where additional genetic clusters are identified only with ddRAD-seq data (see Section 3), we tested the effects of geographic distance on genetic structure on those populations. We calculated the pairwise genetic distances among individuals using either the number of allelic differences between individuals or the proportion of shared alleles (D_{PS}; Bowcock et al., 1994) using the R packages *poppr* (Kamvar et al., 2014) and *adegenet* v2.1.3 (Jombart, 2008), respectively. Euclidean distances between geographic locations were calculated using the R package *fossil* (Vavrek, 2011). The Mantel test was conducted using the R package *vegan* (Dixon & Palmer, 2003).

2.4 | Design of a small panel of informative SNPs

The ability of the panels to reflect the geographic assignment based on the complete set of SNPs was analyzed following a "node approach" based on the dendrogram produced by *fineRADstructure* with the whole dataset (see Section 3). This dendrogram represents the maximum level of geographic resolution supported by the entire set of SNPs ($n = 3182$). To maintain this level of resolution while reducing the number of markers, we considered one node at the time, and ranked each SNP with the aim of identifying SNPs that were most informative for maximizing genetic differentiation between the two groups separated by the node. A comparable number of SNPs was selected from each node (see Section 3). We applied three different methods to select informative SNPs useful for geographic assignment (see Helyar et al., 2011 for review): loadings from PCA, F_{ST}, and random forest (RF; Breiman, 2001). Our aim was to identify the best panel with 48 or 96 SNPs, meaning that a total of six panels were compared (two panel sizes for each method). Only SNPs having at least 50-bp flanking region in the ddRAD-seq loci were selected, as required by the KASP screening protocol (see below).

2.4.1 | PCA-based panel

We performed seven principal component analyses, one per node, using the R package *adegenet* v2.1.3 (Jombart, 2008). For each PCA, a graph was created to visualize the distribution of the data and evaluate which axes would better separate the distinct groups; then, for each informative axis, we saved the SNPs within the top 5% loading value.

2.4.2 | F_{ST}-based panel

The R package *genepop* v1.0.5 (Rousset, 2008) was used to compute pairwise F_{ST} between the two groups branching from each node, retaining the SNPs with the highest F_{ST} values.

2.4.3 | Random forest panel

We finally applied RF (R package *randomforest* v4.6-12; Liaw & Wiener, 2002), a machine learning method used for classification and regression. This method allows to identify a limited number of features (e.g., SNP) that best classify the observations into prior groups by computing the Mean Decrease in Accuracy (MDA). MDA is a measure of worsening of the model in classifying the observations when removing one feature at the time. Higher MDA means that the feature is important to the model. At every run, a forest of decision trees is grown. The out-of-bag (OOB) error is how RF measures misclassification; it is the mean of the prediction error of a bootstrapped training subsample across all the trees. Following the node approach as for the previous methods, we determined for every node our *ntree* parameter (number of trees) by running RF with 100, 500, 1000, 8000, 25,000 trees. The *mtry* parameter (number of features considered at each node) was tested using the function *tuneRF*. We then ran RF 10 times per node, as suggested in Sylvester et al. (2018), computed and exported MDA from every run. We only selected SNPs that were listed in all the runs per node with MDA > 1. We finally created two panels retaining SNPs with the highest MDA.

Venn diagrams (http://bioinformatics.psb.ugent.be/webtools/Venn/) were built to visualize overlapping SNPs selected across all methods within each of the two panel sizes. We calculated expected heterozygosity across seven categories of SNPs selection generated by Venn diagram using ARLEQUIN v3.5 (Excoffier & Lischer, 2010).

We performed self-assignment tests to compare the three methods (each with two panel sizes, 48 and 96 SNPs) in terms of assignment accuracy. Assignment testing was performed by using the R package *AssignPop* (Chen et al., 2018). Initially, the dataset is divided into training (baseline), and test (holdout) datasets using a resampling cross-validation approach by the function *assign.MC*. The user can specify the proportion of individuals from each source "population" to be used in the training dataset. High grading bias is avoided using this method by producing randomly selected, independent training and test datasets (Anderson, 2010). Furthermore, using a PCA, the dimensionality of the training datasets (i.e., the genotypes) is decreased, and the result is utilized to create prediction models using user-chosen classification machine learning algorithms (Chen et al., 2018). Lastly, the models are used to estimate membership probabilities of tested individuals and assign them to a source population. Additionally, the training data are evaluated, and assignment tests are performed to assess the origin of individuals (Chen et al., 2018). Resampling was repeated 500 times for each combination of training individuals and loci. The proportion of individuals from each source population randomly allocated to the baseline dataset was set to 0.5 and 0.7. Finally, the linear discriminant analysis (LDA) was used as a classifier model.

2.5 | Test and validation of the SNP panel

After evaluating the performances and the cost-effectiveness of the 96 and 48 SNP panels, we selected the smaller set of markers

(48 SNPs) identified by the F_{ST} approach and these loci were tested with the KASP-by-Design Fluidigm Assays (LGC Genomics; He et al., 2014). Using this panel, we typed a total of 190 DNA samples including samples previously typed with other markers (only 7 STRs, n = 68; STRs and ddRAD-seq, n = 7; 75 in total) as internal control, samples from captivity (from rescue centers), or from wild populations never typed before (see Table S1–S12 for a detailed description of the samples). *ddRAD-seq reference* (3182 SNPs; 70 samples): We performed an assignment analysis with STRUCTURE (POPFLAG for individuals in the reference database; "update allele frequencies using only individuals with POPFLAG=1" option under a USEPOPINFO without admixture model). K was fixed at its optimal value (see Section 3), while the run length and other parameters were set as above (see STRs analysis with STRUCTURE). We assigned individuals to a source population when the probability of an individual belonging to that population was above 80% (Biello et al., 2021; Lang et al., 2021). Exclusion tests were performed with the partially Bayesian exclusion method (Rannala & Mountain, 1997) implemented in GENECLASS2 (Piry et al., 2004). We compared observed genotypes with an expected likelihood distribution of genotypes generated for each reference population by simulating 1000,000 individuals with Monte Carlo resampling. *STR reference* (7 loci; 292 samples from Biello et al., 2021): We performed assignment analysis on a subset of samples for which STR data were already available using the same methods and parameters used for the *ddRAD-seq reference*.

The diagram in Figures S1–S12 summarizes the workflow of our study.

3 | RESULTS

3.1 | Filtering and quality control

From the 84 individuals sequenced, 14 samples were excluded due to low coverage and high amount of missing data. For the remaining 70 individuals, the sequencing coverage obtained per individual ranged from 89,897 to 412,840 with an average of 236,722 reads per sample (see Table S2). After filtering loci for missing data, by retaining markers that were genotyped in at least 75% of the individuals, we recovered 1179 loci and 3182 SNPs (*full dataset*).

3.2 | Genetic structure with ddRAD-seq and comparison with STRs

The co-ancestry matrix obtained with *fineRADstructure* using the *full dataset* showed the presence of eight main clusters (Figure 1a). The lowest amount of co-ancestry was found between the two subspecies, *T. h. boettgeri* and *T. h. hermanni*. Within the *T. h. hermanni* group, six distinct clusters were identified: Italian Peninsula (ITP), North Calabria (NCA), Central Calabria (CCA), South Calabria (SCA), Sicily (SIC), and Sardinia (SAR). These results support a finer genetic structure than previously suggested using STRs (Biello et al., 2021; Perez et al., 2014). More specifically, SNPs allowed the identification of three clusters within the Calabria region (NCA, CCA, SCA) and a distinction between Sicily and Sardinia (SIC, SAR; Figure 1b). The average co-ancestry between the six clusters of *T. h. hermanni* was generally low, with the higher co-ancestry levels found between ITP and NCA, and between SIC and SAR. Within the *T. h. boettgeri* group, we detected two clusters, namely Greece (GRE), and Bosco Mesola + Croatia (MEC). The comparison between individuals within these two genetic clusters showed the highest level of co-ancestry compared to other groups. See Table S3 for a description of the nomenclature across the reference literature of each genetic cluster described in this study (Biello et al., 2021; Perez et al., 2014).

