Postprint del artículo publicado en: Journal of Heredity 105(5): 597-610 (2014)

Genetic Diversity and Population History of the Endangered Killifish *Aphanius baeticus*

ELENA G. GONZALEZ[1], CARLOS PEDRAZA-LARA[1,2,*] AND IGNACIO DOADRIO[1]

From the Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, MNCN-CSIC, José Gutiérrez Abascal, 2, Madrid, Spain, 28006 (Gonzalez, Pedraza-Lara and Doadrio); and the Instituto de Biología. Universidad Nacional Autónoma de México; Tercer circuito s/n, Ciudad Universitaria, Coyoacán, México D.F, 04510 (Pedraza-Lara).

Address correspondence to C. Pedraza-Lara at the address above, or e-mail: carlospedrazal@yahoo.es

Abstract

The secondary fresh-water fish fauna of the western-Iberian Peninsula basin is primary restricted to local coastal streams, salt water creeks, and man-made salt evaporation ponds, which are susceptible to periodical flood and drought events. Despite their uniqueness in ecological adaptation to high salt-water tolerance, very little is known about their population dynamics and evolutionary history. The killifish, *Aphanius baeticus* (Cyprinodontidae) is an endemic species restricted to river basins on Spain's southern Atlantic coastline and is considered as "Endangered". In this study, the genetic structure, diversity and historical demography of *A. baeticus* were analyzed using mitochondrial and nuclear markers across its entire distribution range. The phylogenetic and networking reconstruction revealed subtle phylogeographic structuring. A scattered expansion at the beginning of the interglacial periods, coupled with posterior events of extinction and colonization caused by periodical cycles of flooding, could explain the absence of well-defined phylogenetic relationships among populations. Moreover, very low genetic diversity values and a weak but significant population differentiation were detected by microsatellites. A possible explanation could be that dispersals allowed by periodic floods connecting river drainages may have promoted a wide genetic exchange among populations and could have contributed to the current genetic relatedness of these populations.

Keywords: Cyprinodotidae, mtDNA, microsatellite, secondary water fishes, Pleistocene glaciations, conservation

Introduction

Understanding how past events and historical processes have contributed to the origin and maintenance of global biodiversity plays a central role in evolutionary biology. The western-Iberia region offers an adequate baseline scenario for studying how different factors such as historical processes (e.g., climate oscillations) or life-history features (e.g., strong seasonal adaptations or short generation times), together with recent human impact, have shaped the current genetic structure of ichthyofauna in this region (Gante, 2011; Gante, et al., 2009; Mesquita, et al., 2005).

Eustatic sea level changes (i.e., marine shore regressions/transgressions) that occurred before and after Pleistocene glaciation events have greatly influenced the population structure and dynamics of onshore rivers and coastal brackish areas within the western-Iberian basin (Mesquita, et al., 2001; Mesquita, et al., 2005; Perdices, et al., 2001). For instance, during the Last Glacial Maximum the sea level dropped 120 m under current sea levels, moving the coastline 15 km from what is the present Strait of Gibraltar (South of the Iberian Peninsula, (Luque, et al., 1999)).

Other postglacial processes occurred within the past c.a. 10,000 years, such as the alteration of inter-seasonal (rainy and dry) climate, characteristic of the Mediterranean Region, and other environmental perturbations, which were accompanied by many local ichthyofauna shifts in that region. This is the case of high floods produced by tsunamis, which have occurred almost periodically over the last 7,000 years, (Lario, et al., 2011). This is true in the case of the last large tsunami (induced by the earthquake in Lisbon in 1755), which caused large washovers in Doñana National Park, Boca do Río and Guadalete estuaries and the Trafalgar Cape. Each of these washovers was confirmed by the existence of marine and boulder deposits that were reported in the area (Ruiz, et al., 2008; Whelan and Kelletat, 2005). For small species with low dispersal capacity, such as the killifish of the genus *Aphanius*, events like this may have severely influenced its

demography. However, being a saltwater-tolerant species, these events may be considered as dispersal events rather than events of extinction.

Aside from historical events, population genetics of secondary water fishes have been more recently complicated by the effect of human pressure, producing in some cases genetic bottlenecks and local extinctions (Perdices, et al., 2001; Schönhuth, et al., 2003). Many cyprinodontid fishes of the Iberian Peninsula are highly endangered due to the degradation of their habitats (mainly because of agricultural or urban disturbances) and the introduction of exotic species (Fernandez-Delgado, et al., 1988; Schönhuth, et al., 2003). In addition to these human-mediated threats, species-specific life history traits (e.g., low effective population size) could be factors that contribute to population declines (Schönhuth, et al., 2003).

The Iberian Peninsula is the western limit of the killifish distribution in Europe, represented by the species *Aphanius baeticus*. Previous works have described the genetic uniqueness of this species, which is clearly separated in genetic and morphological terms from its closer relatives, the Mediterranean *A. iberus* (from southern Spain) (Doadrio, et al., 2002; Doadrio, et al., 1996; Fernandez-Delgado, et al., 1988; Perdices, et al., 2001) and *A. saourensis* (from northern Algeria) (Blanco, et al., 2006). Its distribution is restricted to freshwater lagoons, salt water creeks, and ancient man-made salt evaporation ponds near the coast of the Atlantic Basin in the western Iberian Peninsula (i.e., to the lower reaches of the Guadalquivir River and small streams located on the coastline) (Fig. 1), which are characterized by strong environmental fluctuations and a highly fragmented habitat (Oltra and Todolí, 2000; Schönhuth, et al., 2003). Specifically, the species has been recorded in only eight isolated localities in the Andalucian biogeographical area (Doadrio, et al., 2002; Moreno-Amichi, et al., 1999). As in other killifish within the genus, *A. baeticus* is characterized by a short life span (c.a. two years), high reproductive effort and short generation time (c.a. two per year) (Doadrio, 2011; Fernandez-Delgado, et al., 1988). This species has been assessed

as 'Endangered' under the Red Book of Freshwater Fishes (Doadrio, 2001; Doadrio, 2011) and the IUCN, 2013; http://www.iucnredlist.org/details/61235/0, due to the effect of human-mediated habitat fragmentation and more importantly the introduction of exotic species, such as *Gambusia holbrooki* and *Fundulus heteroclitus* (Doadrio, 2011; Doadrio, et al., 2002; Fernandez-Delgado, et al., 1988).

Given the critical situation of their populations, there was an urgent need to investigate the genetic variation in *A. baeticus* throughout the species distribution range, as knowing whether the population maintains certain genetic diversity levels is crucial for the successful management and conservation of the species. Moreover, genetic data could help to discern how historical processes (e.g., sustained population isolation or dispersion) or more recent events (e.g., human pressure) contributed to the evolutionary history of the species.

This work aims to: 1) investigate the associated phylogeographic and demographic patterns; 2) study the genetic structure of all the populations of *A. baeticus*; 3) determine the degree of genetic flow between populations and/or genetic groups; and 4) provide valuable information in terms of population genetic parameters to estimate future population trends.

