UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

Genetic diversity and population structure in the Iberian endangered *Iberochondrostoma lemmingii* (Steindachner, 1866)

Miguel Antunes Lopes da Cunha

Mestrado em Biologia da Conservação

2009

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

Genetic diversity and population structure in the Iberian endangered *Iberochondrostoma lemmingii* (Steindachner, 1866)

Miguel Antunes Lopes da Cunha

Dissertação orientada por:
Professora Doutora Maria Manuela Coelho
Doutora Maria Ana Aboim

Mestrado em Biologia da Conservação

2009

AGRADECIMENTOS

À professora Maria Manuela Coelho por me ter aceitado como seu mestrando e guiado na execução desta tese.

À Maria Ana Aboim pela sua co-orientação e ajuda na resolução dos problemas mais práticos da tese.

Ao Miguel Santos, Filipa Filipe, Maria Drago e Filipe Ribeiro pela captura dos exemplares utilizados.

À Mónica e à Ana Rita sempre disponíveis para tirar qualquer duvida estatística e funcionamento dos software.

À Silke, Isa e Margarida que tanto aturaram a minha falta de prática laboratorial, sempre disponíveis para tirar qualquer dúvida.

À Dona Branca sempre prestável e sempre pronta a resolver qualquer problema.

Ao Tomás pelas conversas e companhia ao almoço.

À Bárbara por partilhar o desespero à medida que a data de entrega se aproximava.

Um especial obrigado à Professora Maria Manuela Coelho, à Maria Ana Aboim, ao Miguel Santos e à Bárbara pelas suas incansáveis revisões da dissertação.

Aos meus pais por desde pequeno me incutirem o gosto pela biologia.

RESUMO

A região mediterrânica é considerada um dos 25 "hotspots" mundiais, devendo, assim, ser encarada como prioritária para a conservação. Dentro desta região encontra-se o Sul da Península Ibérica que apresenta um elevado número de endemismos, a grande maioria dos quais se encontra ameaçada. O género *Iberochondrostoma*, corresponde a um dos géneros endémicos de peixes dulciaquícolas da família Cyprinidae que pode ser encontrado na Península Ibérica, sendo representado por quatro espécies, todas com estatuto de ameaça. Destas espécies, *Iberochondrostoma lemmingii*, a única para a qual não existe qualquer estudo de genética populacional em Portugal, encontra-se distribuída nas grandes bacias Ibéricas dos rios Tejo, Guadiana e Guadalquivir e nas pequenas bacias hidrográficas de Odiel, Quarteira, Gilão e Almargem. Dado que a sua distribuição dentro destas bacias é muito fragmentada, põe-se a hipótese de que a espécie se encontre numa situação ainda mais preocupante em termos de viabilidade do que a que é sugerida pelo estatuto de conservação "Em Perigo" que foi atribuído às populações portuguesas.

Trata-se de uma espécie que pode ser encontrada em rios e ribeiras de carácter temporário, o que leva à ocorrência de "bottlenecks" das populações durante o Verão, quando o seu habitat fica reduzido a pequenos pegos. A variação anual do número de efectivos leva, assim, a perdas de diversidade genética que podem vir a comprometer o potencial evolutivo da espécie. O facto de esta espécie não ser encontrada em albufeiras é também um factor muito importante a considerar, principalmente, com a construção de novas barragens, como, por exemplo, a Barragem do Alqueva que levou, eventualmente, ao aumento da fragmentação das populações do Guadiana.

Vários estudos já foram efectuados noutras espécies de ciprinídeos demonstrando que as mesmas se encontram bastante ameaçadas principalmente devido a fragmentação de habitats, introdução de espécies invasoras, poluição de águas, construção de barragens e capturas ilegais de águas. Espécies que habitam o mesmo ambiente estão sujeitas aos mesmos constrangimentos ecológicos, sendo, assim, de especial importância compreender a condição genética das populações de outras espécies simpátricas com *I. lemmingii*, de modo a obter uma perspectiva mais alargada e uma melhor interpretação dos padrões de variabilidade genética encontrados na espécie em estudo.

