RESEARCH ARTICLE

Genetic differentiation of a critically endangered population of the limpet *Patella candei candei* d'Orbigny, 1840, in the Canary Islands

Javier Quinteiro[1] · Gustavo González-Lorenzo[2] · Dailo Hernández-Reyes[3] · Lara Quinteiro[1] ·
Rogelio Herrera-Pérez[4] · Juán Martínez-Barrio[4] · José Antonio González[3] · Manuel Rey-Méndez[1] ·
Nieves González-Henríquez[3]

Abstract
The adoption of measures to protect the viability of threatened populations should be supported by empirical data identifying appropriate conservation units and management strategies. The global population of the majorera limpet, *P. candei candei* d'Orbigny, 1840, is restricted to the Macaronesian islands in the NE Atlantic, including near-to-extinct and healthy populations in Fuerteventura and Selvagens, respectively. The taxonomic position, genetic diversity and intra- and interspecific relationships of these populations are unclear, which is hindering the implementation of a recovery plan for the overexploited majorera limpet on Fuerteventura. In this study, ddRAD-based genome scanning was used to overcome the limitations of mitochondrial DNA-based analysis. As a result, *P. candei candei* was genetically differentiated from the closely related *P. candei crenata* for the first time. Moreover, genetic differentiation was detected between *P. candei candei* samples from Selvagens and Fuerteventura, indicating that translocations from the healthy Selvagens source population are inadvisable. In conclusion, the majorera limpet requires population-specific management focused on the preservation of exceptional genetic diversity with which to face future environmental challenges.

Keywords Conservation genetics · Threatened population · ddRAD · Gastropoda

Introduction

Invertebrates inhabiting the upper intertidal zone of oceanic islands are vulnerable to diverse anthropogenic effects, including overexploitation of species for human consumption. Protective strategies should focus on preserving the viability of threatened conservation units, especially in those exploited species with poorly resolved taxonomy and limited geographic distribution. Conservation initiatives and regulatory frameworks rely on appropriate definitions of distinct conservation units to support implementation and efficient resource allocation (Supple and Shapiro 2018). The taxonomic level of a species does not necessarily coincide with a conservation unit but it is a good starting point to have a robust species delineation, with consensus across various species concepts and data types. When empirical data support them, divergent populations that represent the genetic diversity of a species can be considered evolutionary units for conservation (Coates et al. 2018).

The Macaronesian archipelagos constitute the northeastern Atlantic distribution of the endemic limpet species complex classified as *Patella candei* d'Orbigny, 1840, which includes four subspecies: *P. candei gomesii* Drouët (1858) in the Azores; *P. candei ordinaria* Mabille 1888 in Madeira; *P. candei crenata* d'Orbigny 1840 in the Canary Islands and *P. candei candei* d'Orbigny 1840 in the Selvagens and Canary Islands (Christiaens 1973; Hernández-Dorta 1992; Weber and Hawkinks 2002). The majorera limpet, *P. candei candei* d'Orbigny 1840, is endemic to the islands of

✉ Javier Quinteiro
javier.quinteiro@usc.es

1 Molecular Systematics Laboratory, Department of Biochemistry and Molecular Biology, Aquaculture Institute, University of Santiago de Compostela, A Coruña, Galicia, Spain

2 Canary Islands Oceanographic Centre, Spanish Institute of Oceanography (IEO, CSIC), Tenerife, Spain

3 Department of Biology, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

4 Directorate General for the Fight against Climate Change and the Environment. Council for the Ecological Transition, Fight against Climate Change and Territorial Planning, Canary Islands Government, Spain

Fuerteventura and Selvagens. Currently, these populations represent the global distribution of this limpet, which is highly valued locally for human consumption. The critical status of the majorera limpet population on Fuerteventura in the Canary Islands prompted its inclusion in the Spanish Catalogue of Endangered Species under the category 'in danger of extinction' (González-Lorenzo et al. 2015). In contrast, the population on the inhabited archipelago of Selvagens islands is healthy, although the cliff habitat is considered unfavourable for development (Hernández-Dorta 1992; González-Lorenzo et al. 2015).

The Fuerteventuran population has shown a decline in abundance, following a reduction in recolonisation events, loss of connectivity among reproductive populations, low efficacy of conservation strategies and legal protection and uncertainty regarding its taxonomic and historical geographical distribution (González-Lorenzo et al. 2015). The critical conservation status of the *P. candei candei* population in the Canary Islands requires conclusive data for the design, planning and implementation of a recovery plan, supported by the maximum possible knowledge to guarantee its success. This involves elucidation of the taxonomic position of *P. c. candei*, estimation of genetic diversity and assessment of intra- and interspecific relationships.

Previous phylogenetic and population genetic studies have not yet clarified the taxonomic position of the populations of *P. c. candei* and the phylogenetic relationships within this complex. The mitochondrial 12S/16S rRNA molecular phylogeny of the *Patella* genus confirms close relationships among *P. candei*, *P. lugubris* and *P. caerulea* (Koufopanou et al. 1999). The analysis of independent and combined data sets involving COX1, 12S and 16S rRNA sequences revealed the structure of mitochondrial lineages found among the Macaronesian populations: the paraphyly of *P. candei*, a clade containing the Azores, Madeira and Desertas samples, and a second clade including individuals from the Selvagens and Canaries and *P. lugubris* from the Cape Verde Islands (Sá-Pinto et al. 2005). These same lineages were further investigated in detail to elucidate the population structure and gene flow among Macaronesian populations. As a result, mitochondrial data supported two *P. candei* sub-clades related to *P. c. gomesii* and *P. c. ordinaria* subspecies in Azores and Madeira archipelagos, respectively (Sa-Pinto et al. 2008; Carreira et al. 2017). However, the phylogenetic status of *P. c. crenata* in the Canary Islands and *P. c. candei* in the Selvagens and Canary Islands remains unclear.

The analysis of nuclear short tandem repeat (STR) allowed to retrieve three clades: (i) Azores, (ii) Madeira and (iii) Gran Canaria in the Canary Islands, in total congruence with expectations from mitochondrial DNA phylogeography (Faria et al. 2017). In addition, shell shape also showed differentiation among populations (Carreira et al. 2017; Faria et al. 2017).

However, to date, no useful information is available to integrate into the evaluation of conservation strategies for *P. p. candei* since (i) mitochondrial analyses recovered a single clade, without any differentiation between the samples from the Canaries and Selvagens (Sá-Pinto et al. 2005; Sa-Pinto et al. 2008; Carreira et al. 2017) and (ii) STR analysis does not include samples from Selvagens or Fuerteventura, obviating the *P. c. candei* taxa (Faria et al. 2017).

The current accessibility to genomic information allows to obtain data, useful for conservation, at a resolution, sensitivity and precision not previously achievable (Supple and Shapiro 2018). In particular, the use of genome reduced-representation methodological approaches, e.g. restriction-site associated DNA sequencing (RAD), has enabled researchers to obtain thousands of markers dispersed through the genome in a cost-effective way (Baird et al. 2008; Carleton 2011; Peterson et al. 2012; Kess et al. 2015). This high-resolution data is suitable for reviewing hypotheses previously evaluated with partial mitochondrial data (Sá-Pinto et al. 2005; Sa-Pinto et al. 2008; Carreira et al. 2017). In such cases, the conclusions achieved may be affected by a lack of resolution caused by a recent divergence of species, introgressive hybridization events or other features of the evolutionary dynamics of mitochondrial DNA (Avise et al. 1987; Moritz et al. 1987).