Materials and Methods

Sampling collection and DNA extraction

A total of 288 *A. baeticus* individuals were collected during 2009 by dip-netting at 16 locations in the south of the Iberian Peninsula (Figure 1, Table 1), throughout the species distribution range. Among the samples collected, ACB, ACO and ACC correspond to individuals from a breeding program whose stock samples originated from the wild population in Doñana National Park (Schönhuth, et al., 2003). Moreover, COC (Doadrio, personal communication) and VEG (Clavero, et al., 2002; Clavero, et al., 2005) correspond to putative introduced samples from the Guadalquivir River, but previous to this work no molecular analysis has been performed to confirm this statement (Fig. 1).

Individuals were released after a small fin clip sample was taken, tissue was preserved in 90 % ethanol and subsequently stored at -20 ºC. Genomic DNA was extracted using the Qiagen DNeasy tissue kit (Qiagen) following manufacturer's conditions. Voucher samples were stored at the Museo Nacional de Ciencias Naturales of Madrid, Spain (MNCN-CSIC).

Mitochondrial (cob) DNA amplification

The complete mtDNA cytochrome b (cob) gene was amplified for 131 specimens of A. baeticus (GenBank accession numbers: KF854341 - KF854477) (Table 1). Additionally, six *cob* sequences of *A. baeticus* available from GenBank were included in the analyses (accession numbers: AF299280-AF299285). Primers used for amplification were donGlu and donThr (Machordom and Doadrio, 2001). PCR amplifications were carried out in 25 µL reactions containing: 1x PCR buffer, 0.5 µM of each primer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 U *Taq* polymerase (Biotools) and about 50 ng of template DNA. The cycling profile for PCR amplifications was 3 min at 94 ºC (1 cycle), 30 s at 94 ºC, 30 s at 48 ºC and 60 s at 72 ºC (30 cycles), followed by a final extension of 4 min at 72 ºC. Both strands were directly sequenced using the BigDye Terminator Sequencing Ready Reaction v3.1 kit

(Applied Biosystems) following manufacturer's instructions and sequenced on an ABI 3730 automated sequencer. Alignment of the nucleotide sequences was performed using the program MUSCLE (Edgar, 2004) and CLUSTAL X v1.83, and verified manually in order to maximize positional homology.

Population diversity and phylogenetic analyses

This data set was analyzed with Maximum Likelihood (ML; (Felsenstein, 1981)) using PAUP v.4.0b10 (Swofford, 2002) and Bayesian inference (BI; (Huelsenbeck, et al., 2001) using MRBAYES v.3.2.0 (Ronquist, et al., 2012). Best-fit models of nucleotide substitution were estimated using jMODELTEST v.0.1.1 (Posada, 2008) based on Akaike information criteria (AIC) (Akaike, 1973) or BIC for the ML (TPM1uf+I+G) and BI (HKY+I+G; (Hasegawa, et al., 1985)) analyses, respectively. ML was performed using heuristic searches with 10 random additions and TBR branch swapping. BI was performed with four simultaneous chains, each of 2 x 10^8 generations, sampled every 1000 generations (10% of trees were discarded as burn-in). The robustness of ML and BI trees was assessed with non-parametric bootstrap (Felsenstein, 1985) proportions (BPs; 1000 pseudoreplicates) and Bayesian posterior probabilities (BPPs), respectively. The phylogenetic tree was rooted using sequences of *A. iberus* (Genbank accession nº AF299274, AF299278, AF299286), the sister species of the genus (Doadrio, et al., 2002; Hrbek, et al., 2002; Perdices, et al., 2001). Moreover, intraspecific genetic variation (Posada and Crandall, 2001) was determined using a median-joining (MJ) network analyses applying NETWORK v.4.112 (Bandelt, et al., 1999).

Finally, *cob* mtDNA sequences were analyzed with DnaSP v5 (Librado and Rozas, 2009) to determine different descriptive statistics: the number of polymorphic sites (S), mitochondrial haplotype diversity, H_d, (Nei, 1987), and nucleotide diversity, π, (Nei, 1987) were calculated globally and for the lineages previously recognized, using DNASP 5.10 (Librado and Rozas, 2009).

Demographic analyses

Nucleotide and haplotype diversity parameters were estimated using DnaSP 5.10 (Librado and Rozas, 2009). In order to detect possible signatures of demographic changes in *A. baeticus* lineages, deviations from a model of mutation-drift equilibrium were tested. The demographic history of the populations over the *cob* data was inferred using the Tajima's D (Tajima, 1989), Ramos-Onsins & Rozas' R2 (Ramos-Onsins and Rozas, 2006) and Fu's Fs (Fu, 1997) tests as implemented in DnaSP 5.10, and their significance was assessed using 1.000 coalescent simulated re-samplings. Mismatch analyses were also performed in DnaSP 5.10 and Arlequin 3.1 (Excoffier, et al., 2005). We used initial values of θ_0=0, and θ_1=99,999. Finally, time and magnitude of the inferred population expansion was determined by calculating τ, Θ_0 and Θ_1, where $\tau = 2\mu t$ (μ = mutation rate per site per generation; t = time in generations) and Θ_0 and Θ_1 corresponds to the effective female population size before and after the expansion, respectively. The validity of the assumed stepwise expansion model was tested with a parametric bootstrapping approach (1.000 permutations).

Finally, past population dynamic reconstructions were performed with Bayesian coalescent-based methods as implemented in BEAST V1.7.2 software package (Drummond, et al., 2012). Bayesian skyline plots (BSPs) were generated using Markov chain Monte Carlo (MCMC) sampling to infer past changes in effective population size ($N_e\tau$) of the phylogenetic clades obtained. Simulations were run with a strict molecular clock using uniformly distributed priors and a divergence *cob* mtDNA rate of 0.4 % MY (Perea, et al., 2010) for the clock calibration. The most likely mutation model was estimated with jMODELTEST v.0.1.1, using the Bayesian information criterion (BIC), for each of the mitochondrial lineages obtained. To ensure convergence of the posterior distributions, two independent Markov Chain Monte Carlo (MCMC) runs of 4.000,000 generations, with 4.000 generations of burn-in followed by sampling every 2.000 generations, were performed and combined subsequently in the module LOGCOMBINER (Drummond, et al., 2012). The Bayesian skyline reconstructions were generated using TRACER (Drummond, et al., 2012). A generation time of two times per year was used to estimate N_e (Fernandez-Delgado, et al., 1988; Schönhuth, et al., 2003).