While *fineRADStructure* utilizes haplotype linkage information, other approaches such as PCA may be biased by the association between markers. Thus, population clustering was further investigated by means of PCA using only independent SNPs (*single SNP dataset*; n = 1179 markers). The first three principal components, PC1, PC2, and PC3, together accounted for 52% of the total variance in the dataset, with eigenvalues of 39.4%, 8.4%, and 4.2%, respectively (Figure 2a,b). PC1 separated the two subspecies, whereas PC2 separated the western subspecies *T. h. hermanni* in two groups, the first composed by three of the four mainland Italian populations (ITP, NCA, and CCA) and the second by south Calabria together with the two islands (SCA, SIC, and SAR; Figure 2a). Within the first group, ITP, NCA, and CCA were distributed along a "north-to-south" gradient, as were the corresponding regions of origin of the populations; conversely within the second group no clear spatial structure could be identified. The third axis separated the two *T. h. boettgeri* clusters, GRE and MEC (Figure 2b).

Pairwise estimates of F_{ST} among western subspecies populations ranged from 0.05 (between SIC and SAR) to 0.35 (between the ITP and SCA; Figure S2, lower diagonal). Within the two subspecies, the highest F_{ST} value was observed between the two eastern subspecies populations MEC and GRE (F_{ST} = 0.47). When populations belonging to different subspecies are compared, F_{ST} values reached values between 0.7 and 0.8. All F_{ST} values were significant at the nominal significance level of 0.05, and most p-values were lower than 0.01 (see Table S4). When we analyzed in detail the populations from Calabria (NCA, CCA and SCA), where additional clusters were identified, we found that geographic distance contributes significantly to genetic differentiation between populations in this region (number of allelic differences, Mantel test: r = 0.7413, p = 1e-04; D_{PS}, Mantel test: r = 0.8404, p = 1e-04; Figure S5).

The SNPs markers isolated with the ddRAD-seq approach appeared thus able to genetically discriminate populations with different geographic distribution, with a precision that was not the same using different statistical approaches. To further confirm that the SNPs markers contained relevant information to discriminate individuals with different geographic origin, at least at the macro-areas level, we compared the geographic structure inferred with the SNPs

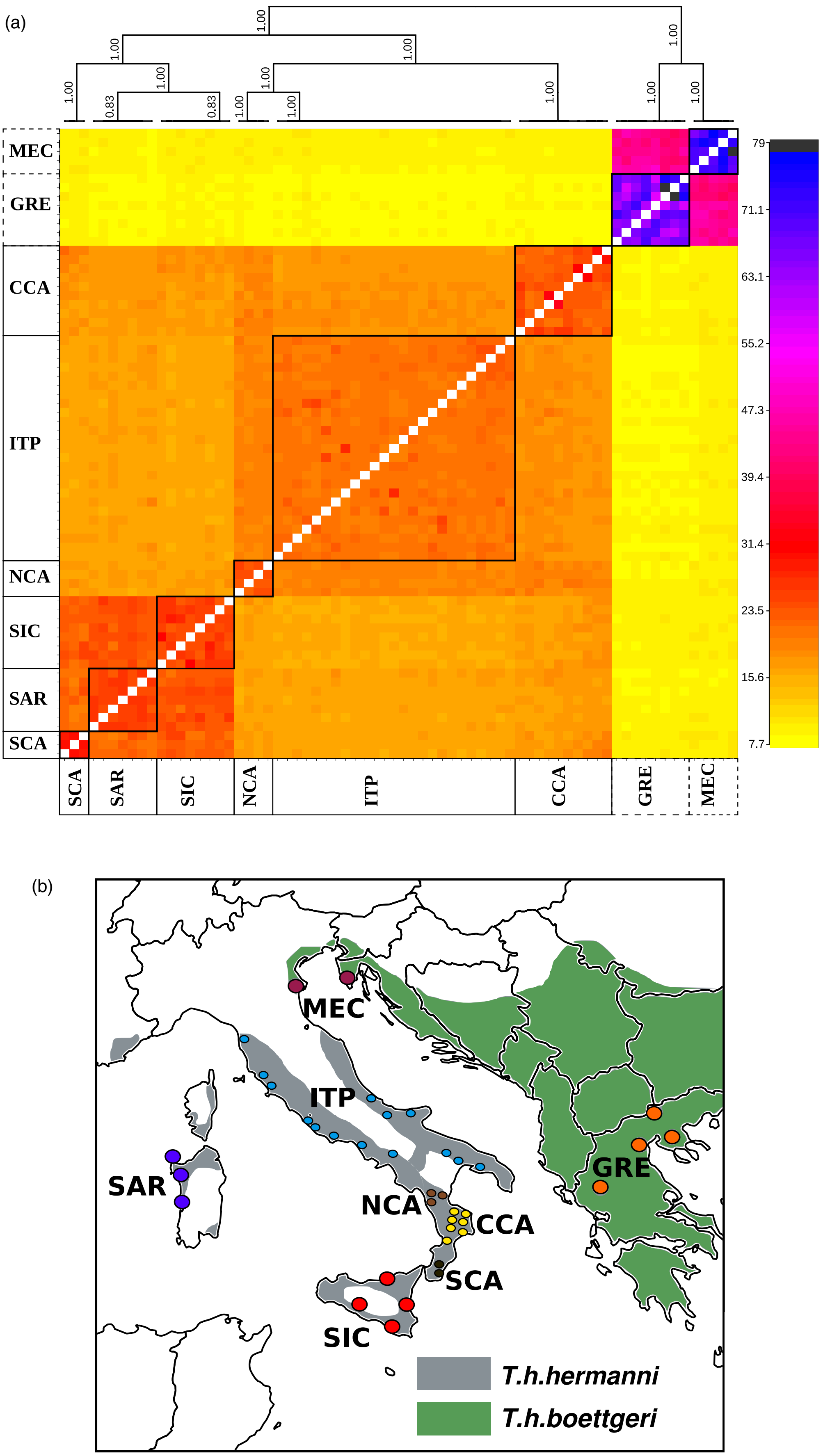
(a)
1.00
1.00
1.00
0.83
0.83
1.00
1.00
1.00
1.00
1.00
1.00
1.00
1.00
MEC
GRE
CCA
ITP
NCA
SIC
SAR
SCA
SCA
SAR
SIC
NCA
ITP
CCA
GRE
MEC
79
71.1
63.1
55.2
47.3
39.4
31.4
23.5
15.6
7.7
(b)
MEC
ITP
SAR
NCA
CCA
SCA
SIC
GRE
T.h.hermanni
T.h.boettgeri

FIGURE 1 (a) Clustered *fineRADstructure* co-ancestry matrix. The analysis included individuals from 8 areas (GRE = Greece; MEC = Croatia and Bosco Mesola; ITP = Italian Peninsula; NCA = Northern Calabria; CCA = Central Calabria; SCA = Southern Calabria; SIC = Sicily; SAR = Sardinia) and two subspecies (*Testudo hermanni hermanni*: solid line, *Testudo hermanni boettgeri*: dashed line). The highest levels of co-ancestry are indicated in black/purple and the lowest in yellow. (b) Geographic distribution of the sampled populations