This study focused in (i) the delimitation of the closely related taxa inhabiting the Selvagens and Canary Islands archipelagos (*P. c. candei* and *P. c. crenata*), and (ii) the population genetics of a threatened endemic Fuerteventurean limpet (*P. c. candei*) for assistance in conservation and management. Accordingly, we assessed species delimitation patterns using all available mitochondrial COX1 data and performed a genome scan based on double digest restriction-site associated DNA (ddRAD) to infer an alternative species delimitation and the population structure pattern relevant to the conservation of the majorera limpet.

Materials and methods

Sample collection

The available tissue included frozen (-20ºC) ethanol-preserved samples initially collected in 1995, 2010 and 2014 for mitochondrial DNA analysis of *P. candei gomesii* Drouet, 1858; *P. candei ordinaria* Mabille, 1888; *P. candei crenata* d'Orbigny, 1840; and *P. candei candei* d'Orbigny, 1840. Additional samples from *P. candei crenata* and *Patella lugubris* Gmelin, 1791, were sampled in 2020–2021. The sample subset selected for genome-wide data

analysis included *P. candei candei* samples from Selvagens ($N=30$) and *P. candei crenata* from Fuerteventura ($N=15$) and Tenerife ($N=15$). Specimens from the *P. candei candei* threatened population from Fuerteventura (Canaries) were available from (i) a 1995 collection when no restrictions on capture were implemented ($N=5$), (ii) the seizure of illegal captures ($N=3$) and (iii) non-invasive mucosal swab sampling ($N=2$) (Table 1).

DNA isolation, COX1 amplification and sequencing

Total DNA was isolated with an E.Z.N.A.® Mollusc DNA kit from a muscle tissue sample (10–30 mg) and mucosal swab samples, spectrophotometrically quantified with a Biophotometer Eppendorf D30, electrophoretically characterised and stored at -20ºC.

A fragment (573 bp length) located in the 5' portion of the COX1 gene was amplified with the PATCOX1F (5'-CCTAGGRATATGRGCAGGTTTAGTAGG-3') and PATCOX1R (5'-TCAGTTAATARTATAGTAATTGCYC-CRGCTA-3') primers. Reactions were performed in 15 µL volume containing GoTaq Flexi Buffer (Promega, Madison, WI, USA). PCR products were purified and sequenced following previously described protocols (Quinteiro et al. 2020). After chromatogram edition, sequences were aligned using Clustal X (Thompson et al. 1997) implemented in Geneious Prime (http://www.geneious.com).

Table 1 Sampling locations and data generated for Macaronesian limpet taxa

Specie	Sampling location	N	mtDNA seqs	ddRAD data
Patella candei gomesii	São Jorge (Azores)	14	14	
	Faial (Azores)	9	8	
Patella candei ordinaria	Madeira	6	6	
	Porto Santo (Madeira)	5	5	
	Desertas (Madeira)	6	6	
Patella candei crenata	Gran Canaria (Canaries)	10	10	
	Tenerife (Canaries)	34	32	15
	Fuerteventura (Canaries)	35	35	15
Patella candei candei	Selvagens	48	48	30
	Fuerteventura (Canaries)	10	9	10
Patella lugubris	Santiago (Cabo Verde)	6	6	
	Santa Luzia (Cabo Verde)	8	8	

Elaboration and sequencing of ddRAD libraries

Four reduced representation genomic libraries were elaborated using the ddRAD methodology (Baird et al. 2008; Carleton 2011; Peterson et al. 2012; Kess et al. 2015). In the preliminary methodology set-up, no significant differences among DNA samples replicates ($N=3$, 4 samples) were observed (Mastretta-Yanes et al. 2015). A total of 70 DNA samples were processed, to reduce genotyping error, in population specific libraries for *P. candei candei*, sampled in Selvagens ($N=30$, 2014) and Fuerteventura ($N=10$, 1995; 2021), and *P. candei crenata*, collected on Fuerteventura ($N=15$, 2020) and Tenerife ($N=15$, 2020).

The estimated genome size for Acmaeidae limpet species (Patellogastropoda) (C value $=0.60$, 586.8 Mb) (www.genomesize.com) and a draft assembly of transcriptome data (Werner et al. 2013) of *Patella vulgata* were used to estimate the most appropriate restriction enzyme combination for the selected size range and fragment yield with ddRADseqTools (Mora-Marquez et al. 2017). Total DNA was double digested with High Fidelity (HF®) *Eco*RI and *Hind*III restriction enzymes (NEB). The enzyme specific adapters were ligated to the fragments with T4 DNA Ligase (NEB). Adapter-fragments were gel-based size selected (300–500 bp) and purified with Nucleospin (Machery-Nagel). The fragments for each individual were labelled using dual-barcoded index PCR with Phusion™ High-Fidelity DNA Polymerase (Thermo Scientific) and gel purified. The individual libraries were quantified with KAPA Library Quantification Kit (Illumina®) universal qPCR (Roche) to elaborate an equimolar pool. A 12.5 nM library was used for sequencing with MiSeq Reagent kit v2, 2 x 250 bp (Illumina).

Mitochondrial COX1 DNA analysis

A complete dataset was elaborated from 807 available Genbank COX1 sequences from the *Patella* genus. A phylogenetic neighbor-joining tree was inferred from distances computed using the Tamura-Nei 93 (Tamura and Nei 1993) selected model, with a shape parameter $=0.37$ for the gamma distribution of the rate variation among sites, using MEGA7 (Kumar et al. 2016). The p-distances were then estimated within and among those clades/species phylogenetically supported by a bootstrapping value >70 (2000 replicates). The sequences belonging to the *candei/lugubris* clades were excluded for the estimation of inter- and intraspecific distance reference values estimated from another *Patella* spp.

The mitochondrial COX1 phylogenetic inference within the *candei/lugubris* clade was based on all available Gen-Bank data for the involved taxa and the sequences obtained. *Patella caerulea* and *P. ferruginea* were established as

outgroups. The data set included 263 taxa with 515 sites. The phylogenetic relationships were analysed by Bayesian analysis with BEAST v2.6 (Drummond et al. 2012) and visualised with DensiTree (Bouckaert and Heled 2014). The previously selected Tamura-Nei 93 model (Tamura and Nei 1993), was implemented in BEAST adding the substitution rate parameter under a gamma distribution and the proportion of invariable sites, with the gamma and invariant sites options, respectively. A Yule tree prior was used, with default values, for tree reconstruction under strict molecular clock. Other options The MCMC chain length was 1 000 000.