Microsatellite loci amplification

To assess nuclear genetic variation in *A. baeticus*, 19 polymorphic loci previously shown to amplify polymorphic microsatellite in other *Cyprinodontidae* (Babbucci, et al., 2007; Strecker, 2006) were first tested in 20 samples of *A. baeticus*. PCR was performed in 12 µl reaction volumes containing 10x Taq PCR Buffer (Eppendorf, 20 mM Tris-HCl, pH 8, 100mM KCl, 0.1 EDTA, 1 mM DTT), 1.0-2.0 mM $MgCl_2$, 2.02 mM each forward and reverse primer, 0.1 mM of each dNTP, 0.25 U of Taq DNA polymerase (Eppendorf), and approximately 10 ng of template DNA. PCR amplifications consisted of one cycle of denaturing at 95ºC for 5 min; 35 cycles of 94ºC for 1 min, annealing for 45 s at 52ºC - 60 ºC and extension at 72ºC for 1 min; followed by one cycle of 7 min extension at 72ºC (Table S1). Forward primers were labeled with fluorescent dyes (Invitrogen), and amplified PCR products were run on an ABI Prism 3730 DNA Analyzer (250-500 LIZ size standard). Eight out of the 19 loci tested were successfully amplified after PCR optimization. Of those, four were polymorphic and were used in full screening of the rest of the samples (total N=288, Table 1) following the protocol described before. Allele scoring was performed using GeneMapper v3.7 (Applied Biosystems). Approximately 5 % of the samples were re-amplified to ensure scoring repeatability.

Genetic diversity analyses

Microsatellite genetic diversity was quantified for locus and sampling site based on the average number of alleles per locus (N_A), number of alleles standardized to those of the population sample with the smallest size (N_S) (Nei and Chesser, 1983) and the observed (H_O) and expected (H_E) heterozygosities (Nei, 1978), using GENETICS 4.05 (Belkir, 2000) and FSTAT 2.9.3 (Goudet, 2001). Unbiased expected heterozygosity (Nei, 1987) was retained because of the small sample size of some populations. Deviations from Hardy-Weinberg (HW) proportions were tested using the Exact Probability Test for multiple alleles (Guo and Thompson, 1992) available in GENEPOP 4.0.1 (Raymond and Rousset, 1995) at each locus for each population and over all loci for each population. Genotypic

linkage disequilibrium between each pair of loci was estimated by Fisher's exact tests with GENEPOP 4.0.1 software. Both tests for deviations from HW proportions and for linkage disequilibrium used a Markov chain (10000 dememorization steps, 1000 batches, 2000 iterations per batch) (Guo and Thompson, 1992). Correction for multiple testing (type I error rates) was performed using the false discovery rate (FDR) approach (Benjamini and Hochberg, 1995) using the R package QVALUE (Storey, 2002).

A bimodal test for each locus and sampling site was performed to detect possible genotyping errors due to preferential amplification of one of the two alleles, misreading of bands or transcription errors, using the program DROPOUT (McKelvey and Schwartz, 2005). Additionally, MICRO-CHECKER v2.23 (Van Oosterhout, et al., 2004) was used to explore the existence of null alleles, and to evaluate their impact on the estimation of genetic differentiation.

Bayesian clustering analyses

Two different Bayesian clustering methods were used to determine the population genetic structure of *A. baeticus*. The number of populations (*K*) with the highest posterior probability (mean lnProb (D)) was calculated with the program STRUCTURE 2.0 (Pritchard, et al., 2000), assuming an admixed model and a uniform prior probability of the number of populations, *K*. MCMC simulations consisted of 5 x 10^6 burn-in iterations followed by 5 x 10^5 sampled iterations. Furthermore, the modal value of lambda, ΔK (Evanno, et al., 2005) was also calculated to infer the best value of *K*. Ten runs for each value of *K* were conducted to check for consistency in the results. In addition, GENECLASS 2.0 (Piry, et al., 2004) was applied to verify congruence among results obtained with STRUCTURE. The computation followed the partial exclusion method (Rannala and Mountain, 1997), and simulation consisted of 10,000 individuals.

Population structure and gene flow analyses

To determine the amount of genetic structuring among grouping levels, an analysis of molecular variance (AMOVA) (Excoffier, et al., 1992) was performed with GENODIVE, v2.0b17 (Meirmans and Van Tienderen, 2004). Meirmans' standardization of F statistics (by their maximum possible values) was conducted following data transformation in RECODEDATA (Meirmans, 2006). AMOVA was used to assess the partitioning of microsatellite variation in allelic frequencies (F_{ST}) between and within different grouping schemes of geographical differentiation (i.e., among sampling sites and river basins). To visualize the relationship between population samples using microsatellite data, a correspondence analysis (Guinand, 1996) was implemented in GENETIX v.4.05.2 (Belkir, 2000).

A Bayesian approach was also used to determine contemporary gene flow among *A. baeticus* populations using BAYESASS+ v1.3 (Wilson and Rannala, 2003). Two independent runs were performed to test for congruence, and a likelihood ratio test was employed to determine whether the prior and posterior probabilities of migration rates are significantly different from each other (Wilson and Rannala, 2003).

Finally, possible severe reductions in effective population size were assessed using BOTTLENECK (Dirienzo, et al., 1994; Piry, et al., 1999). Analyses were carried out assuming three different mutation models, (i) an infinite allele (IAM), (ii) a stepwise mutation (SMM) and (iii) a two-phase (TPM, with 70 % stepwise, 30 % variable), and applying the Wilcoxon signed rank test for statistical detection of H_e excess. Estimation was based on 10,000 replicates. Also, the mode-shift test (Luikart, et al., 1998) contained in BOTTLENECK was performed to determine whether the observed distribution of allele frequencies among *A. baeticus* populations differed from that expected under drift-mutation equilibrium.

Results

Mitochondrial phylogeographic analyses

Sequences for the *cob* gene were trimmed to the size of the smallest fragment and alignments produced a data set of 1009 base pairs (bp) that yielded 3 unique haplotypes defined by 62 invariable sites and 2 variable sites, from which one was a singleton variable, and one was a parsimony-informative site. The average of G+C frequencies was 0.452. The most common haplotype (93.8 %) was shared between individuals of all drainages. The second most common haplotype (5.3 %) was exclusive of IRO individuals. All methods were able to identify four different lineages. The reconstructed ML and BI phylogeny for the *cob* dataset is shown in Fig. 2 and describes a phylogenetic relationship with no clear geographical pattern (in terms of river drainages), even when only the natural sample populations were included in the analyses (Fig. 2B). Most nodes are highly supported with posterior probabilities > 0.95, with the exception of two interior nodes (i.e., Guadalquivir 1 and 2 lineages). The southern lineage (referred to as South lineage), supported by a high posterior probability, includes individuals from three different drainages (Iro, Roche and Conil), but geographically very close to each other (COPA, ROM and IRO sites). Another lineage, also with a high posterior probability, (referred to as Guadalete) included four haplotypes and was geographically restricted to PEC and PEZ localities. Samples collected at the Guadalquivir and Odiel Rivers were included in the two main lineages (referred to as Guadalquivir 1 and 2). One of these lineages (Guadalquivir 1) is present in salt rivers located upstream and on the left side of the lower reaches of the Guadalquivir River. (Fig. 2A). The other lineage (Guadalquivir 2) is composed by estuarine populations of the Guadalquivir River (San Lucar sampling site) (Fig 2 B). The populations from Doñana National Park that belongs to the conservation programme of breeding and management of the National Park shared haplotypes with the Guadalquivir upstream (Guadalquivir 1) and downstream (Guadalquivir 2) populations. The population of the Odiel Basin is known to be an artificial population founded with Doñana individuals (Doadrio, Personal communication). The individuals from Vega River (VE site) were suspected to be