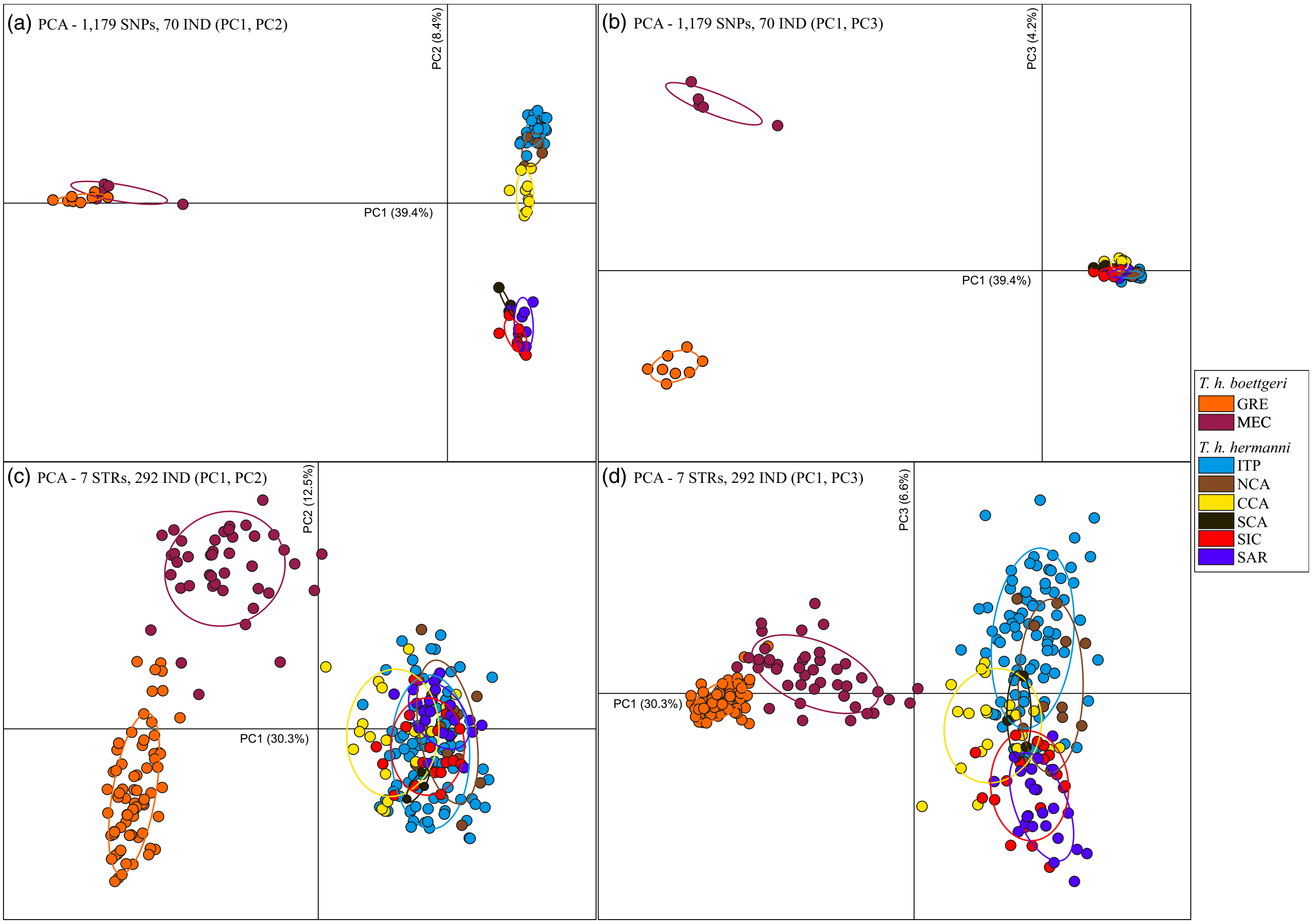

FIGURE 2 Principal component analysis (PCA) of (a, b) 1179 SNPs (from ddRAD-seq) with 70 samples, and (c, d) 7 STRs with 292 samples. Colored by eight groups found with *fineRADstructure* (see Figure 1). GRE = Greece; MEC = Croatia and Bosco Mesola; ITP = Italian Peninsula; NCA = Northern Calabria; CCA = Central Calabria; SCA = Southern Calabria; SIC = Sicily; SAR = Sardinia

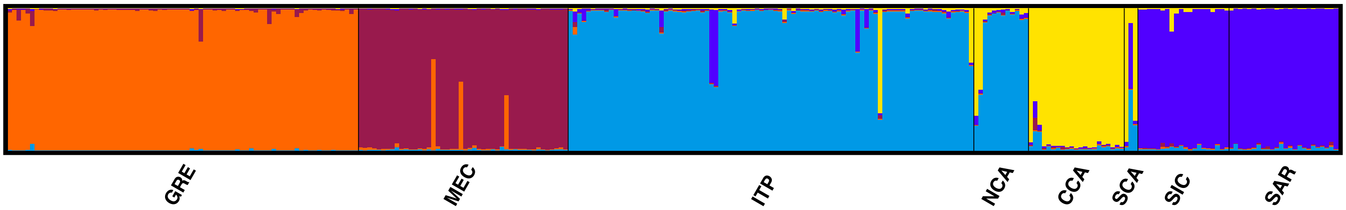

FIGURE 3 Genetic structure at seven STR loci of wild *Testudo hermanni* populations estimated using STRUCTURE, 292 samples and K = 5 with 8 population groups. GRE = Greece; MEC = Mesola and Croatia; ITP = Italian Peninsula; NCA = Northern Calabria; CCA = Central Calabria; SCA = Southern Calabria; SIC = Sicily; SAR = Sardinia

with that estimated from a small STR panel of proven efficacy for conservation and forensic applications (Biello et al., 2021). The goal here was not to analyze in detail the difference between the different markers, but to additionally validate the SNPs dataset (based on 70 individual and >1000 independent biallelic loci) by showing that it contains the same geographic partition of the genetic variation inferred from a different type of molecular marker (almost 300 individuals typed at 7 STRs). We can confidently conclude that the geographic structure identified by the SNPs is very similar, and slightly more resolved, than that based on a much larger sample size because: (i) the pairwise F_{ST} values between the groups inferred by the SNPs and the STRs (Figure S2, upper diagonal) were highly correlated (see Figure S3); (ii) the PCA results on the STR dataset confirmed a clear subdivision of the two subspecies, as well as the two *T. h. boettgeri* clusters GRE and MEC (Figure 2c,d); overall, the first three components separated some of the *T. h. hermanni* populations, but with higher overlap between individuals, as observed in the SNPs dataset (Figure 2a,b); (iii) the major genetic clusters identified by the STRUCTURE analysis on the STR markers (see Figure 3 and Figure S4) corresponded to those observed with the SNPs dataset, but without separating the three pairs of areas with the smallest F_{ST} values (ITP and NCA; CCA and SCA; SIC and SAR).

3.3 | Isolation of a small panel of informative SNPs

The SNPs were selected following a "node approach" based on the dendrogram produced by *fineRADstructure* with the whole dataset (Figure 4), with the panels including a comparable number of SNPs from each node.

In the first method (PCA-based panel), the highest loadings from the top 5% of the most informative principal components (PC) were selected. For all the seven nodes, PC1 was always the only axis to highlight differences in our data. In the second method (F_{ST}-based panel), the loci with the highest F_{ST} resulting from pairwise comparisons between the two clades at each node were selected. F_{ST} values ranged from 1 to 0.39 in the 48 SNPs panel and from 1 to 0.33 for the 96 SNPs panel. Finally, in the last method (RF) the loci with the highest MDA were included. MDA values ranged from 7.99 to 1.45 in the 48 SNPs panel and from 7.99 to 1.26 in the 96 SNPs panel. The distribution of SNPs across the three SNP selection methods is summarized by means of a Venn diagram in Figure S6. The six different SNP panels (48 or 96 SNPs, three selection criteria) were tested for accuracy in providing the assignment of individuals to their actual group. Assignment scores (Table 1) were high and ranged from 73.3% to 94%.

When the 48 SNP panels were tested, individuals were assigned to their original cluster at a mean accuracy of 75.5%, 83.2%, and 73.3%, for PCA, F_{ST}, and RF methods, respectively (Table 1, Figure S7). Assigning individuals to the two subspecies was extremely accurate, ranging from 99.1% to 100% depending on the method used (see Figure S8). Accuracy was very high also when we attempted to assign individuals to the two main groups in *T. h. hermanni*, Italian peninsula (including ITP, NCA, and CCA clusters) and the Mediterranean Islands (including SCA, SIC, and SAR). The mean accuracy ranged from 91.6% (PCA) to 97.6% (RF; see Figure S9).

The 96 SNPs panels assignments showed a mean accuracy of 94%, 88.5% and 88.2%, for PCA, F_{ST}, and RF methods, respectively (Table 1, Figure S10). Assigning individuals to the 2 subspecies was 100% accurate independently from the method used (see Figure S11). The mean accuracy when we assigned individuals to the two main groups in *T. h. hermanni* was extremely accurate, ranging from 98.5% (F_{ST}) to 99.1% (RF; see Figure S12).

3.4 | Test and validation of the SNP panel

Considering the relatively low power increase obtained by doubling the size of the SNPs panel, but the relevant additional costs of the larger compared to the smaller panel (see Section 4), we decided to empirically test the smaller panel that has a greater chance to be incorporated in future large scale genetic testing in this species.

Of the 48 SNPs initially selected with the F_{ST} approach, seven were dropped due to technical problems. Thus, 41 SNPs were left for further assignment analyses.