Variant calling on ddRAD data

The raw and demultiplexed Illumina data were trimmed for quality and adapter content with Trimmomatic v0.35 (Bolger et al. 2014) and arranged into four population samples: Selvagens and Fuerteventura populations of *P. candei candei*, and Tenerife and Fuerteventura populations of *P. candei crenata*. These initial data were submitted to two different and alternative pipelines to minimize the effects of genotyping errors, filtering strategies and algorithms assumptions on the biological conclusions. The Stacks v2.60 pipeline (Catchen et al. 2013) started with an additional trimming with the *process radtags* v.1.44 program in a subset of 52 specimens with the topmost read number to minimize missing data (< 0.5%). The *denovo.pl* script was used for sequential running of the *ustacks*, *cstacks*, *sstacks* and *populations* modules. The *ustacks* (-*m* 3, -*M* 8, *N* 9) and *populations* (-*p* 4, -*r* 0.8) parameters were selected after evaluation of the number of single nucleotide polymorphisms (SNPs) and genetic differentiation estimates achieved from different values and combinations of *m* (3), *M* (2 to 8) and *N* (=M±1), and *p* (2 to 4), *R* (0.5–0.7) and *r* (0.8) (Paris et al. 2017; Diaz-Arce and Rodriguez-Ezpeleta 2019).

The dDocent v2.7.8 pipeline (Puritz et al. 2014) included read merging with the OL paired-end sequencing option in PEAR v0.9.8. CD-HIT (Fu et al. 2012) clustered representative sequences with similarity > 90%, followed by reads mapping to those sequences using the MEM algorithm of BWA (Li and Durbin 2009). Variant calling was performed with Freebayes v1.0.2 (Garrison and Gabor 2012). Finally, the SNP data were filtered using a script (https://github.com/jpuritz/dDocent/blob/master/scripts/dDocent_filters) for sites with missing data (≤ 0.01), quality scores (≥ 30), minor allele count (3), individuals with missing data (> 5%), minimum allele frequency (0.01) and including population specific test for HWE.

Population structure

Using the package SNAPP v2.4.8 (Bryant et al. 2012) within the BEAST2 platform (Bouckaert et al. 2019), a coalescent-based tree inference was conducted in a Bayesian framework directly from biallelic markers without estimating a gene tree for each locus. The Beauti prior settings included α = 2, β = 20 and θ = 0.1. The trees recorder during the MCMC analysis (chain length = 100 000) were visualized using the DensiTree program, reflecting the entire posterior distribution of the species tree. *Bitwise* distances among individuals were calculated using poppr (1000 bootstrapping replicates) and visualised with a dendrogram constructed with ape R programs (Kamvar et al. 2015; R-Core-Team 2021). Principal component analysis (PCA) of SNP data was performed with SNPRelate v1.16.0 (Zheng et al. 2012).

Genetic structure was also inferred by Bayesian model-based clustering of individuals with Structure v.2.3 (Pritchard et al. 2000) for the assignment of individuals to detected populations. Selected options included the admixture model for the ancestry estimates among individuals, the sampling locations as prior information (LOCPRIOR option) and the correlated allele frequencies model. The length of burn-in period was 200 000 and the number of MCMC was 2 000 000. The parameter $K = 1$ to $K = 5$ was sequentially evaluated with Structure Harvester v0.6.94 using 5 replicates (Earl and vonHoldt 2011) for the complete and taxon-specific datasets being duplicated for the alternative stacks and dDocent SNP datasets. Then, the Evanno method (Evanno et al. 2005) was used for detecting the number of K groups that best fit these datasets.

Mitochondrial sequences obtained were deposited in GenBank/EMBL/DDBJ and are available with accession numbers MZ605952 - MZ606134. VCF files containing filtered SNP data and MEGA and BEATS input alignment COX1 files were deposited in Dryad (https://doi.org/10.5061/dryad.b5mkkwhfn).

Results

Analysis of genetic distances based on COX 1 mitochondrial sequences

Analysis of the available mitochondrial COX1 sequences belonging to the *Patella* genus, from this work and mined from GenBank (accessed 03/11/2019), enabled a comparison of the pairwise distance values at different taxonomical levels and among consistent mitochondrial clades within the *P. candei* complex (Sá-Pinto et al. 2005; Sa-Pinto et al. 2008). The mean net distance between pairs of *Patella* species, excluding those of the *P. candei* complex and *P.*

lugubris, was 0.112 (SD = 0.015). A lower estimate, 0.041 (SD = 0.024), was obtained evaluating the inter-clade distances within the *P. candei* complex, including *P. lugubris*. The distance between clade A (containing *P. c. gomesii* and *P. c. ordinaria*) and clade B (including *P. c. candei*, *P. c. crenata* and *P. lugubris*) was 0.05. The mean estimated intra-specific (species or clade) value was similar for all groups (0.007; SD = 0.005). A lower distance estimation (0.0125; SD = 0.004) was obtained from a comparison between clades A.1 (Azores, *P. c. gomesii*) and A.2 (Madeira, *P. c. ordinaria*), in the range of intraspecific range of genetic distances as estimated for the *Patella* genus (Fig. 1; Table 2).

Table 2 Estimation of distances (average p-distance) among clades (below diagonal, blue), within clades (diagonal, green), and SD values (above diagonal, orange)

Clade	A.1	A.2	B.1	B.2
P. candei A.1	*0.0035*	0.0040	0.0090	0.0090
P. candei A.2	0.0125	*0.0047*	0.0090	0.0090
P. candei B.1	0.0564	0.0566	*0.0046*	0.005
P. lugubris B.2	0.0526	0.0528	0.0189	*0.0011*

Phylogenetic relationships based on COX 1 mitochondrial sequences

The complete set of available COX1 sequences for the *P. candei/P. lugubris*-related taxa were consistently grouped into two major clades (bootstrapping value = 99), labelled A and B. Clade A was subdivided into two closer subclades, A.1 and A.2. Clade A.1 included all the sequences from
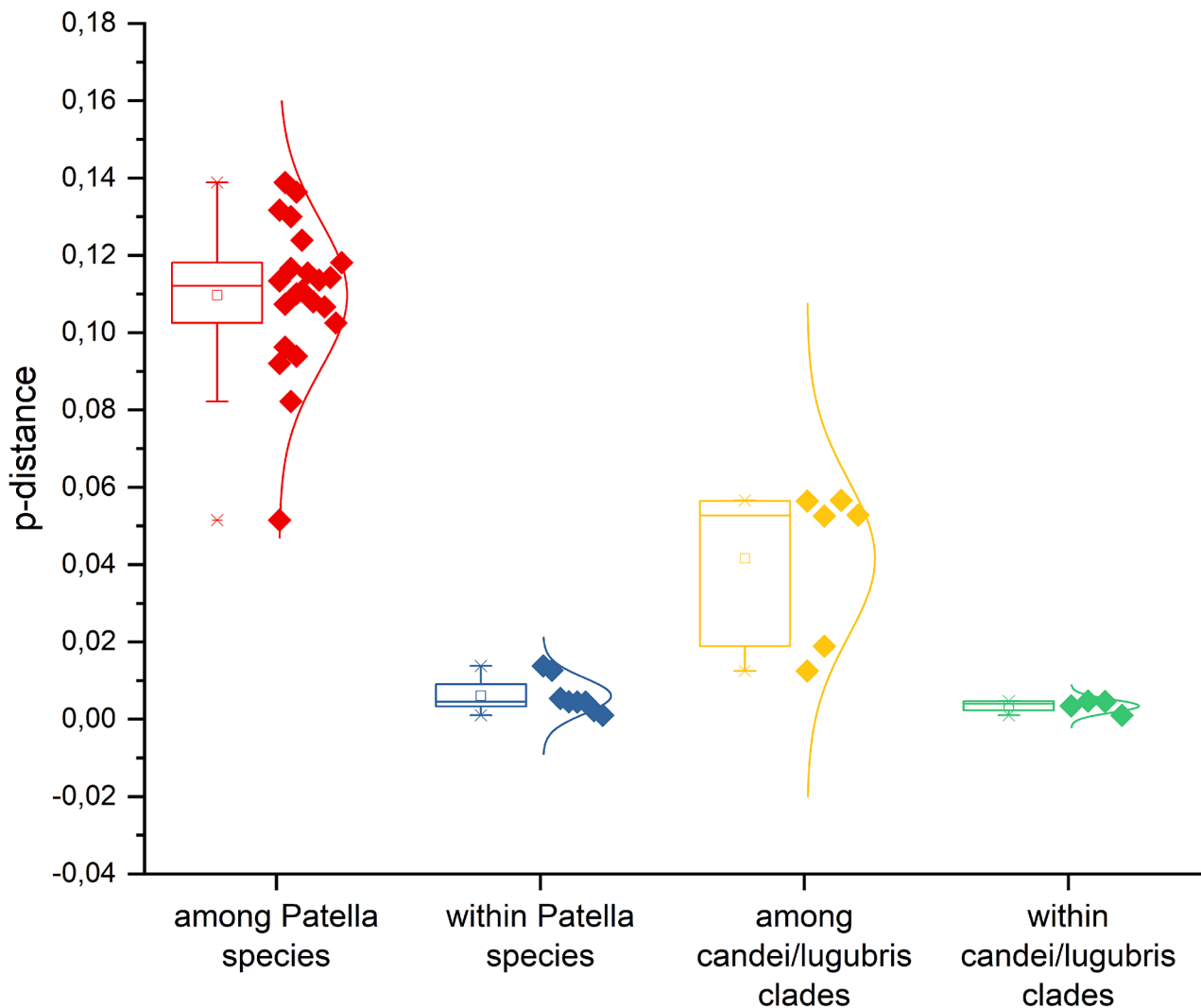