introduced from some unknown population (Clavero, et al., 2002; Clavero, et al., 2005). Our data seem to be in accordance with an origin of human introduction for the Vega population, apparently coming from Guadalquivir 1 populations (Fig. 2A). When we remove from the analyses the introduced populations and those of the Doñana National Park breeding programme, the populations of Guadalquivir upstream (Guadalquivir 1) are more closely related to the populations of the Guadalete Basin than to those of the Guadalquivir estuary (Guadalquivir 2).

The haplotype network gave complementary information and confirmed the existence of two main phylogroups, one representing the southern populations and the other representing the rest, each of them interconnected by two subgroups of haplotypes that correspond with the four main phylogenetic clades obtained previously. All phylogroups showed a slightly deep star-like topology, somewhat indicating an exponential growth of populations from a small number of individuals. Again, the network pattern did not fully conform to the geographic distribution in terms of river drainages (Fig. 1).

Mitochondrial genetic diversity and historical demography

Genetic diversity indices (π, Hd and S) for the lineages previously recognized are summarized in Table 2. The overall nuclear (π= 0.0017 ± 0.001) and haplotype and Hd= 0.352 ± 0.036) diversities were very low, with the South lineage showing higher *h* and π values, although Guadalquivir 1 displayed a higher number of polymorphic sites (*S*), probably due to the larger number of samples assayed (Table 2).

Hypotheses of demographic expansion were investigated by means of Tajima's *D* (Tajima, 1989), Ramos-Onsins & Rozas' R^2 (Ramos-Onsins and Rozas, 2006) and *Fu's Fs* (Fu, 1997) statistical test of neutrality (Table 2), and also by examining the frequency distribution of pairwise differences between *A. baeticus* sequences (Fig. 3). The neutrality test gave significant negative results for all the

lineages (Table 2) and hence a sudden demographic expansion demographic model cannot be rejected for this species. Moreover, the unimodal distribution (Fig. 3) and the low and significant Tajimas' *D* index obtained for each lineage are also consistent with a model of recent expansion (Table 2). Conversely, the unimodal distribution of the complete data set was not significantly different (as measured by the sum of pairwise differences; $p = 0.402$) from that predicted by the growth expansion model (not shown). Estimated effective female population size for all clades after expansion (Θ_0) was higher than before (Θ_1) expansion indicating a process of expansion (not shown). However, this difference was almost 20 times higher for the South lineage in comparison with the rest of the populations. Moreover, estimated τ -values were very different (and lower for this lineage), indicating that population expansion in that clade may date from different historical periods.

The BSP shows a relatively good fit to a climate change trend since the Late Pleistocene glaciations (LPG), and indicates a continuous population expansion of the lineages Guadalquivir 1 and 2 over the last interglacial 15 Kyr. Population growth of the other two mitochondrial lineages suggests lack of evidence for population expansion since they show a flat line across time (Fig. 4). However, the presence in our data of a common haplotype in all of the lineages may result in underestimation of effective population sizes and time-scales, and thus interpretation of the results should be made with caution. The analysis of a larger mtDNA fragment may reveal additional phylogeographic structure, especially among the recently diverged lineages and may overcome some of the uncertainties present in our dataset.

Nuclear genetic diversity

For the microsatellite diversity estimates and further analyses, SAB, SAI, and SAC sites and ACB, ACO and ACC sites were pooled together due to the reduced number of individuals and their geographical proximity. Two loci showed departure from Hardy-Weinberg (HW) equilibrium due to significant heterozygote deficiency; however, this pattern changed when the analysis was performed

across sample sites, where only COPA showed HW disequilibrium. Tests for linkage disequilibrium showed a very low number of significant pairwise comparisons, which suggests independence of all examined loci.

Overall, the number of alleles detected was 16 (ranging from two to six, mean N_A = 4.2), with the IRO sampling site being the only one with unique alleles. Consistent with the mtDNA data, the amount of genetic variability was very low (mean observed heterozygosity H_O = 0.266) across loci and sampling sites, as would be expected for a species that is genetically depressed (Table 3), with Acebuche (ACB, ACO and ACC sampling sites) being the locality with the lowest genetic diversity (H_O = 0.081). However, these values are slightly higher than those reported previously for the same species (Schönhuth, et al., 2003) or for the closely related killifish, *A. fasciatus* (Angeletti, et al., 2010) at the allozyme variability level, which could indicate a possible genetic recovery of the species.

Corrected estimates of F_{ST} by null allele presence were highly similar to non-corrected F_{ST} values (data not shown) putatively harboring null alleles. Additionally, the bimodal test rendered a non-significant bimodal test indicating no evidence of unspecific locus amplification or genotyping errors, which could have resulted in null alleles. Thus, all loci where used in further analyses.

Nuclear population structure, gene flow and decline

All but ten F_{ST} pairwise comparisons rendered moderate, but significant, results. The non-significant pairwise comparisons mainly involved samples of the Guadalquivir River drainage (e.g. LEP and LEC sites), and conversely the higher F_{ST} values corresponded to comparisons involving El Acebuche (in the Doñana National Park) and the Iro river drainage (Table 4). When populations were grouped into four clusters (corresponding to the four mtDNA lineages), comparisons were also significant in all cases (result not shown). Results of AMOVAs also support a similar pattern of genetic structure. The hierarchical AMOVA reveals overall significant genetic structuring (P <0.005) of the analyzed

samples, (overall F_{ST} = 0.35). However, we observed that the genetic variation of the allele diversity was mainly explained by intra-population variation (24.55 %) whereas a low percentage (1.57 %) of variance was attributable to interpopulation differences.

The number of populations (and the assignment of individuals to each population) was estimated using Bayesian inferences (Pritchard, et al., 2000). Our results detected the highest likelihood for the model with K = 4. However, the modal value of ΔK (Evanno, et al., 2005) was shown at K = 3 (Fig. 5). According to the average proportions of memberships of each pre-defined population (Q), the probability of assignment of all *A. baeticus* individuals was generally low (Q < 70 %). The samples with a higher Q value corresponded to IRO, COPA, VEG and ACB sites, with the samples collected from the Salado River (IRO and COPA) clustering in the sample group, whereas VEG and ACB were assigned to two different clusters. The rest of the individuals could not be assigned exclusively to one cluster, suggesting low genetic structure and large gene flow between river basins. Concordantly, the Bayesian assignment test of GENECLASS, correctly assigned only 29.2 % of the individuals to their own source location, a similar percentage of individuals with a high probability of assignment as detected with STRUCTURE.