The genotypes of 189 DNA samples (one sample failed) were assigned to the most probable geographic area of provenance included in the *ddRAD-seq reference* (see Section 2) using STRUCTURE. We

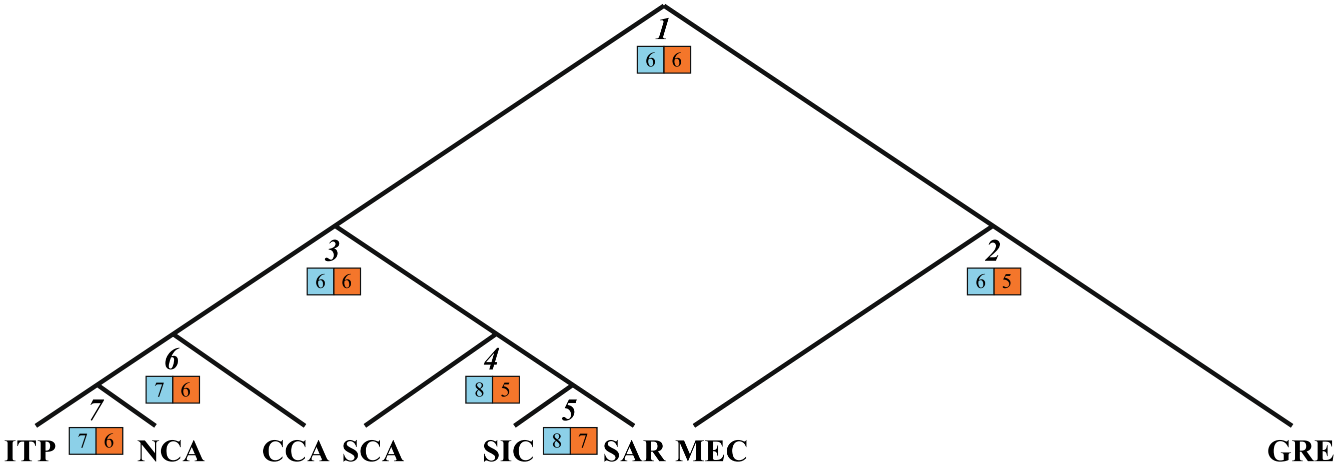

FIGURE 4 Genetic relationships among the clusters identified with *fineRADstructure* based on co-ancestry. The numbers in the squares indicate for each node the SNP included in the panel (left, light blue) and those successfully typed (right, orange). GRE = Greece; MEC = Croatia and Bosco Mesola; ITP = Italian Peninsula; NCA = Northern Calabria; CCA = Central Calabria; SCA = Southern Calabria; SIC = Sicily; SAR = Sardinia

TABLE 1 Mean percent assignment scores using *AssignPop* with 0.7 as proportion of training individuals and 500 repetitions of Monte Carlo cross validation

	Method	Assignment score (%mean ± %SD)
48 SNPs	PCA	75.5 ± 7.1
	F_{ST}	83.2 ± 7.9
	RF	73.3 ± 6.5
96 SNPs	PCA	94.0 ± 4.6
	F_{ST}	88.5 ± 6.7
	RF	88.2 ± 6.2

Note: 48 or 96 SNPs selected with PCA, F_{ST} or RF.

used K = 8 and membership coefficients (*Q*-value) > 0.80 as the assignment threshold. Using these parameter values, we were able to assign more than 78% of samples to one of the eight clusters. 32.8% of the tortoises were assigned to the GRE cluster, 24.3% to ITP, 18% to MEC, 2.1% to the SAR, 0.5% to NCA, and 0.5% to CCA, while 21.7% of the individuals were not assigned to any predefined cluster (NA).

Furthermore, we compared the geographical assignment of samples that were genotyped at 41 SNPs and 7 STR loci (see Table S1), using first the *ddRAD-seq reference* and then the *STR reference* (see Biello et al., 2021; Figure 5). In order to minimize the differences between the two references, we decided, in this analysis, to group together the clusters SIC and SAR inferred by SNPs following the STRs cluster ISS (see Biello et al., 2021; Table S3). Assigning individuals to their area of origin was extremely accurate when we considered samples from wild populations with known geographic origin and for which the reference (*ddRAD-seq reference*) was available. Both markers assigned all the samples to the same cluster (Figure 5). By contrast, for the captive populations, the results were not completely concordant between the two set of markers. The *ddRAD-seq reference* and the *STR reference* assigned 83% of samples to the same clusters (Figure 5). Among the unassigned samples (17%), four samples assigned to GRE with the SNPs were assigned to MEC with the STR data (clusters from the same subspecies, *T. h. boettgeri*), and three samples assigned to ITP with the STRs were assigned to ITP and CAL with the SNPs (clusters from the same subspecies and adjacent geographic regions, *T. h. hermanni*).

Finally, we performed assignment analysis for the captive samples (from rescue centers), and from wild populations never typed before (Dune; Figure 6). Captive samples showed evidence of long-distance translocations and potential hybridization events. For example, 46% of the tortoises from the Veneto center were not assigned to any cluster, 27% were classified as imported from Greece (GRE), 19% were classified within the *T. h. hermanni* subspecies (Italian Peninsula, ITP), and only 8% were assigned to the MEC cluster (typically found in northeastern Italy). In the Emilia-Romagna center, 58% of the tortoises had Greek origins, and 6% were assigned to the MEC cluster. Approximately one-third of the tortoises from this center was not assigned. The centers of Emilia-Romagna and Lazio show a similar proportion of unassigned individuals (33% and 22%, respectively). On the contrary, the two centers show an opposite trend in terms of individuals assigned to the GRE (i.e., non-native origin) and ITP (i.e., local origin) clusters: 58% GRE and 6% ITP for the Emilia-Romagna center, 11% GRE and 56% ITP for the Lazio center. Interestingly, the tortoises from a wild and isolated population located in a natural reserve with fossil dunes corresponding to the former coastline of the Northern Adriatic Sea (Dune), were mainly assigned to the MEC cluster (78%; Figure 6). Populations from this cluster are typical in nearby areas (northeast Italy), showing a likely native component of this population.

4 | DISCUSSION

In this study, we present the development and the application of a small panel of SNP markers useful to investigate population genetic structure and geographical assignment in the endangered species Hermann's tortoise, *Testudo hermanni*. We initially performed a ddRAD-seq experiment on 70 individuals from different locations, and we showed that these loci, in terms of genetic structure and power to identify the macro-area of origin, outperform those obtained with a sample of 292 wild individuals previously typed at 7 STRs. Starting from these results, we designed a small panel of SNPs, retaining the most informative markers in terms of maximizing F_{ST} between the clusters inferred by the population genetic analysis. Finally, we tested the small SNP panel in 190 individuals using the KASP genotyping chemistry, a relatively inexpensive SNP genotyping approach. This cost-effective molecular tool enables a large number of confiscated individuals to be genotyped for geographical assignment, necessary for their management and their possible reallocation in the wild.

4.1 | Population genetic structure of *T. hermanni*: From STRs to SNPs

The overall results of the genetic structure of *T. hermanni* wild populations using SNPs confirmed the pattern based on the analysis of 7 STR loci found by Biello et al. (2021) but also revealed a more detailed structure. First, SNPs data were able to identify three genetically distinct groups in Calabria, a small region in the South of Italy, corresponding to samples collected in Northern, Central, and Southern areas, respectively. This result is not surprising given that this geographically heterogeneous region is a hotspot of genetic diversity for many temperate species and likely acted as a single or multiple glacial refugia (Canestrelli et al., 2010; Chiocchio et al., 2017). Considering however that the samples are not covering homogenously the whole region, and the clustering algorithm we applied tends to overestimate genetic structure under isolation by distance (Frantz et al., 2009), we cannot exclude that these clades do not refer to genetic isolates but to genetically distinct groups sampled within an area of large genetic variation but limited barriers and genetic discontinuities. Second, SNP data