Fig. 1 Estimation of genetic distances (*p*-distances) based on mitochondrial COX1 gene sequences at different taxonomic levels and detected clustering patterns within the *Patella* genus and the *P. candei* complex, including *P. lugubris*

individuals sampled at the Azores, grouped together with data mined from GenBank and related to individuals of the same geographic origin. The sister clade included individuals from the Madeira archipelago: Madeira, Porto Santo and Desertas (Fig. 2). Clade B.1 included specimens of *P. candei candei* sampled at the Selvagens and Canary Islands. No subclade organization or subdivision related to taxa or sampling site was supported within this broad clade. The close sister clade B.2 included all *P. lugubris* samples analysed from the distant Cabo Verde archipelago.

ddRAD variants

A total of Illumina 17 312 424 paired-end reads was generated from 70 individuals. The use of both recent and carefully preserved samples, with adequate DNA integrity and concentration (mean > 800 ng), in combination with long-time archived and swab samples, resulted in different sequencing output patterns. Thus, both initial mean reads number and depth coverage per specimen were lower for *P. c. candei* from Fuerteventura (Pcan FV) (162369.6 and 9.48x mean values) and *P. c. candei* from Selvagens (Pcan SV) (177892.1 and 11.1x) than those achieved for *P. c. crenata* (Pcre FV) (346026.6 and 15.2x) and *P. c. crenata* from Tenerife (Pcre TF) (226757.4 and 13.5x).

dDocent assembled 19 366 sequences into 8494 contigs, a number close to the theoretically expected ddRAD fragments. After initial SNP calling and filtering, 3092 SNPs from 54 individuals were retained. To take into consideration the effect of alternative data on achieved inferences, a
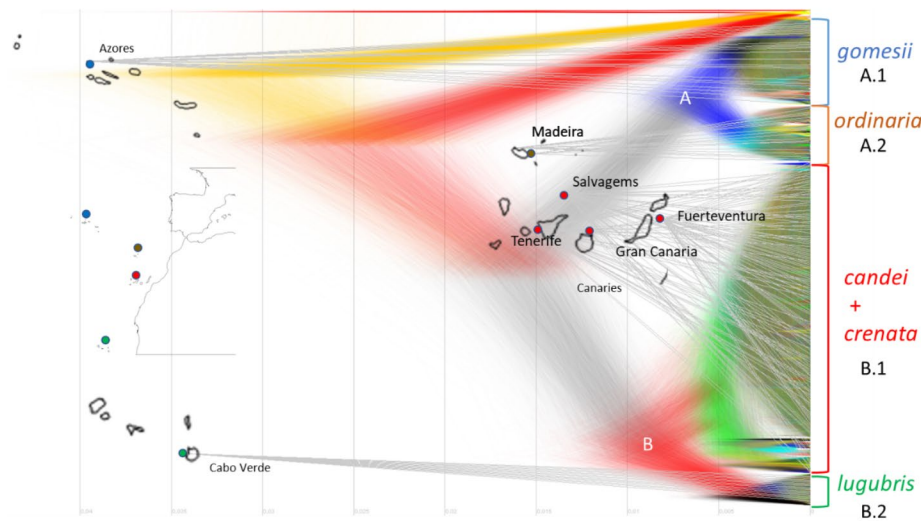

Fig. 2 Density tree based on all available mitochondrial COX 1 sequences, from this study and mined from GenBank, reflecting the relationships among taxa included in the *Patella candei* complex, including *P. c. gomesii*, *P. c. ordinaria*, *P. c. candei* and *P. c. crenata* from Macaronesia, and *P. lugubris* from Cabo Verde. The specimens sequenced in this study are traced to the sampling site (not drawn to scale) by grey lines. Grey vertical lines represent 0.005 p-distance intervals. The *P. caurelea* sequences located in the upper position (no label) were used as an outgroup. An embedded map displays the four sampling locations in the northeastern Atlantic

Table 3 Number of alleles, loci and nucleotide diversity (nuc. div.) estimates for the two SNP data sets obtained using Stacks and dDocent pipelines in *P. candei candei* (Pcan) and *P. candei crenata* (Pcre) sampled on Fuerteventura (FV), Tenerife (TF) and the Selvagens Islands (SV)

Stacks data set	Populations			
	Pcan FV	Pcan SV	Pcre FV	Pcre TF
N	18	24	30	20
Loci < 5, 0% missing data	1300	1074	1479	1280
Number of substitutions	616	395	615	492
Nuc. Div.	0.116456	0.094196	0.079348	0.088351
SD	0.058577	0.046752	0.038982	0.044231
dDocent data set				
N	18	38	30	22
Loci < 5, 0% missing data	3481	3199	3597	3638
Number of substitutions	2103	2174	1961	1825
Nuc. Div.	0.175335	0.177643	0.141664	0.138636
SD	0.08787	0.086176	0.069258	0.068699

highly stringent, with minimum missing data (<5%), Stacks data set included 46 specimens and 2136 SNPs.

Genetic diversity and differentiation

The estimated nucleotide diversity values for *candei* were higher than for the *crenata* taxon, with larger values estimated for the small and threatened Fuerteventura population (Table 3).

The results of the exact G test in both datasets, for genotype differentiation among populations, suggested a differentiation for all pairwise comparisons ($P < 0.0001$), except in the case of *crenata* populations.