The Wilcoxon test detected recent bottlenecks (p<0.05) for all populations in IAM and SMM, and for all populations with the exception of PEZ and PEC when the TPM model was assumed. Additionally, the mode-shift indicator test was employed to detect genetic bottlenecks. Our results indicated that a mode-shift distortion of allele frequency distributions (Luikart, et al., 1998), characteristic of a recent bottleneck, was observed for the majority of the populations (with the exception of IRO and VEG).

Discussion

The present study aimed to clarify the population and genetic structure within and among the endangered killifish *Aphanius baeticus* populations at different timescales. To do so, the genetic structure of *A. baeticus* was explored using two different markers, mtDNA (*cob* gene) and nuclear (microsatellite) markers, the latter being used for the first time in this species. Also, we attempted to cover the entire distribution range of the species, increasing the number of individuals studied per locality with respect to previous molecular studies (Perdices, et al., 2001; Schönhuth, et al., 2003).

Previous studies in killifish species with similar habitat requirements to those of *A. baeticus* (i.e., coastal and high-salinity water bodies) show striking differences in mitochondrial and nuclear genetic distribution, shifting from high to low genetic variability (Angeletti, et al., 2010; Ferrito, et al., 2013; Maltagliati, et al., 2003; Tatarenkov, et al., 2011; Whitehead, 2009). Consequently, it has been argued that these differences in genetic variability patterns are not only determined by habitat adaptation but also by specific life-history traits and historical processes (Fuller, et al., 2007; Strand, et al., 2012; Whitehead, 2009). Overall, the detected mtDNA genetic variability was very low in relation to other non-threatened killifish species (e.g. *A. fasciatus*, (Ferrito, et al., 2013; Pappalardo, et al., 2008; Rocco, et al., 2007). On the other hand, the nuclear genetic variation detected was at similar levels to those reported for the endangered *A. iberus* (unpublished data), and to *A. baeticus* based on 12 polymorphic allozymes and a partial sampling overlap with our study (Perdices, et al., 2001; Schönhuth, et al., 2003), but lower than those reported for *A. fasciatus* (Angeletti, et al., 2010; Maltagliati and Camilli, 2000). Our results also indicated that the patterns of genetic variation throughout the species range are not equally distributed for each of the different molecular datasets used. Mitochondrial data showed some population structuring of genetic variation in populations of *A. baeticus*. On the other hand, the highly sensitive microsatellite data indicated a subtle population structure at a small geographic scale.

Overall, the *A. baeticus* populations have a weaker mtDNA geographical structure in comparison to primary freshwater fish populations of the Cyprinidae and Cobitidae families occurring in the southern part of the Iberian Peninsula (Mesquita, et al., 2005; Perdices and Doadrio, 2001). It has been proposed that the strong geographical structure in these populations of primary fishes have been greatly determined by glacial refugia (Gante, et al., 2009; Gomez and Lunt, 2007). Current distribution for these fish taxa restricted to freshwater environments have been mostly configured by their occurrence in permanent riverbeds when the glacial/interglacials of the Late Pleistocene occurred (Pujolar et al., 2011). Such dispersal limitations to hydrographical networks led to a high degree of genetic differentiation among locations, often resulting in strong population structure after isolation even when posterior contact events may have happened (Mesquita, et al., 2001; Pujolar, et al., 2011; Volckaert, et al., 2002). In contrast, the poor geographical structure of mtDNA haplotypes and the low haplotype diversity detected in our study, suggests that Pleistocene glaciations had a different effect on the population structure of *A. baeticus* (see Doadrio 1988; Rodriguez et al. 1993, and references cited therein) than on other primary freshwater fishes. More tolerant to salinity, *A. baeticus* (as well as other secondary killifish species) inhabits a very restricted range of conditions in rivers' lower reaches, in a fluctuant zone of just a few kilometres wide along the coastline, where conditions are determined by the constant interaction between river input, a precipitation-evaporation regime and sea levels, which defines a ‘tolerance belt’ that the species can inhabit. Specifically, this tolerance belt includes areas with the highest levels of salt (even higher than sea water), such as salt marshes or salt exploitation ponds close to the coast, that are not exclusively dependent on historical processes of river drainages. In addition, these areas are highly susceptible to cyclical floods alternating with drought periods. Thus, genetic variation in secondary fish fauna must endure especially strong seasonal periods involving isolation and founder effects and consequently impacting genetic diversity, population structure and adaptation (Bartakova, et al., 2013; Maltagliati and Camilli, 2000).

The effect of river basin isolation on mitochondrial structure is supported only by the occurrence of lineages exclusive to the Guadalete and the Salado-Roche river basins (Guadalete and South lineages, respectively), which show some level of isolation. This could be explained by a series of changes in eustatic levels, which have been proposed as a main cladogenetic force in several species (Mesquita, et al., 2001; Mesquita, et al., 2005; Perdices, et al., 2001). During the Last Glacial Maximum, the coastline at the Strait of Gibraltar receded more than 120 m under current sea levels (Luque, et al., 1999). Under this scenario, the depletion of sea levels may have allowed contact between river basins and populations, while their rise may have affected their isolation. Cladogenesis in different river basins could have occurred and expansions northwards likely allowed colonization of new areas by ancestral haplotypes. The mitochondrial structure observed today, although weak, could reflect a series of this mechanism along the coast. That is the case for the individuals from the South clade, which retained and held the highest levels of genetic diversity, including unique haplotypes, supporting the hypothesis of survival of individuals in this region. Moreover, the fact that the coastal populations from the Guadalete lineage were more closely related to the inland populations of the Guadalquivir 1 lineage is also in agreement with this hypothesis, as the Guadalquivir 2 could correspond to an ancient splitting event. However, colonization waves from alternative refuge areas, not visible in our sequence variation, should not be discarded.

The habitat expansion hypothesis was further supported by the significant results of the Tajima's D, R2, and Fu's F neutrality tests and mismatch pairwise distributions that rendered a significant signal for population expansion for most of the haplotype groups. The Bayesian coalescent methods also revealed a subtle demographic change that occurred after the last interglacial (15 kya) period. The population expansion of Guadalquivir 1 and 2 (and the weaker signal of expansion for the other lineages), correspond to a warmer and wet period in south Iberia.