Populações de *I. lemmingii* compreendidas na sua área de distribuição em Portugal foram amostradas com o objectivo de delinear um plano para a sua gestão. Dos exemplares amostrados foram retiradas pequenas porções das barbatanas pélvicas para extracção e análise de ADN, com o objectivo de se obter dados sobre a variabilidade genética, filogeografia e história demográfica da espécie, através da análise do gene mitocondrial do citocromo *b* e vários *loci* de microssatélites.

Estes marcadores encontram-se entre os mais usados em estudos de Genética da Conservação, principalmente devido ao facto de, presumivelmente, não serem afectados por fenómenos de selecção, sendo sua diversidade resultante da interacção entre fenómenos de mutação e deriva genética, o que proporciona uma descrição mais correcta dos fenómenos que deram origem ao actual estatuto de conservação de uma espécie ou população. Ambos os marcadores são obtidos por reacções de PCR (*Polimerase Chain Reaction*), sendo apenas necessária uma pequena porção de ADN que pode ser obtida através de processos de amostragem não invasivos, o que é de especial importância no estudo de espécies ameaçadas.

De modo a compreender a variabilidade e a estrutura genética de *I. lemmingii*, indivíduos de cinco localidades diferentes foram amostrados: Rio Caia e Ribeira de Foupana (bacia hidrográfica do Guadiana, respectivamente, N=51 e N=45), Ribeira de Quarteira (bacia hidrográfica de Quarteira, N=42), Rio Aravil (bacia hidrográfica do Tejo, N=34) e Rio Almargem (bacia hidrográfica de Almargem, N=15). Todos os indivíduos foram genotipados para um total de 7 *loci* microssatélites, tendo-se ainda sequenciado 1013 pares de bases do gene mitocondrial do citocromo *b* para um total de 63 indivíduos (14 de Foupana, 13 de Caia, 12 de Quarteira, 13 de Aravil e 12 de Almargem).

Os resultados apontaram para a ocorrência de maior variabilidade nas populações de *I. lemmingii* quando comparadas com as outras espécies de ciprinídeos presentes no Sul de Portugal para as quais já foram efectuados estudos de genética das populações. A amostra de Foupana, proveniente da região Sul do Guadiana, revelou os maiores níveis de variabilidade, tanto em termos de diversidade haplotídica como em termos de riqueza alélica. Já a amostra de Quarteira, apresentou apenas um haplótipo, não tendo esta baixa variabilidade sido corroborada pelos dados dos microssatélites que demonstraram o segundo maior valor em termos de heterozigotia esperada. Embora todos os valores *pairwise* de F_{ST} tenham sido significativos, o que revela a existência de diferenciação entre todas as amostras, os valores de Φ_{ST} de comparação entre as