The population pairwise F_{ST} from both datasets suggested a strong differentiation between the samples classified as *P. candei candei* (Pcan) and *P. candei crenata* (Pcre) taxa ($F_{ST} = 0.5, 0.7; P = 0$). Within *P. candei candei* the populations from the islands of Fuerteventura and Selvagens were significantly differentiated ($F_{ST} = 0.05, 0.10; P = 0$).

Also, Tenerife and Fuerteventura populations of *P. candei crenata* showed a low but significant level of differentiation. The values estimated from both datasets were correlated, but those estimated based on Stacks pipeline SNPs data were generally higher (Table 4).

The phylogenetic relationships among individual genotypes suggested the monophyletic and divergent clustering into two main taxa-specific clades for *P. c. candei* and *P. c. crenata*. These clades were well supported and recovered under the Bayesian criteria (Fig. 3. A, B) and distance clustering (Fig. 3. C, D), as well as when alternative Stacks (Fig. 3. A, C) and dDocent (Fig. 3. B, D) data sets were considered. A high level of intraspecific cluster differentiation was observed in *P. c. candei* respect to the *P. c. crenata* (Fig. 3. A, B) density trees. Population-specific clades were only well supported (>93%) for the Fuerteventura and Selvagens populations of *P. c. candei* in distance trees based in the dDocent data set (Fig. 3. D) whereas that a low support (<50%) was estimated in Stacks data set based trees.

Table 4 F_{ST} pairwise values among populations of *P. candei candei* (Pcan) and *P. candei crenata* (Pcre) sampled on Fuerteventura (FV), Tenerife (TF) and the Selvagens Islands (SV). Estimates based on the SNP dataset derived from the Stacks pipeline are below the diagonal, whereas those from the dDocent pipeline are above the diagonal

Populations	Pcan FV	Pcan SV	Pcre FV	Pcre TF
Pcan FV		0.05044	0.52510	0.52694
Pcan SV	0.10120		0.52280	0.52151
Pcre FV	0.73178	0.74298		0.02346
Pcre TF	0.73010	0.74372	0.01609	
All values are significant ($P < 0.05$)				

Fig. 3 Trees estimated for limpets belonging to the Fuerteventura and Selvagens populations of *Patella candei candei* (candei) and to the Fuerteventura and Tenerife populations of *Patella candei crenata* (crenata). Both density trees based on Bayesian criteria (A and B) as dendrograms based in bitwise distance (C and D) were estimated from Stacks (A and C) and dDocent SNPs (B and D) data sets. Vertical lines in density trees represent a 0.01 distance value
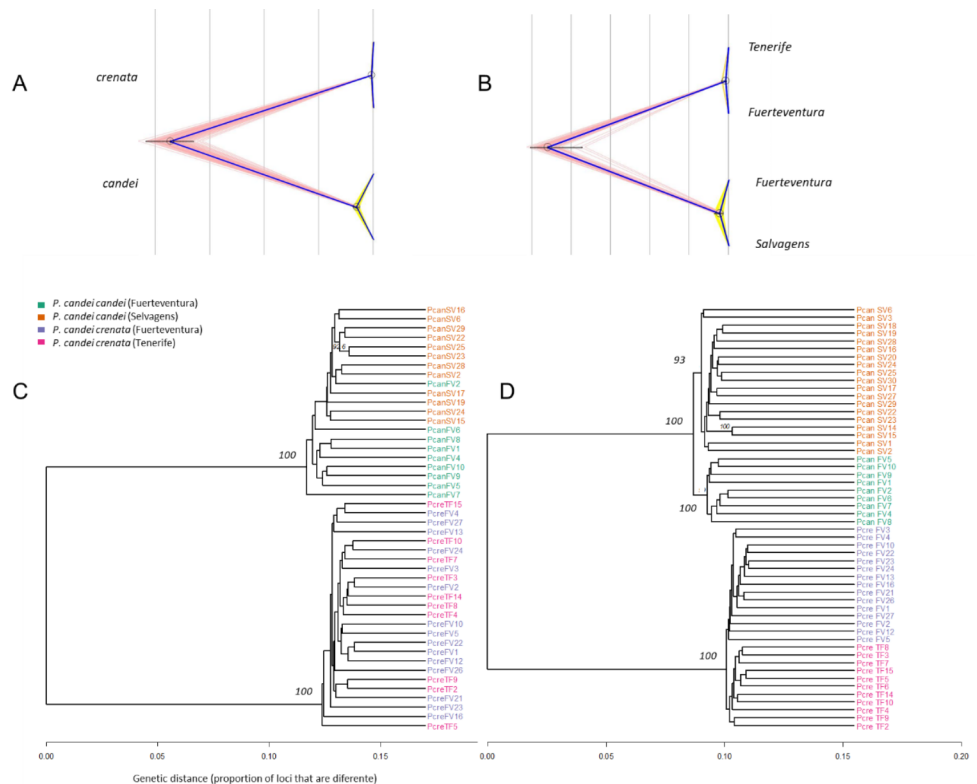
A similar pattern is observed for population-specific clades in *P. c. crenata*. The population specific clustering observed

Fig. 4 PCA plot of the first two component axes calculated from a Stacks pipeline dataset (green colour; 1384 LD filtered SNPs) and dDocent data set (blue colour; 2756 LD filtered SNPs) of SNPs from insular populations of *P. candei candei* (Fuerteventura and Selvagens) and *P. candei crenata* (Fuerteventura and Tenerife)
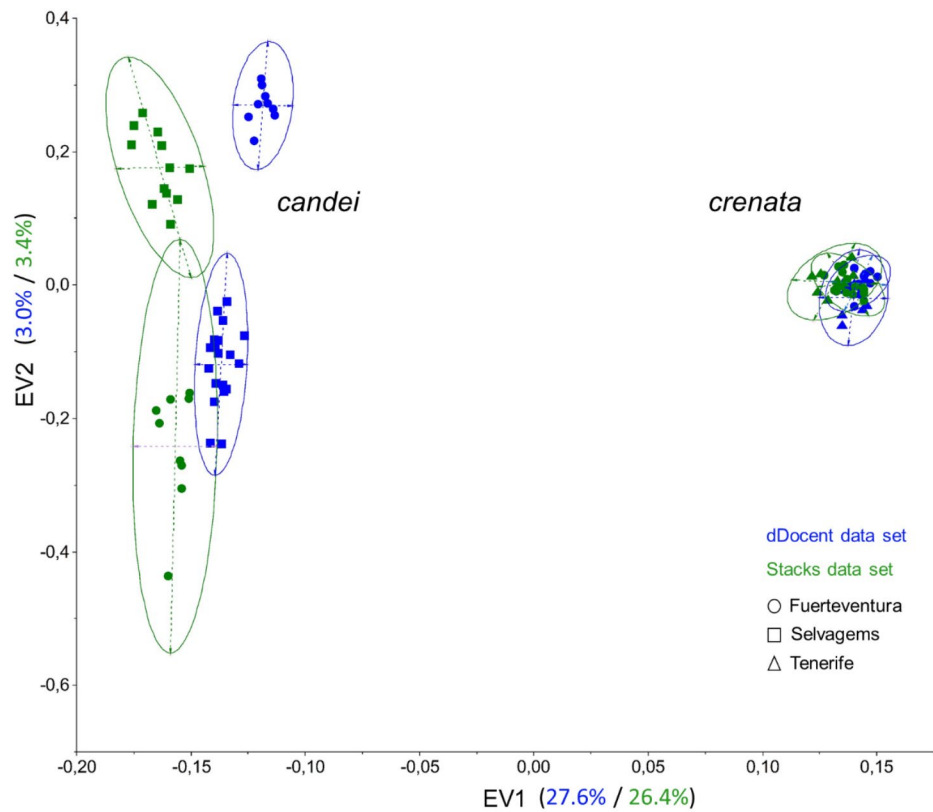