This scenario was more recently modified by a succession of secondary contacts and admixture events. Guadalquivir I and Guadalete's clades show the effects of these more recent dispersal waves, likely triggered by the recurrent tsunamis that occurred mainly in the Guadalquivir River drainage region. These populations are situated near larger rivers (even navigable ones, such as the Guadalquivir River), at risk of flooding and with larger discharges from rivers, outfalls, canals or other watercourses, allowing connectivity and subsequent gene flow between populations. However, the short life span of *A. baeticus* (c.a. two years, (Doadrio, 2011; Fernandez-Delgado, et al., 1988)) may endanger the viability of its fragmented populations. As commonly argued for wild populations, demography is likely to be of more importance in population maintenance than genetics alone (Lande, 1988). In this case, environmental stochasticity could lead an entire population to extinction in just two consecutive years without favorable conditions for reproduction, and may make *A. baeticus* populations especially susceptible to local extinction. In addition, in species showing population subdivision as was observed in this case, persistence may greatly depend on the balance between local extinction and the colonization of new suitable habitats, and thus dispersal ability through different habitat patches could be determining (see discussion below).

Microsatellite markers show an overall low genetic diversity, with low allele numbers and heterozygosity values. Moreover, Bayesian clustering analyses and the classic F_{ST} and AMOVA statistics revealed weak but significant genetic structuring across their distribution range, suggesting low levels of gene flow and population admixture. In fact, we observed that most of the samples (except from the COPA population site) were in HW equilibrium and did not show significant inbreeding, suggesting random mating and gene flow between individuals. Furthermore, it was impossible to predict an exact number of inferred populations, which ranged from 3 to 4, and assignment of individuals to the inferred cluster was poor regardless of the different methods used (STRUCTURE or GENECLASS software). This indicates that *A. baeticus* is capable of some dispersion across different populations, and is concordant with the relatively low F_{ST} values detected overall.

However, genetic flow is especially restricted in some cases such as in the IRO sample, where private alleles were observed. Interestingly, the private alleles and the higher structuring of the IRO and COPA populations (detected with Bayesian clustering and classic F_{ST} statistics) is in agreement with the results derived from population genetic analyses based on *cob* sequence data which also detected the higher haplotype diversity and existence of unique haplotypes in this region, highlighting the uniqueness of their population.

Today, human pressures on freshwater fish fauna have been well documented and are considered one of the main causes of population declines and extinctions of many species (such as Mediterranean Cyprinids (Alves, et al., 2001; Mesquita, et al., 2001; Mesquita, et al., 2005; Perdices, et al., 2001)). The results of the equilibrium-based methods (as implemented in BOTTLENECK) support the hypothesis of recent population reduction in several *A. baeticus* populations. However, it is unlikely that human threats could solely account for such a low genetic diversity in the species. It is possible that an interplay with species-specific factors and historical processes may also have contributed to the genetic features of the species as observed today (Gillespie, et al., 2008; Lexer and Fay, 2005). *A. baeticus* is adapted to harsh environmental conditions with extreme fluctuations (Doadrio et al., 1996; Fernández-Pedrosa, 1997), where phenomena such as river floods and seasonal storms can alter the levels of water salinity and temperature in a matter of hours. Under these ecological conditions, it is expected that *A. baeticus* will suffer some fluctuations in population sizes due to the effect of periodic bottlenecks and founder events (Schönhuth, et al., 2003). Most founding events entail a reduction in population size, which could lead to genetic drift and could ultimately compromise the long-term viability of a species (Frankham, et al., 2011). On the other hand, studies have shown that low diversity does not necessary imply the extinction of a species (Groombridge, et al., 2009; Maltagliati, 2002; Reed, 2010). In our case, some species-specific life history traits, such as a high reproductive output, could act to maintain larger population sizes and hence diminish the impact of population declines on genetic variation. Moreover, dispersal of *A. baeticus* individuals may occur

during occasional large-scale floods (e.g. periodic tsunamis). The genetic exchange between populations in a cyclical manner, coupled with the high reproductive output, could provide conditions that favour the survival of the species even under depleted overall genetic diversity. These factors should be considered in future management plans aimed at improving the viability of this endangered killifish species.

Funding

This work was supported by the project "*Asistencia técnica para el análisis genético y el estudio filogeográfico de las poblaciones de salinete (Aphanius baeticus)*" from La Junta de Andalucía and by the projects of the Ministerio de Ciencia e Innovación (CGL2010-15231_BOS), to Ignacio Doadrio.

Acknowledgements

We are grateful to J. Galindo, J. M. Pérez and J. López-Santiago for providing us with samples and C. Fernández provided with helpful insight to the study. Thanks to P. Garzón, J. L. González and P. Ornelas for collection support and L. Alcaraz for her assistance at the laboratory.

References

Akaike H. 1973. Information theory as an extension of the maximum likelihood principle. In 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary: Csaksi, B. N. P. a. F., p. 267-281.

Alves MJ, Coelho H, Collares-Pereira MJ, Coelho MM. 2001. Mitochondrial DNA variation in the highly endangered cyprinid fish *Anaecypris hispanic*a: importance for conservation, Heredity, 87:463-473.

Angeletti D, Cimmaruta R, Nascetti G. 2010. Genetic diversity of the killifish *Aphanius fasciatus* paralleling the environmental changes of Tarquinia salterns habitat, Genetica, 138:1011-1021.

Babbucci M, Pappalardo AM, Ferrito V, Barbisan F, Patarnello T, Tigano C. 2007. Isolation and characterization of eight polymorphic microsatellite markers in *Aphanius fasciaus* (Teleostei: Cyprinodontidae), Molecular Ecology Notes, 7:293-295.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies, Molecular biology and evolution, 16:37 - 48.

Bartakova V, Reichard M, Janko K, Polacik M, Blazek R, Reichwald K, Cellerino A, Bryja J. 2013. Strong population genetic structuring in an annual fish, *Nothobranchius furzeri*, suggests multiple savannah refugia in southern Mozambique, BMC evolutionary biology, 13:196.

Belkir K, P. Borsa, L. Chickhi, N. Raufaste and F. Bonhomme (ed.). 2000. *GENETIX 4.04 Logici el sous Windows TM, pour la Génétique des Populations*. In. Montpellier, France.: Université de Montpellier II, Laboratoire Génome, Populations, Interactions, CNRS UMR 5000.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society Series B-Methodological, 57:289-300.

Blanco JL, Hrbek T, Doadrio I. 2006. A new species of the genus Aphanius (Nardo, 1832) (Actinopterygii, Cyprinodontidae) from Algeria, Zootaxa, 1158:39-53.

Clavero M, Blanco F, Rebollo A, Delibes M, Prenda J. 2002. Nueva población de fartet atlántico hallada en el Campo de Gibraltar, Quercus, 201:12.

Clavero M, Blanco-Garrido F, Rebollo A. 2005. El salinete (Aphanius baeticus) en el río de la Vega (Tarifa): Caracterización de la población y factores de amenaza., Almoraima, 31:87-92.

Dirienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB. 1994. Mutational processes of simple-sequence sepeat loci in human-populations, Proceedings of the National Academy of Sciences of the United States of America, 91:3166-3170.