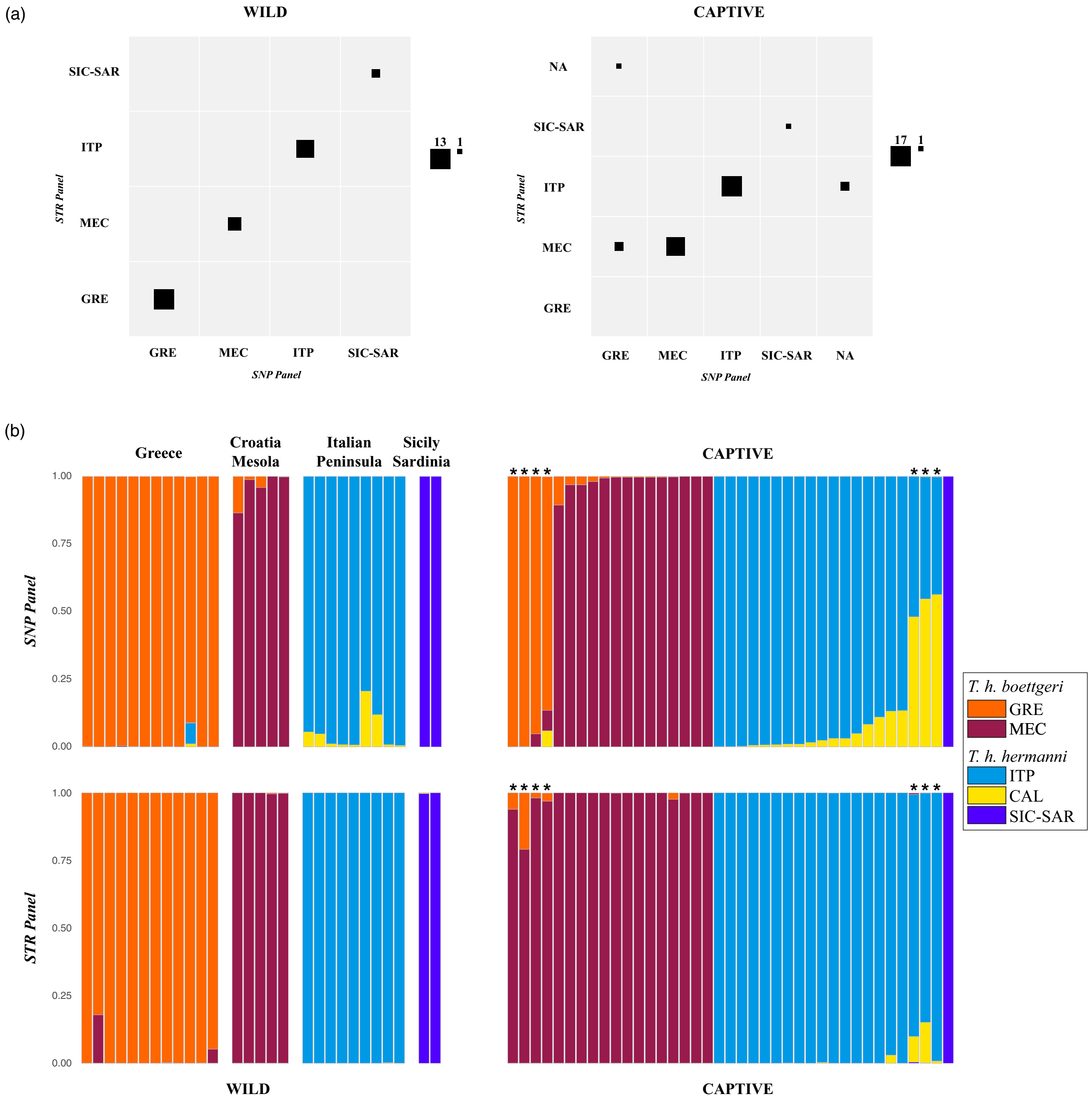

FIGURE 5 Contingency tables (a) and bar plots (b) of the geographical assignment of Wild (left) and Captive (right) samples to the *ddRAD-seq reference* and *STR reference* (see Table S1). (a) Columns correspond to the number of individuals assigned to original clusters using the *ddRAD-seq reference*, while rows correspond to the number of assignments using the *STR reference* (Biello et al., 2021): Square sizes are scaled by number of assignments to the cluster. SNP/STR Panels: GRE = Greece; MEC = Croatia and Bosco Mesola; ITP = Italian Peninsula; SIC-SAR = Sicily and Sardinia; NA = not assigned. * individuals assigned to different clusters by the two panels

suggest, differently from STRs, the existence of different genetic pools in Sicily and Sardinia. Finally, we note that even with a descriptive analysis such as the PCA, the SNPs dataset produces groups more easily distinguishable from each other, supporting the higher discriminatory power for these markers when compared to seven STRs. In summary, even if the samples present in the ddRAD-seq dataset included only 23% of the individuals included in the STR dataset, these were sufficient to understand better the genetic structure in this species, especially in emphasizing the fine-scale structure among wild populations.

4.2 | Design and validation of the SNP panel

Custom species-specific SNP panels, including high-ranking loci, have been shown to be highly informative for assignment

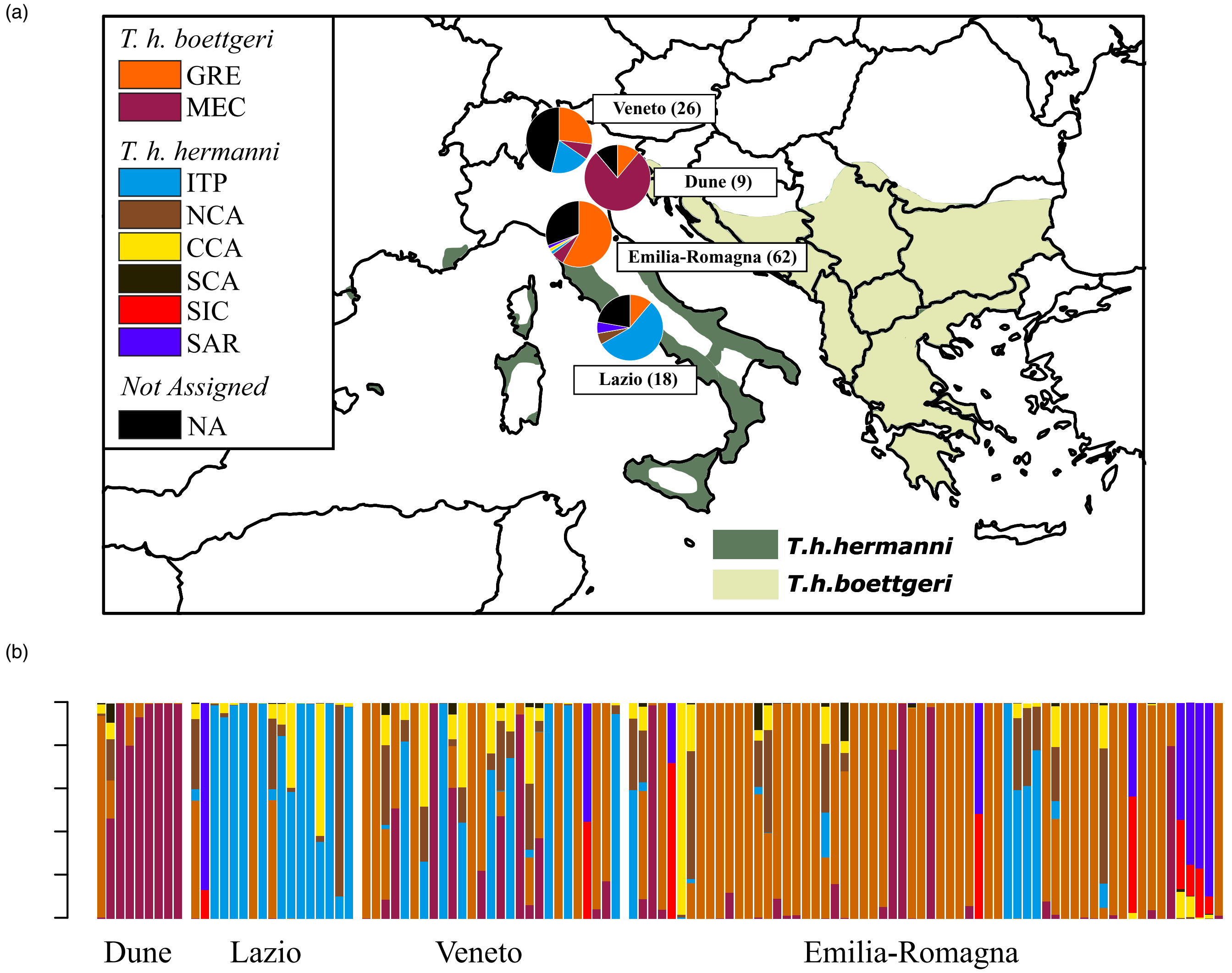

FIGURE 6 Geographic assignment of captive samples from three Italian rescue centers (Veneto, Emilia-Romagna, and Lazio) and one wild population (Dune). Local assignments for each location are showed in the pie charts on the map (in brackets the samples size). GRE = Greece; MEC = Croatia and Bosco Mesola; ITP = Italian peninsula; NCA = northern Calabria; CCA = Central Calabria; SCA = Southern Calabria; SIC = Sicily; SAR = Sardinia; NA = not assigned samples

studies (Förster et al., 2018; Jenkins et al., 2019; Kleinman-Ruiz et al., 2017) and prove to be an important resource for identifying the geographic origin of species in a conservation and forensic perspective.