Fig. 5 Summary plot of estimates of ancestry (*Q*). Plots were divided into population specific segments (Fuerteventura, Selvagens and Tenerife). Each individual is represented by a single vertical line broken into K coloured segments, with lengths proportional to each of the K inferred clusters. The plots for the complete combined data set (A and B) (*P. candei candei* and *P. candei crenata*) assumed K = 3 in A (Stacks data set) and B (dDocent data set). In addition, plots were obtained separately for each taxa assuming K = 2. The *P. c. candei* results were separately plotted (left) from Stacks (C) and dDocent (D) data sets. Similarly, the specific *P. c. crenata* estimates were plotted (right) from Stacks (C) and dDocent (D) data sets
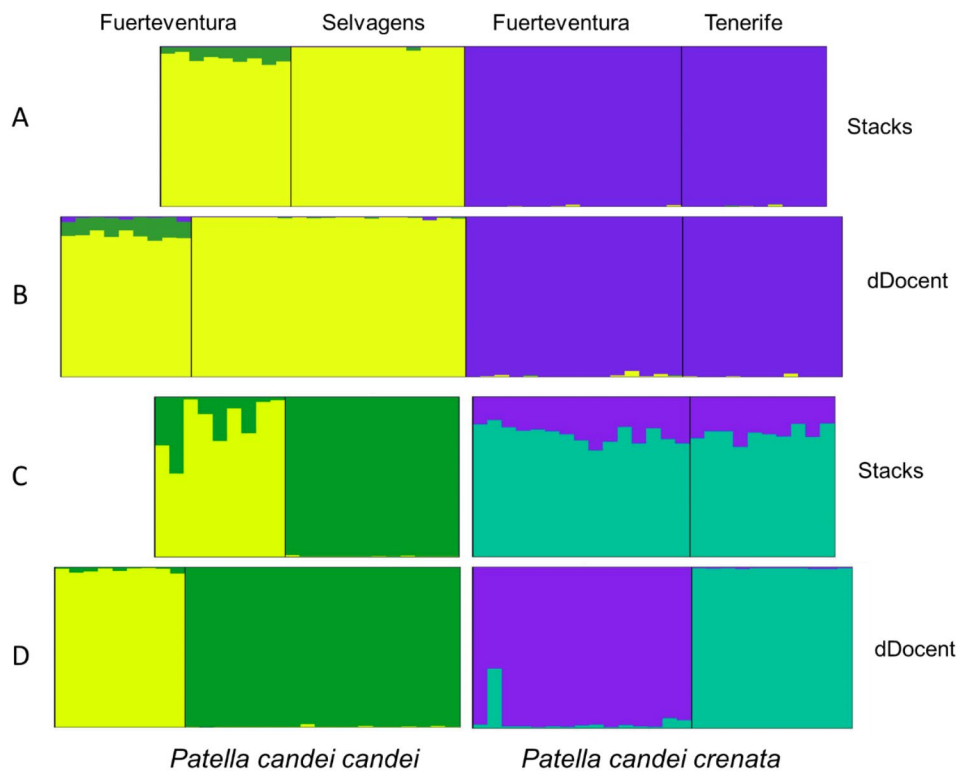
in trees derived from dDocent data set, however, showed a low support (Fig. 3. D), whereas no clustering was observed from the Stacks data set for *P. c. crenata*. (Fig. 3. C).

The PCA revealed a clear divergence between the *candei* and the *crenata* populations, along the first axis and explained 26.4–27.6% of the total variance. Furthermore, two different patterns emerged when variation within these taxa was considered. The *crenata* populations (Fuerteventura and Tenerife) showed a homogeneous and near compact cluster. Conversely, the *candei* populations, from Fuerteventura and Selvagens, displayed an insular-specific divergent clustering. The second axis principally separated these *candei* samples with 3.0–3.4% of the total variance explained. Similar patterns were obtained from the alternative Stacks and dDocent LD filtered data sets (Fig. 4).

The Structure model-based clustering method for inferring population structure suggested the presence of three differentiated populations based on diverse analysis of both the Stacks and dDocent data sets.

For the complete data set, involving all populations, a $K=2$ was selected, where individuals from *P. c. candei* and *P. c. crenata* taxa were separately clustered. However, the test of a $K=3$, with a slightly lower likelihood value in Structure Harvest, suggested a population differentiation within the main *P. c. candei* cluster between Selvagens and Fuerteventura samples. The individual membership coefficient for Fuerteventura individuals shows a different pattern respect to the Selvagens individuals including a significant memberships coefficient for an alternative cluster. In contrast, individuals from Tenerife and Fuerteventura populations of *P. c. crenata* showed identical and homogeneous ancestry estimates (Fig. 5. A, B).

In a subsequent analysis, the *P. c. candei* and *P. c. crenata* taxa were discretely analysed. The Structure Harvest results suggested a K value of 2 and the analysis revealed a partition of *P. c. candei* into two clusters, for both Stacks and dDoc SNPs data sets, in agreement with the Selvagens and Fuerteventura geographic origin of individuals. The analysis ($K=2$) of the *P. c. crenata* based on the Stacks SNPs data set reflected a homogeneous membership coefficient for the sampled individuals. In contrast, when ancestry was evaluated with the dDoc data set, a different proportion of membership of each *P. c. crenata* samples from Tenerife and Fuerteventura was obtained (Fig. 5. C, D).

Discussion

Analysis of the ddRAD-based SNPs data provided the resolution required to infer taxonomic and population relationships at the level where mitochondrial data showed critical limitations. Moreover, the data achieved from archived samples and non-invasive mucosal swab sampling suggest the efficacy of this protocol under non-ideal conditions. The subspecies *P. c. candei* and *P. c. crenata* were genetically differentiated, for the first time, based on the diverse SNPs datasets and alternative analysis. At the intraspecific level, differentiation has also been observed between the Selvagens and Fuerteventura populations of *P. c. candei*, which represents a significant contribution to the planning of conservation strategies for the endangered population.

Disjunct ranges for inter- and intraspecific genetic distances are essential for drawing species boundaries under the COX1 DNA barcoding concept, allowing for the molecular identification of known species and the discovery of previously disregarded taxa (Hebert et al. 2003). An evaluation of estimated genetic distances is particularly pertinent in the absence of a taxonomic consensus involving threatened populations. The mean interspecific COX1-based distance among *Patella* species, excluding *candei* and *lugubris* clades, was 11.2%. The mean intraspecific or intra-clade distance among *Patella* species (0.7%) agreed with expected values of >1%, as reported for Canadian marine molluscs (0.49%) (Zou et al. 2011; Layton et al. 2014). The distance among the taxa of the *candei* complex and *P. lugubris* is at an intermediate level, with a mean value of 4.1%. However, there was no divergence between the putative taxa *P. candei crenata* and *P. candei candei*.