Doadrio I. 2001. Atlas y libro rojo de los peces continentales de España. Madrid: Ministerio de Medio Ambiente, Consejo Superior de Investigaciones Cientìficas (CSIC).

Doadrio I. 2011. Ictiofauna continental española. Bases para su seguimento. Madrid: DG Medio Natural y Politica Forestal. MARM.

Doadrio I, Carmona JA, Fernandez-Delgado C. 2002. Morphometric study of the Iberian *Aphanius* (Actinopterygii, Cyprinodontiformes), with description of a new species, Folia Zoologica, 51:67-79.

Doadrio I, Perdices A, Machordom A. 1996. Allozymic variation of the endangered killifish *Aphanius iberus* and its application to conservation, Environmental Biology of Fishes, 45:259-271.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7, Molecular biology and evolution, 29:1969-1973.

Earl D, vonHoldt B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, Conservation Genet Resour, 4:359-361.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic acids research, 32:1792-1797.

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.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol Bioinform Online, 1:47-50.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data, Genetics, 131:479-491.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach, J Mol Evol, 17:368-376.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap, Evolution, 39:783-791.

Fernandez-Delgado C, Hernando JA, Herrera M, Bellido M. 1988. Age, growth and reproduction of *Aphanius iberus* (Cuv. & Val., 1846) in the lower reaches of the Guadalquivir river (south-west Spain), Freshwater Biology, 20:227-234.

Ferrito V, Pappalardo A, Canapa A, Barucca M, Doadrio I, Olmo E, Tigano C. 2013. Mitochondrial phylogeography of the killifish *Aphanius fasciatus* (Teleostei, Cyprinodontidae) reveals highly divergent Mediterranean populations, Mar Biol:1-16.

Frankham R, Ballou JD, Eldridge MDB, Lacy RC, Ralls K, Dudash MR, Fenster CB. 2011. Predicting the probability of outbreeding depression, Conservation Biology, 25:465-475.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection, Genetics, 147:915-925.

Fuller RC, McGhee KE, Schrader M. 2007. Speciation in killifish and the role of salt tolerance, Journal of evolutionary biology, 20:1962-1975.

Gante HF. 2011. Diversification of circum-Mediterranean barbels. In Grillo O ed. Changing diversity in changing environment. Rijeka, Croatia: inTech, p. 283-298.

Gante HF, Micael J, Oliva-Paterna FJ, Doadrio I, Dowling TE, Alves MJ. 2009. Diversification within glacial refugia: tempo and mode of evolution of the polytypic fish *Barbus sclateri*, Molecular Ecolology, 18:3240-3255.

Gillespie RG, Claridge EM, Roderick GK. 2008. Biodiversity dynamics in isolated island communities: interaction between natural and human-mediated processes, Molecular ecology, 17:45-57.

Gomez A, Lunt DH. 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula, Phylogeography of Southern European Refugia:155 - 188.

Goudet J. 2001. FSTAT, a programe to estimate and test gene diversities and fixation indices, Version 2.9.3. Available from: http://www.unil.ch/izea/softwares/fstat.html. In.

Groombridge JJ, Dawson DA, Burke T, Prys-Jones R, Brooke MD, Shah N. 2009. Evaluating the demographic history of the Seychelles kestrel (*Falco araea*): Genetic evidence for recovery from a population bottleneck following minimal conservation management, Biological Conservation, 142:2250-2257.

Guinand B. 1996. Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations, Biological Journal of the Linnean Society, 58:173-195.

Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles, Biometrics, 48:361-372.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA, Journal of molecular evolution, 22:160-174.

Hrbek T, Kucuk F, Frickey T, Stolting KN, Wildekamp RH, Meyer A. 2002. Molecular phylogeny and historical biogeography of the *Aphanius* (Pisces, Cyprinodontiformes) species complex of central Anatolia, Turkey, Molecular Phylogenetics and Evolution, 25:125-137.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology, Science, 294:2310-2314.

Lande R. 1988. Genetics and demography in biological conservation, Science, 241:1455-1460.

Lario J, Zazo C, Goy JL, Silva PG, Bardaji T, Cabero A, Dabrio CJ. 2011. Holocene palaeotsunami catalogue of SW Iberia, Quaternary International, 242:196-200.

Lexer C, Fay MF. 2005. Adaptation to environmental stress: a rare or frequent driver of speciation?, Journal of evolutionary biology, 18:893-900.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics, 25:1451-1452.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks, Journal of Heredity, 89:238-247.

Luque L, Zazo C, Recio JM, Dueñas MA, Goy JL, Lario J, González-Hernández F, Dabrio CJ, González-Delgado A. 1999. Evolución sedimentaria de la laguna de la Janda (Cádiz) durante el Holoceno, Cuaternario y Geomorfología, 13:43-50.

Machordom A, Doadrio I. 2001. Evidence of a cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (Luciobarbus, Cyprinidae), Molecular phylogenetics and evolution, 18:252-263.

Maltagliati F. 2002. Genetic monitoring of brackish-water populations: the Mediterranean toothcarp *Aphanius fasciatus* (Cyprinodontidae) as a model, Marine Ecology Progress Series, 235:257-262.

Maltagliati F, Camilli L. 2000. Temporal Genetic Variation in a Population of *Aphanius fasciatus* (Cyprinodontidae) from a Brackish-water Habitat at Elba Island (Italy), Environmental Biology of Fishes, 57:107-112.

Maltagliati F, Domenici P, Fosch CF, Cossu P, Casu M, Castelli A. 2003. Small-scale morphological and genetic differentiation in the Mediterranean killifish *Aphanius fasciatus* (Cyprinodontidae) from a coastal brackish-water pond and an adjacent pool in northern Sardinia, Oceanologica Acta, 26:111-119.

McKelvey KS, Schwartz MK. 2005. DROPOUT: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework, Molecular Ecology Notes, 5:716-718.

Meirmans P. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure, Evolution, 60:2399-2402.

Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms, Molecular Ecology Notes, 4:792-794.

Mesquita N, Carvalho G, Shaw P, Crespo E, Coelho MM. 2001. River basin-related genetic structuring in an endangered fish species, *Chondrostoma lusitanicum*, based on mtDNA sequencing and RFLP analysis, Heredity, 86:253-264.

Mesquita N, Hanfling B, Carvalho GR, Coelho MM. 2005. Phylogeography of the cyprinid Squalius aradensis and implications for conservation of the endemic freshwater fauna of southern Portugal, Molecular ecology, 14:1939-1954.

Moreno-Amichi R, Planelles-Gomis M, Fernández-Delgado C, García-Berthou E. 1999. Distribución geográfica de los ciprinodontiformes en la Península Ibérica. In Planelles-Gomis M ed. Peces ciprinodóntidos Ibéricos, Fartet y Samaruc. Valencia: Generalitat Valenciana: Conselleria de Medio Ambiente, p. 33-57.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals, Genetics, 89:583-590.