We tested different methods for selecting highly informative SNPs and assuming that the target of the SNP number in the panel should be low to reduce costs. Using a panel of 48 SNPs, the F_{ST} method outperformed the Random Forest and PCA. The F_{ST} panel had a consistent higher self-assignment accuracy compared to those selected by PCA and RF. However, self-assignment of the largest panel (96 SNPs) showed a better accuracy of PCA method over F_{ST} and RF. This pattern is in discordance with what was found in Sylvester et al. (2018), where RF-based panels almost always outperformed F_{ST}-based panels. The higher accuracy of the F_{ST} panel over the RF panel may be due to the fact that genetic structure in *T. hermanni* species is higher than in salmon populations, resulting in a higher fixation index (Sylvester et al., 2018). Interestingly, approximately half of the loci identified by each method in the 48 SNP panel are not shared among methods (approximately one third in the 96 SNPs panel), indicating that algorithms optimized the selection in very different ways (although the results were not dramatically different).

Given the relatively high overall self-assignment accuracy of the 48 SNPs panel based on F_{ST} (above 80%) and considering that the genotyping cost for a panel with twice as much of SNPs implied a twofold cost increase, we selected the smallest panel for further genotyping to achieve a trade-off between accuracy and costs.

For the genotyping of our SNP panel, we adopted the SNP typing platform, KASP-by-Design Fluidigm Assays (LGC Genomics; He et al., 2014). The SNP panel we have developed is extremely cost-effective: users only need to extract less than 500 nanograms of DNA (10 ng per sample per SNP for a genome size ranging from 2.0 to 3.5 Gbp) from any tissue and send it to the genotyping service, which will return easy-to-interpret table format output files. The genotyping cost is affordable, and data for 190 samples is provided for 1338 € (7 € per sample). By comparison, for STRs, genotyping costs include fluorescently labeled oligos, PCR chemistry, and hot start polymerase, with subsequent fragment analysis on a capillary sequencer. Since there are currently no multiplex PCRs for the 7 *T. hermanni*'s STRs, outsourcing the entire service is extremely expensive (Table 2). One option to reduce the cost of STR analysis is to carry out part of the wet lab internally, although KASP remains much cheaper (Table 2).

4.3 | Test of the SNP panel

We observed a high genotyping success rate when we tested the panel with 190 samples, even with a low DNA yield (500 ng), which is significant in the case of noninvasively collected samples. We showed that missing data were below 1.9% for every SNP in our final panel, which is low compared with higher rates of missing data observed especially in STR genotyping of species of conservation concern where the sampling should not be invasive (Kraus et al., 2015). A few SNPs (seven) failed due to technical problems, most probably caused by nonspecific annealing of primers in multiple genomic locations. The 41 SNPs performed very well in terms of individual identification.

Assigning individuals to their area of origin was highly accurate (100%) when we considered samples from wild populations with known geographic origin and for which the reference (*ddRAD-seq reference*) was available. These results were also confirmed by using a panel of STR markers (Biello et al., 2021). This is in accordance with previous studies, which showed that reduced panels of highly informative SNPs (<100 SNPs) performed as well as or better than the traditional number (10–20) of STR loci used for individual identification (Glover et al., 2010). We also showed that assigning individuals from captive populations to their location of origin was not completely concordant between the *ddRAD-seq reference* and the *STR reference*. Approximately 17% of samples were not assigned to the same groups or not assigned to any group. These samples could represent descendent of crosses in captivity between individuals with different origin (including hybrids between subspecies), or to the lack of major genetic clusters in the reference populations.

Among the tortoises with unknown origin that we genotyped for the first time and are currently hosted in Italian rescue centers, we found evidence of long-distance translocations of individuals from Balkan regions and individuals with unclear genotype (i.e., potentially hybrids). This pattern confirms the Balkan areas as a source of illegal trading and the presence of hybrids in the Italian rescue centers due to mating occurring in captivity between individuals with different origins (see Biello et al., 2021).

Most of the tortoises from a wild population located in a natural reserve with fossil dunes corresponding to the former coastline of the Northern Adriatic Sea (Dune), which was never sampled before, were assigned to the genetic cluster presents also in other areas in North-East Italy (Bosco Mesola) and Croatia. However, we also found non-native tortoises belonging to genetic cluster observed in Greek individuals.

4.4 | Implications for management and perspectives

Approximately 2 million tortoises were exported from the former Yugoslavia to different European countries in the last century, especially after the Second World War and until the 80's (Ljubisavljević et al., 2011). A large fraction of them were *Testudo hermanni* harvested for the pet trade, and Italy was a major destination and virtually the only one where local populations already existed and introgression of non-native genomes could have occurred. Besides the possible translocation in recent and historical times of native individuals across the Italian peninsula and the Mediterranean islands (Perez et al., 2014), several thousands of individuals of Eastern origin, and their descendants, are therefore likely present in Italy in captivity (private and public seizures, including those with confiscated animals) and, probably at low frequencies (Biello et al., 2021; Perez et al., 2014), in some wild population. Bech et al. (2022) revealed a relatively high level of hybridization, between the two subspecies, as compared to previous estimations (Perez et al., 2014; Zenboudji et al., 2016) in a wild population in the Var district (France). The massive release (more than 4300 captive individuals; Devaux, 1990) carried out in the 1990s to reinforce declining populations most

TABLE 2 Cost per sample for 7 STR markers (Biello et al., 2021; Perez et al., 2014) and 41 SNPs (present work) for an increasing number of samples

Number of samples	7 STRs (in-house PCR, outsourced fragment analysis)	7 STRs (outsourced PCR and fragment analysis)	41 SNP panel (outsourced KASP, LGC genomics)
12	23.3€	34.4€	–
24	23.3€	34.4€	14.2€
48	23.3€	34.4€	10.6€
96	18.8€	30.1€	9.4€
192	18.8€	30.1€	7.0€

Note: STRs: fragment analysis quotation with or without PCR amplification (second and third columns, respectively) by Macrogen Europe; KASP quotation by LGC Genomics, UK (October 2021). For KASP, a minimum of 22 samples is required to determine the genotype of all the data points by means of genotyping cluster generation.

likely introduced *T. h. boettgeri* individuals or hybrids in this area. This highlights the importance of monitoring the genetic composition of wild populations and selecting genetically suitable individuals for reintroduction projects. Considering that this species is endangered with a very patchy natural distribution in Italy (as well as in Spain and France), these measures should be carefully evaluated as a necessary step in conservation plans.

In a previous paper we introduced a panel of 7 STR markers useful to reach this goal (Biello et al., 2021). It was an important advancement for the assignment of individuals of unknown origin, based on a large reference dataset of 461 wild individuals collected across most of the distribution range of this species. Here, we showed a new development based on a preliminary ddRAD-seq genomics study that allowed the introduction of an informative and cost-effective panel of biallelic markers suitable for the *T. hermanni* protection. This study, describing in detail the necessary methodological steps of the process, represents also a useful example for the development of similar tools in other species with similar management needs.

The SNP panel introduced in this study allows the genotyping of a large number of samples at low cost. Compared to the development and the typing of a set of similarly informative STR markers, this approach is cheaper, faster; requires less handling; and provides immediate standardization across laboratories. At the moment, however, the reference database of wild populations for the SNPs panel is still smaller and geographically less representative of the distribution range of this species compared to the STRs database, and we therefore recommend, until more wild individuals will be typed, the use of the SNPs panel integrated for difficult and especially forensic assignments with the analysis of the STR markers.