Mitochondrial-based phylogenetic studies of Macaronesian species of the genus *Patella* revealed highly consistent patterns (Koufopanou et al. 1999; Sá-Pinto et al. 2005; Sa-Pinto et al. 2008; Carreira et al. 2017). In congruence, our analysis grouped *P. candei* and *P. lugubris* as the closest taxa within the clade I (Sá-Pinto et al. 2005), denominated the *P. candei–P. lugubris* complex (Sa-Pinto et al. 2008). Within clade I, clade MAZ, following Sá-Pinto et al. (2008) nomenclature, was defined here as clade A.

The division of clade A into the subclades A.1 and A.2, with individuals from the Azores and Madeira archipelagos, respectively, reflects the taxonomic delineation of *P. c. gomesii* (Azores) and *P. c. ordinaria* (Madeira), elevated to the species range (Faria et al. 2017). Similarly, clade B (clade CANCBV following Sá-Pinto et al. (2008) nomenclature is subdivided into clade B.1, containing *P. candei candei* individuals from the Selvagens and Canary archipelagos, and clade B.2, including *P. lugubris*.

Clade B1 includes two putative and partially sympatric subspecies *P. c. candei* and *P. c. crenata* (Christiaens 1973). No genetic discontinuity and monophyletic clustering were observed within this clade, here and in other studies (Carreira et al. 2017) from mitochondrial data. In contrast, their species status has been assumed in a STR-based study although, unfortunately, no STR data have been obtained from *P. candei candei* taxa (Faria et al. 2017). However,

morphological differentiation has been described between *candei* and *crenata* shells based on the position and shape of the apex and radiation ridges and, notably, a smooth edge without ribs in *candei* (Christiaens 1973; Hernández-Dorta 1992; Titselaar 1998; González-Lorenzo et al. 2015). This description resembled the eroded phenotype described in shells of *P. c. candei* from Selvagens (Carreira et al. 2017). Despite the non-cryptic appearance, the absence of previous mitochondrial or nuclear data supporting *P. c. candei* and *P. c. crenata* clades resulted in unresolved phylogenetic relationships and in the absence of a robust taxonomic frame required for the conservation and management of the putative *candei* and *crenata* taxa.

The subspecies *Patella c. candei* and *P. c. crenata* individuals were divergent and monophyletically clustered in well-supported clades (100%). The F_{ST} pairwise values among populations from different taxa ($F_{ST} > 0.5$) suggested an absence of gene flow, with high values when comparing those populations sympatrically occurring in the Fuerteventura rockeries. Finally, most of the total variance estimated from the SNP data was related to differences in allelic frequencies between *P. c. candei* and *P. c. crenata.*

Diverse events have been suggested as a cause of incongruence between nuclear and mitochondrial data. Thus, introgressive hybridization events introduce female mtDNA lineages of a species in the genomic composition of a second species. Such a case of putative hybridization between *P. depressa* and *P. vulgata* has been reported (Koufopanou et al. 1999). However, in the *P. candei* complex the two main mitochondrial lineages, clades A and B, diverged at intermediate values between intra- and interspecific rates, suggesting a recent speciation process within the complex and an incomplete mitochondrial lineage sorting affecting the closely related *candei* and *crenata* taxa.

The SNP data suggested the delimitation of *candei* and *crenata* sister taxa as divergent evolutionary units. A definitive conclusion on their taxonomic status, however, requires a phylogenetic analysis of homologous data involving all taxa and populations within the *candei* complex and related species. In any taxonomic context, the observed delimitation and high divergence between these morphologically similar taxa must be considered for the adoption of problematic specific management programmes for a resource inhabiting the same rockeries on the Fuerteventura coast (Titselaar 1998; González-Lorenzo et al. 2015).

The resolution provided by the SNP data enabled the observation of a different population structure pattern within the *candei* and *crenata* taxa. The populations of *crenata* sampled from Tenerife and Fuerteventura belong to a single gene pool. Only when using the wider SNPs dataset was a slight population differentiation observed.

In the case of *candei*, the isolation between the only two worldwide populations of *P. c. candei* enables us to delineate intraspecific units for conservation. The significant genetic differentiation between the Selvagens and Fuerteventura samples reflects an absence of gene flow congruent with the absence of natural recruitment in Fuerteventura from a putative source population is Selvagens (González-Lorenzo et al. 2015). The pattern observed for *P. c. candei* resembles that detected for the endangered limpet *P. ferruginea*. Throughout the Mediterranean, non-differentiated mtDNA lineages were detected. In contrast, northwestern local populations were differentiated by ISSR nuclear markers analysis, with implications for the planning of conservation initiatives (Casu et al. 2006, 2011).

As stated by the Convention of Biological Diversity, intraspecific diversity is a fundamental part of overall biological diversity (CBD 1992) since the conservation of distinct local populations is vital for maximizing evolutionary potential and minimizing the probability of extinction (Allendorf et al. 2013). The restoration plan for the "majorera limpet" suggests the possibility of restoring the threatened population of Fuerteventura with the Selvagens population as a source. However, this possibility would only be considered after the analysis and evaluation of resolutive genetic data. In consequence, this study responded to the Plan proposal for an action involving the genetic characterization of the "majorera limpet" and the analysis of interpopulation gene flow (Gobierno de Canarias 2015).

Source populations for translocations should have genetic similarity to the recipient population and high genetic diversity (Allendorf et al. 2013). The SNP data acquired from a reduced representation of the genome of *P. c. candei* suggested a genetic differentiation between samples from the Selvagens and Fuerteventura co-specific populations. The introduction of individuals from Selvagens carrying differentiated genotypes into the threatened Fuerteventura populations may have hardly predictable consequences.

The genetic diversity estimated (N alleles $= 18$) for the near-to-collapse population in Fuerteventura population is unexpectedly slightly higher or similar to that estimated (N alleles $= 24/38$) for the healthy unexploited Selvagens population. Nevertheless, erosion of the current genetic diversity in the Fuerteventura population would be expected after supplemental translocations under conditions of climate change and habitat degradation.

The genetic diversity values do not support the role suggested for the Selvagens population of *P. c. candei* as the ancestral population of *P. c. candei*, or the ancestral centre of species radiation (Weber and Hawkins 2002; Sa-Pinto et al. 2008) if we consider its relationship with *Patella lugubris* based on mitochondrial phylogenies (Sá-Pinto et al. 2005; Sa-Pinto et al. 2008). In contrast, both Selvagens and

Fuerteventura likely constitute relic populations, healthy and threatened, respectively. Selvagens was likely peripherally differentiated, and the Fuerteventura population partially retains the genetic diversity of a pre-human ancestral distribution that existed before excessive exploitation (Weber and Hawkinks 2002). However, González-Lorenzo (2015) rejected the relictual hypothesis based on the archaemalacological register derived from deposits of aboriginal origin where *P. candei* was abundant only in Fuerteventura and Lanzarote. The relictual hypothesis will be congruent with the archaeological register only if natural factors are responsible for the current distribution. Precise molecular dating, which is currently unavailable, will aid in clarifying this theory.

Without any inference about the evolutionary forces shaping the detected genetic variability in *P. c. candei*, the observed patterns are valuable and unique evidence to support urgent decisions on the conservation management and strategy for the threatened Fuerteventura population. We conclude that the "majorera limpet" is a significant evolutionary unit requiring closed and population-specific management focused on the preservation of exceptional genetic diversity with which to face future environmental challenges.