Nei M. 1987. Molecular evolutionary genetics. New York, USA.

Nei M. 1987. Molecular evolutionary genetics. . New York: Columbia Univ. Press.

Nei M, Chesser R. 1983. Estimation of fixation indices and gene diversities, Annals of Human Genetics, 47:253-259.

Oltra R, Todolí R. 2000. Reproduction of the endangered killifish *Aphanius iberus* at different salinities, Environmental Biology of Fishes, 57:113-115.

Pappalardo AM, Ferrito V, Messina A, Guarino F, Patarnello T, De Pinto V, Tigano C. 2008. Genetic structure of the killifish *Aphanius fasciatus*, Nardo 1827 (Teleostei, Cyprinodontidae), results of mitochondrial DNA analysis, Journal of Fish Biology, 72:1154-1173.

Perdices A, Carmona JA, Fernandez-Delgado C, Doadrio I. 2001. Nuclear and mitochondrial data reveal high genetic divergence among Atlantic and Mediterranean populations of the Iberian killifish *Aphanius iberus* (Teleostei: Cyprinodontidae), Heredity, 87:314-324.

Perdices A, Doadrio I. 2001. The molecular systematics and biogeography of the European cobitids based on mitochondrial DNA sequences, Molecular phylogenetics and evolution, 19:468-478.

Perea S, Bohme M, Zupancic P, Freyhof J, Sanda R, Ozulug M, Abdoli A, Doadrio I. 2010. Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data, BMC evolutionary biology, 10:265.

Piry S, Alapetite A, Cornuet J, Paetkau D, Baudouin L, Estoup A. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection, Journal of Heredity, 95:536.

Piry S, Luikart G, Cornuet JM. 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data, Journal of Heredity, 90:502-503.

Posada D. 2008. JModelTest: phylogenetic model averaging, Molecular biology and evolution, 25:1253 - 1256.

Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks, Trends in Ecology & Evolution, 16:37-45.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data, Genetics, 155:945-959.

Pujolar JM, Lucarda AN, Simonato M, Patarnello T. 2011. Restricted gene flow at the micro- and macro-geographical scale in marble trout based on mtDNA and microsatellite polymorphism, Frontiers in Zoology, 8:7.

Ramos-Onsins SE, Rozas J. 2006. Statistical properties of new neutrality tests against population growth (vol 19, pg 2092, 2002), Molecular biology and evolution, 23:1642-1642.

Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes, Proceedings of the National Academy of Sciences of the United States of America, 94:9197-9201.

Raymond M, Rousset F. 1995. An exact test for population differentiation, Evolution, 49:1280-1283.

Reed DH. 2010. Albatrosses, eagles and newts, Oh My!: exceptions to the prevailing paradigm concerning genetic diversity and population viability?, Animal Conservation, 13:448-457.

Rocco L, Ferrito V, Costagliola D, Marsilio A, Pappalardo AM, Stingo V, Tigano C. 2007. Genetic divergence among and within four Italian populations of *Aphanius fasciatus* (Teleostei, Cyprinodontiformes), Italian Journal of Zoology, 74:371-379.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space, Systematic biology, 61:539-542.

Ruiz F, Abad M, Rodríguez-Vidal J, Cáceres LM, Carretero MI, Pozo M. 2008. The Holocene record of tsunamis in the southwestern Iberian Margin: date andconsequences of the next tsunami. In 6ª Assembleia Luso-Espanhola de Geodesia e Geofísica. Tomar, Portugal, p. 365-366.

Schönhuth S, Luikart G, Doadrio I. 2003. Effects of a founder event and supplementary introductions on genetic variation in a captive breeding population of the endangered Spanish killifish, Journal of Fish Biology, 63:1538-1551.

Storey JD. 2002. A direct approach to false discovery rates, Journal of the Royal Statistical Society Series B-Statistical Methodology, 64:479-498.

Strand AE, Williams LM, Oleksiak MF, Sotka EE. 2012. Can diversifying selection be distinguished from history in geographic clines? A population genomic study of killifish (*Fundulus heteroclitus*), PloS one, 7:e45138.

Strecker U. 2006. Characterization and cross-species amplification of microsatellite loci in a Cyprinodon species flock, Molecular Ecology Notes, 6:843-846.

Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b 10.

Tajima F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism, Genetics, 123:585-595.

Tatarenkov A, Lima SM, Avise JC. 2011. Extreme homogeneity and low genetic diversity in *Kryptolebias ocellatus* from south-eastern Brazil suggest a recent foundation for this androdioecious fish population, Journal of Fish Biology, 79:2095-2105.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data, Molecular Ecology Notes, 4:535-538.

Volckaert FAM, Hänfling B, Hellemans B, Carvalho GR. 2002. Timing of the population dynamics of bullhead Cottus gobio (Teleostei: Cottidae) during the Pleistocene, Journal of evolutionary biology, 15:930-944.

Whelan F, Kelletat D. 2005. Boulder deposits on the southern spanish Atlantic coast: possible evidence for the 1755 AD Lisbon tsunami?, Science of tsunami hazards, 23:25.

Whitehead A. 2009. Comparative mitochondrial genomics within and among species of killifish, BMC evolutionary biology, 9:11.

Wilson GA, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes, Genetics, 163:1177-1191.

Figure legends

Figure 1.

Map of the sample location of the *Aphanius baeticus* individuals used in this study. The potential distribution area (hatched) has been drawn according to the IUCN. The codes indicate the sampling localities and are specified in Table 1. The median-joining network showing the relationships among mtDNA haplotypes of *A. baeticus* is also indicated. For the network representation, circles represent haplotypes and sizes are proportional to the haplotype frequency. Haplotype circles are filled corresponding to the colour code for each sample site.

Figure 2.

Phylogenetic relationships of *A. baeticus* based on mtDNA *Cob* data using Bayesian inference (BI) and Maximum Likelihood (ML), including all the samples (**A**) or only the natural populations (**B**). Numbers represent support for maximum likelihoods bootstrap and posterior probabilities (up and down each node, respectively)

Figure 3.

Observed (bars), growth-decline population model (continuous line) and constant population model mismatch distribution (dotted line) for all pairwise differences of the *A. baeticus cob* gene for: (A) Guadalquivir 1; (B) Guadalete; (C) Guadalquivir 2 and (D) South lineages.

Figure 4.

Bayesian skyline plot (BSP) showing historical demography of each mtDNA lineage obtained from the *A. baeticus* analyses. The y-axis equals changes in effective population sizes (Ne) and the x-axis

measures time in thousand of years (Kya) before the present. The black line is the median estimate and the blue lines show the 95% highest posterior density intervals.

Figure 5.

Number of *A. baeticus* populations with the highest posterior probability expressed as **A**) the mean likelihood (log P ($X|K$)), and **B**) ΔK, over 10 runs for each K (one to 16). Graph has been created using Harvester Structure (Earl and vonHoldt, 2012).