Hundreds of tortoises are currently maintained in captivity in breeding and rescue centers, providing a highly valuable source of individuals for reintroduction and reallocation projects. A fraction of these individuals, selected considering their health conditions, sex, age, and genetic compositions, could represent not only a demographic and genetic supplementation of small natural populations, but also the founders of new wild populations in ecologically suitable areas where this species was present in the past. These interventions would favor the well-being of the captive tortoises but also the human well-being by creating more natural and biodiverse environments (Fuller et al., 2007; Kuo, 2015).

ACKNOWLEDGMENTS

We thank Parco Natura Viva (Verona, Italy), Maurizio Bellavista, Michele Capasso, Giovanni Nobili, Fabrizio Bellini, and all the Carabinieri Forestali at Punta Marina (Ravenna) for helping and providing samples from captive populations. We also thank Emanuele Lubian and Annalaura Mancia for their help in collecting the specimens from Riserva Naturale Orientata Dune Fossili Di Massenzatica (Ferrara, Italy). We also thank Noemi Giurintano, Camilla Broggini and Maria Luisa Boglino for their help in laboratory work. This study was funded by the University of Ferrara, the Tuscia University (Viterbo), and the Carabinieri Forestali.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

Raw demultiplexed sequences are available on the Sequence Read Archive (SRA) on the study accession number: PRJNA777783.

ORCID

Roberto Biello https://orcid.org/0000-0002-5916-884X
Silvia Fuselli https://orcid.org/0000-0002-1442-6171
Roberta Bisconti https://orcid.org/0000-0002-0600-7436
Andrea Chiocchio https://orcid.org/0000-0002-0067-7025
Emiliano Trucchi https://orcid.org/0000-0002-1270-5273
Daniele Canestrelli https://orcid.org/0000-0001-9351-4972
Giorgio Bertorelle https://orcid.org/0000-0002-2498-2702

REFERENCES

Anderson, E. C. (2010). Assessing the power of informative subsets of loci for population assignment: Standard methods are upwardly biased. *Molecular Ecology Resources*, *10*, 701–710. https://doi.org/10.1111/j.1755-0998.2010.02846.x

Barbosa, S., Hendricks, S. A., Funk, W. C., Rajora, O. P., & Hohenlohe, P. A. (2020). *Wildlife population genomics: Applications and approaches*. Springer. https://doi.org/10.1007/13836_2020_83

Bech, N., Nivelle, D., Caron, S., Ballouard, J. M., Arnal, V., Arsovski, D., Golubović, A., Bonnet, X., & Montgelard, C. (2022). Extent of introgressive hybridization in the Hermann's tortoise (*Testudo hermanni hermanni*) from the south of France. *European Journal Wildlife Research*, *68*, 37. https://doi.org/10.1007/s10344-022-01585-8

Bertolero, A., Cheylan, M., Hailey, A., Livoreil, B., & Willemsen, R. E. (2011). *Testudo hermanni* (Gmelin 1789)— Hermann's tortoise. Conservation biology of freshwater turtles and tortoises: A compilation project of the IUCN/SSC tortoise and freshwater turtle specialist group. *Chelonian Research Monographs*, *5*, 059–061. https://doi.org/10.3854/crm.5.059.hermanni.v1.2011

Biello, R., Zampiglia, M., Corti, C., Deli, G., Biaggini, M., Crestanello, B., & Canestrelli, D. (2021). Mapping the geographic origin of captive and confiscated Hermann's tortoises: A genetic toolkit for conservation and forensic analyses. *Forensic Science International: Genetics*, *51*, 102447. https://doi.org/10.1016/j.fsigen.2020.102447

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, *368*, 455–457. https://doi.org/10.1038/368455a0

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Campbell, E. O., Brunet, B. M. T., Dupuis, J. R., & Sperling, F. A. H. (2018). Would an RRS by any other name sound as RAD? *Methods in Ecology and Evolution*, *9*, 1920–1927. https://doi.org/10.1111/2041-210X.13038

Canestrelli, D., Aloise, G., Cecchetti, S., & Nascetti, G. (2010). Birth of a hotspot of intraspecific genetic diversity: Notes from the underground. *Molecular Ecology*, *19*, 5432–5451. https://doi.org/10.1111/j.1365-294X.2010.04900.x

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*, 171–182. https://doi.org/10.1534/g3.111.000240

Chen, K. Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). AssignPOP: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution*, *9*, 439–446. https://doi.org/10.1111/2041-210X.12897

Chiocchio, A., Bisconti, R., Zampiglia, M., Nascetti, G., & Canestrelli, D. (2017). Quaternary history, population genetic structure and diversity of the cold-adapted alpine newt *Ichthyosaura alpestris* in peninsular Italy. *Scientific Reports*, *7*, 1–12. https://doi.org/10.1038/s41598-017-03116-x

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*, 499–510. https://doi.org/10.1038/nrg3012

Devaux, B. (1990). Réintroduction de tortues d'Hermann (*Testudo hermanni hermanni*) dans le massif des Maures. *Revue d'Ecologie*, *Terre et Vie*, *5*, 291–297.

Dixon, P., & Palmer, M. W. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eriksson, C. E., Ruprecht, J., & Levi, T. (2020). More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. *Molecular Ecology Resources*, *20*, 1505–1516. https://doi.org/10.1111/1755-0998.13208

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

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Feutry, P., Devloo-Delva, F., Tran Lu, Y. A., Mona, S., Gunasekera, R. M., Johnson, G., & Kyne, P. M. (2020). One panel to rule them all: DArTcap genotyping for population structure, historical demography, and kinship analyses, and its application to a threatened shark. *Molecular Ecology Resources*, *20*, 1470–1485. https://doi.org/10.1111/1755-0998.13204

Förster, D. W., Bull, J. K., Lenz, D., Autenrieth, M., Paijmans, J. L. A., Kraus, R. H. S., Saveljev, A. P., Nowak, C., Bayerl, H., Kuehn, R., Sindičić, M., Hofreiter, M., Schmidt, K., & Fickel, J. (2018). Targeted re-sequencing of coding DNA sequences for SNP discovery in non-model species. *Molecular Ecology Resources*, *18*, 1356–1373. https://doi.org/10.1111/1755-0998.12924

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, *46*, 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Fuller, R. A., Irvine, K. N., Devine-Wright, P., Warren, P. H., & Gaston, K. J. (2007). Psychological benefits of greenspace increase with biodiversity. *Biology Letters*, *3*, 390–394. https://doi.org/10.1098/rsbl.2007.0149

Garvin, M. R., Saitoh, K., & Gharrett, A. J. (2010). Application of single nucleotide polymorphisms to non-model species: A technical review. *Molecular Ecology Resources*, *10*, 915–934. https://doi.org/10.1111/j.1755-0998.2010.02891.x

Glover, K. A., Hansen, M. M., Lien, S., Als, T. D., Høyheim, B., & Skaala, Ø. (2010). A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics*, *11*, 1–12. https://doi.org/10.1186/1471-2156-11-2

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in alpine ibex. *Nature Communications*, *11*, 1–12. https://doi.org/10.1038/s41467-020-14803-1

He, C., Holme, J., & Anthony, J. (2014). SNP genotyping: The KASP assay. In D. Fleury & R. Whitford (Eds.), *Crop breeding: Methods and protocols* (pp. 75–86). Springer. https://doi.org/10.1007/978-1-4939-0446-4_7

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R., & Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, *11*, 123–136. https://doi.org/10.1111/j.1755-0998.2010.02943.x

Henriques, D., Parejo, M., Vignal, A., Wragg, D., Wallberg, A., Webster, M. T., & Pinto, M. A. (2018). Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). *Evolutionary Applications*, *11*, 1270–1282. https://doi.org/10.1111/eva.12623

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, *11*, 117–122. https://doi.org/10.1111/j.1755-0998.2010.02967.x

Holderegger, R., Balkenhol, N., Bolliger, J., Engler, J. O., Gugerli, F., Hochkirch, A., Nowak, C., Segelbacher, G., Widmer, A., & Zachos, F. E. (2019). Conservation genetics: Linking science with practice. *Molecular Ecology*, *28*, 3848–3856. https://doi.org/10.1111/mec.15202

Humble, E., Paijmans, A. J., Forcada, J., & Hoffman, J. I. (2020). An 85K SNP array uncovers inbreeding and cryptic relatedness in an Antarctic fur seal breeding colony. *G3: Genes*, *Genomes*, *Genetics*, *10*, 2787–2799. https://doi.org/10.1534/g3.120.401268

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

Jenkins, T. L., Ellis, C. D., Triantafyllidis, A., & Stevens, J. R. (2019). Single nucleotide polymorphisms reveal a genetic cline across the north-East Atlantic and enable powerful population assignment in the European lobster. *Evolutionary Applications*, *12*, 1881–1899. https://doi.org/10.1111/eva.12849

Johnston, S. E., Huisman, J., Ellis, P. A., & Pemberton, J. M. (2017). A high-density linkage map reveals sexual dimorphism in recombination landscapes in red deer (*Cervus elaphus*). *G3: Genes*, *Genomes*, *Genetics*, *7*, 2859–2870. https://doi.org/10.1534/g3.117.044198

Jombart, T. (2008). *Adegenet*: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J., & Godoy, J. A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, *18*, 1–12. https://doi.org/10.1186/s12864-017-3946-5

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*, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kraus, R. H. S., vonHoldt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., Förster, D. W., Fickel, J., Roos, C., & Nowak, C. (2015).