Data Availability Data obtained in the study was deposited in GenBank.

Code Availability Not applicable.

Declarations

Conflicts of interest/Competing interests Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All authors approve the manuscript and give their consent for submission and publication.

References

Allendorf F, Luikart G, Aitken SN (2013) Conservation and the Genetics of Populations, 2nd edn. Wiley-Blackwell

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography - the Mitochondrial-DNA bridge between population-genetics and systematics. Annu Rev Ecol Syst 18:489–522

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376. https://doi.org/10.1371/journal.pone.0003376

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

Bouckaert RR, Heled J (2014) DensiTree 2: Seeing Trees Through the Forest. 012401. https://doi.org/10.1101/012401

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J, Jones G, Kuhnert D, De Maio N, Matschiner M, Mendes FK, Muller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 15:e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol 29:1917–1932. https://doi.org/10.1093/molbev/mss086

Carleton KL (2011) Quantification of transcript levels with quantitative RT-PCR. Methods Mol Biol 772:279–295. https://doi.org/10.1007/978-1-61779-228-1_17

Carreira GP, Shaw PW, Gonçalves JM, McKeown NJ (2017) Congruent Molecular and Morphological Diversity of Macaronesian Limpets: Insights into eco-evolutionary Forces and Tools for Conservation. Front Mar Sci. 4https://doi.org/10.3389/fmars.2017.00075

Casu M, Casu D, Lai T, Cossu P, Curini-Galletti M (2006) Inter-simple sequence repeat markers reveal strong genetic differentiation among populations of the endangered mollusc *Patella ferruginea* (Gastropoda: Patellidae) from two Sardinian marine protected areas. Mar Biol 149:1163–1174. https://doi.org/10.1007/s00227-006-0255-y

Casu M, Rivera-Ingraham GA, Cossu P, Lai T, Sanna D, Dedola GL, Sussarellu R, Sella G, Cristo B, Curini-Galletti M, Garcia-Gomez JC, Espinosa F (2011) Patterns of spatial genetic structuring in the endangered limpet *Patella ferruginea*: implications for the conservation of a Mediterranean endemic. Genetica 139:1293–1308. https://doi.org/10.1007/s10709-012-9631-3

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Christiaens J (1973) Révision du genre *Patella* (Mollusca, Gastropoda). Bulletin du Museum National d´Histoire Naturelle, 3, 1305–1392

Coates DJ, Byrne M, Moritz C (2018) Genetic Diversity and Conservation Units: Dealing With the Species-Population Continuum in the Age of Genomics. Frontiers in Ecology and Evolution, 6. https://doi.org/10.3389/fevo.2018.00165

Diaz-Arce N, Rodriguez-Ezpeleta N (2019) Front Genet 10:533. https://doi.org/10.3389/fgene.2019.00533. Selecting RAD-Seq Data Analysis Parameters for Population Genetics: The More the Better?

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973. https://doi.org/10.1093/molbev/mss075

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Faria J, Martins GM, Pita A, Ribeiro PA, Hawkins SJ, Presa P, Neto AI (2017) Disentangling the genetic and morphological structure of *Patella candei* complex in Macaronesia (NE Atlantic). Ecol Evol 7:6125–6140. https://doi.org/10.1002/ece3.3121

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Garrison E, Gabor M (2012) Haplotype-based variant detection from short-read sequencing. In:.

Gobierno de Canarias (2015) Decreto 161/2015, de 3 de julio, por el que se aprueba el Plan de Recuperación de la Lapa Majorera (*Patella candei*). In: Gd C (ed) 136. Boletín Oficial de Canarias

González-Lorenzo G, Mesa Hernández E, Pérez-Dionis G, Brito Hernández A, Galván Santos B, Barquín Diez J (2015) Ineffective conservation threatens *Patella candei*, an endangered limpet endemic to the Macaronesian islands. Biol Conserv 192:428–435. https://doi.org/10.1016/j.biocon.2015.11.005

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270:313–321. https://doi.org/10.1098/rspb.2002.2218

Hernández-Dorta FJ (1992) Revisión del género *Patella* Linné, 1758 (Mollusca, Gastropoda) en las Islas Canarias. Vieraea 21:109–135

Kamvar ZN, Brooks JC, Grunwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:208. https://doi.org/10.3389/fgene.2015.00208

Kess T, Gross J, Harper F, Boulding EG (2015) Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. J Molluscan Stud 82:104–109. https://doi.org/10.1093/mollus/eyv042

Koufopanou V, Reid DG, Ridgway SA, Thomas RH (1999) A molecular phylogeny of the patellid limpets (Gastropoda: Patellidae) and its implications for the origins of their antitropical distribution. Mol Phylogenet Evol 11:138–156. https://doi.org/10.1006/mpev.1998.0557

Layton KK, Martel AL, Hebert PD (2014) Patterns of DNA barcode variation in Canadian marine molluscs. PLoS ONE 9:e95003. https://doi.org/10.1371/journal.pone.0095003

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Pinero D, Emerson BC (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol Ecol Resour 15:28–41. https://doi.org/10.1111/1755-0998.12291

Mora-Marquez F, Garcia-Olivares V, Emerson BC, Lopez de Heredia U (2017) ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. Mol Ecol Resour 17:230–246. https://doi.org/10.1111/1755-0998.12550

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. AnnRevEcolSyst 18:269–292

Paris JR, Stevens JR, Catchen JM, Johnston S (2017) Lost in parameter space: a road map for stacks. Methods Ecol Evol 8:1360–1373. https://doi.org/10.1111/2041-210x.12775

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. 155, 945–959

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. https://doi.org/10.7717/peerj.431

R-Core-Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Sa-Pinto A, Branco M, Sayanda D, Alexandrino P (2008) Patterns of colonization, evolution and gene flow in species of the genus Patella in the Macaronesian Islands. Mol Ecol 17:519–532. https://doi.org/10.1111/j.1365-294X.2007.03563.x

Sá-Pinto A, Branco M, Harris DJ, Alexandrino P (2005) Phylogeny and phylogeography of the genus *Patella* based on mitochondrial DNA sequence data. J Exp Mar Biol Ecol 325:95–110

Supple MA, Shapiro B (2018) Conservation of biodiversity in the genomics era. Genome Biol 19:131. https://doi.org/10.1186/s13059-018-1520-3

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Titselaar F (1998) A revision of the recent European Patellidae (Mollusca: Gastropoda): 1. The Patellidae of the Azores, Madeira, the Selvagens and the Canary Islands. Vita Mar 45:21–62

Weber LI, Hawkinks SJ (2002) Evolution of the limpet Patella candei d'Orbigny (Mollusca, Patellidae) in Atlantic archipelagos: human intervention and natural processes. Biol J Linn Soc 77:341–353. https://doi.org/10.1046/j.1095-8312.2002.00102.x %J Biological Journal of the Linnean Society

Werner GD, Gemmell P, Grosser S, Hamer R, Shimeld SM (2013) Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. Mar Biotechnol (NY) 15:230–243. https://doi.org/10.1007/s10126-012-9481-0

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zou S, Li Q, Kong L, Yu H, Zheng X (2011) Comparing the usefulness of distance, monophyly and character-based DNA barcoding methods in species identification: a case study of neogastropoda. PLoS ONE 6:e26619. https://doi.org/10.1371/journal.pone.0026619