A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources*, *15*, 295–305. https://doi.org/10.1111/1755-0998.12307

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., Gomulkiewicz, R., Klein, E. K., Ritland, K., Kuparinen, A., Gerber, S., & Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, *15*, 378–392. https://doi.org/10.1111/j.1461-0248.2012.01746.x

Kuo, M. (2015). How might contact with nature promote human health? Promising mechanisms and a possible central pathway. *Frontiers in Psychology*, *6*, 1093. https://doi.org/10.3389/fpsyg.2015.01093

Lang, A. R., Weller, D. W., Burdin, A. M., Robertson, K., Sychenko, O., Urbán, J., Martinez-Aguilar, S., Pease, V. L., LeDuc, R. G., Litovka, D. I., Burkanov, V. N., & Brownell, R. L., Jr. (2021). Population structure of North Pacific gray whales in light of trans-Pacific movements. *Marine Mammal Science*, *38*, 433–468. https://doi.org/10.1111/mms.12875

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*, 18–22.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*, 298–299. https://doi.org/10.1093/bioinformatics/btr642

Ljubisavljević, K., Džukić, G., & Kalezić, M. L. (2011). The commercial export of the land tortoises (*testudo* spp.) from the territory of the former Yugoslavia: A historical review and the impact of overharvesting on wild populations. *North-Western Journal of Zoology*, *2*, 250–260.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution*, *35*, 1284–1290. https://doi.org/10.1093/molbev/msy023

Meek, M. H., Baerwald, M. R., Stephens, M. R., Goodbla, A., Miller, M. R., Tomalty, K. M. H., & May, B. (2016). Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. *Ecology and Evolution*, *6*, 7706–7716. https://doi.org/10.1002/ece3.2493

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*, 240–248. https://doi.org/10.1101/gr.5681207

Ogden, R., Baird, J., Senn, H., & McEwing, R. (2012). The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: A case study from Arabian and scimitar-horned oryx. *Conservation Genetics Resources*, *4*, 471–473. https://doi.org/10.1007/s12686-011-9577-2

Ogden, R., & Linacre, A. (2015). Wildlife forensic science: A review of genetic geographic origin assignment. *Forensic Science International: Genetics*, *18*, 152–159. https://doi.org/10.1016/j.fsigen.2015.02.008

Perez, M., Livoreil, B., Mantovani, S., Boisselier, M.-C., Crestanello, B., Abdelkrim, J., Bonillo, C., Goutner, V., Lambourdiere, J., Pierpaoli, M., Sterijovski, B., Tomovic, L., Vilaca, S. T., Mazzoti, S., & Bertorelle, G. (2014). Genetic variation and population structure in the endangered Hermann's tortoise: The roles of geography and human-mediated processes. *Journal of Heredity*, *105*, 70–81. https://doi.org/10.1093/jhered/est071

Pertoldi, C., Tokarska, M., Wójcik, J. M., Demontis, D., Loeschcke, V., Gregersen, V. R., Coltman, D., Wilson, G. A., Randi, E., Hansen, M. M., & Bendixen, C. (2009). Depauperate genetic variability detected in the American and European bison using genomic techniques. *Biology Direct*, *4*, 1–7. https://doi.org/10.1186/1745-6150-4-48

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. https://doi.org/10.1371/journal.pone.0037135

Piry, S., Alapetite, A., Cornuet, J. M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, *95*, 536–539. https://doi.org/10.1093/jhered/esh074

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

Rajora, O. P. (2019). In O. P. Rajora (Ed.), *Population genomics*. Springer International Publishing. https://doi.org/10.1007/978-1-4471-5304-7_17

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 9197–9201. https://doi.org/10.1073/pnas.94.17.9197

Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in isle Royale wolves, a population on the threshold of extinction. *Science Advances*, *5*, eaau0757. https://doi.org/10.1126/sciadv.aau0757

Roques, S., Chancerel, E., Boury, C., Pierre, M., & Acolas, M. L. (2019). From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. *Ecology and Evolution*, *9*, 7017–7029. https://doi.org/10.1002/ece3.5268

Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for windows and Linux. *Molecular Ecology Resources*, *8*, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, *11*, 1–8. https://doi.org/10.1111/j.1755-0998.2010.02979.x

Sinding, M.-H. S., Gopalakrishan, S., Vieira, F. G., Castruita, J. A. S., Raundrup, K., Jørgensen, M. P. H., Meldgaard, M., Petersen, B., Sicheritz-Ponten, T., Mikkelsen, J. B., Marquard-Petersen, U., Dietz, R., Sonne, C., Dalen, L., Bachmann, L., Wiig, O., Hansen, A. J., & Gilbert, M. T. P. (2018). Population genomics of grey wolves and wolf-like canids in North America. *PLoS Genetics*, *14*, e1007745. https://doi.org/10.1371/journal.pgen.1007745

Stubbs, D., Swingland, I. R., Hailey, A., & Pulford, E. (1985). The ecology of the Mediterranean tortoise *Testudo hermanni* in northern Greece (the effects of a catastrophe on population structure and density). *Biological Conservation*, *31*, 125–152. https://doi.org/10.1016/0006-3207(85)90045-X

Sylvester, E. V. A., Bentzen, P., Bradbury, I. R., Clément, M., Pearce, J., Horne, J., & Beiko, R. G. (2018). Applications of random forest feature selection for fine-scale genetic population assignment. *Evolutionary Applications*, *11*, 153–165. https://doi.org/10.1111/eva.12524

Vavrek, M. J. (2011). Fossil: Palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, *14*, 16.

von Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T. E., Alves, P. C., Lyons, L. A., Mattucci, F., Randi, E., Cragnolini, M., Galian, J., Hegyeli, Z., Kitchener, A. C., Lambinet, C., Lucas, J. M., Molich, T., Ramos, L., Schockert, V., & Cocchiararo, B. (2020). Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Molecular Ecology Resources*, *20*, 662–680. https://doi.org/10.1111/1755-0998.13136

VonHoldt, B. M., Cahill, J. A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J. P., Shapiro, B., Wall, J., & Wayne, R. K. (2016). Whole-genome sequence analysis shows that two endemic species of north American wolf are admixtures of the coyote and gray wolf. *Science Advances*, *2*, e1501714. https://doi.org/10.1126/sciadv.1501714

Wang, J. (2017). The computer program structure for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources*, *17*, 981–990. https://doi.org/10.1111/1755-0998.12650

Zenboudji, S., Cheylan, M., Arnal, V., Bertolero, A., Leblois, R., Astruc, G., Bertorelle, G., Pretus, J. L., Lo Valvo, M., Sotgiu, G., & Montgelard, C. (2016). Conservation of the endangered Mediterranean tortoise *Testudo hermanni hermanni*: The contribution of population genetics and historical demography. *Biological Conservation*, *195*, 279–291. https://doi.org/10.1016/j.biocon.2016.01.007

Zimmerman, S. J., Aldridge, C. L., & Oyler-McCance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics*, *21*, 1–16. https://doi.org/10.1186/s12864-020-06783-9

SUPPORTING INFORMATION

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

How to cite this article: Biello, R., Zampiglia, M., Fuselli, S., Fabbri, G., Bisconti, R., Chiocchio, A., Mazzotti, S., Trucchi, E., Canestrelli, D., & Bertorelle, G. (2022). From STRs to SNPs via ddRAD-seq: Geographic assignment of confiscated tortoises at reduced costs. *Evolutionary Applications*, *15*, 1344–1359. https://doi.org/10.1111/eva.13431