amostras de Quarteira e Almargem não foram significativos, o que pode ser indicativo de uma origem comum mais recente em relação às restantes amostras. De facto, os valores de F_{ST} e Φ_{ST} revelaram uma estruturação significativa entre as amostras de Caia e Foupana, ambas pertencentes à bacia hidrográfica do Rio Guadiana tendo esta sido suportada pelos resultados obtidos com o programa STRUCTURE que agrupou as amostras do Caia (região Norte do rio Guadiana) com as amostras do Aravil (bacia hidrográfica do rio Tejo) e as amostras do Foupana (região Sul do rio Guadiana) com as amostras de Almargem (uma pequena bacia hidrográfica da região Sul de Portugal). A estrutura observada no interior da bacia hidrográfica do rio Guadiana pode ser explicada tanto pela baixa mobilidade de *I. lemmingii*, como pelo limite de intrusão de águas marinhas no rio principal, impossibilitando, assim, a existência de fluxo genético entre espécies de diferentes tributários, o que, consequentemente, leva a um aumento da diferenciação entre as populações aí presentes devido a fenómenos de deriva genética. No entanto, a aparente fraca diferenciação entre populações de diferentes bacias indica a existência de uma estrutura populacional que não é definida pelos limites das próprias bacias hidrográficas, um caso que não se encontra em concordância com a maioria dos resultados obtidos para outras espécies de ciprinídeos, especialmente aquelas já estudadas no sul de Portugal. Esta baixa diferenciação entre populações de diferentes bacias pode ser explicada pela história das mesmas: por exemplo, as bacias hidrográficas do Tejo e Guadiana estiveram ligadas até há 80 000 anos, o que pode explicar a baixa diferenciação das populações presentes em ambas as bacias. Uma origem comum para as populações das bacias hidrográficas de Quarteira e Almargem é também proposta através da análise da rede de haplótipos criada que demonstra uma grande semelhança entre ambas as amostras. Para além disso, uma vez que os haplótipos presentes nestas bacias parecem ter tido origem nos encontrados na amostra de Foupana, é ainda possível que as mesmas partilhem uma origem comum com a população da bacia hidrográfica do Rio Guadiana.

Assim, tendo em conta os resultados obtidos, todas as populações amostradas foram sugeridas como MU's (Management Units *sensu* Moritz, 1994) diferentes.

Palavras chave: Sul da Península Ibérica; Cyprinidae; *Iberochondrostoma lemmingii*; Genética da Conservação; Microssatélites; ADN mitocondrial.

ABSTRACT

The Mediterranean region is considered one of the 25 global hotspots for biodiversity. The south, of the Iberian Peninsula, is part of this hotspot and holds a great number of endemic species, most of which are currently threatened. The genus *Iberochondrostoma* belonging to the Cyprinidae family is currently represented by four species, all of them threatened. Of these, *I. lemmingii* is the only one with any type of genetic study for conservation purposes. This species occurs in the Tagus, Guadiana and Guadalquivir large Iberian drainages as well as in the smaller basins of Quarteira, Gilão, Almargem and Odiel.

The mitochondrial citochrome *b* gene and seven microsatellite *loci* were used to investigate the genetic variability and differentiation of samples from Aravil (Tagus drainage), Caia and Foupana (Guadiana drainage), Almargem and Quarteira (both small Mediterranean type drainages).

The obtained results revealed higher levels of genetic variation in *I. lemingii* than those previously described for other cyprinid species in the south region of Portugal. The Foupana population from the Guadiana basin revealed the higher levels of haplotype diversity and allelic richness and the Quarteira population, presented only one mitochondrial haplotype despite high allelic diversity for microsatellites. Despite a high degree of differentiation between all samples, the population structure for *I. lemmingii* revealed a strong intra-drainage differentiation as well as some inter-drainage similarities, with the samples from Caia (northern Guadiana drainage) being more similar to the samples from Aravil (Tagus drainage) than to the samples of Foupana (southern Guadiana drainage). Although the differentiation within the Guadiana drainage was expected as it had already been reported for *A. hispanica*, the seemingly week inter-drainage differentiation contrasts with what has been reported for other cyprinids species in the southern region of Portugal. Although all pairwise F_{ST} values were significant, the Φ_{ST} values for the comparison between Quarteira and Almargem were non significant, which might be indicative of a common origin. No evidences of bottlenecks were detected and the expansion events that were detected seem to correspond to inter-glacial periods. With the obtained results it was suggested that all populations should be considered different MUs (Management Units).

Keywords: Iberian Peninsula South; Cyprinidae; *I. lemmingii*; Conservation Genetics; microsatellite; mitochondrial DNA.

CONTENTS

1. INTRODUCTION

The Circum-Mediterranean region is characterized by a great number of endemic species and, simultaneously, excessive losses of habitat, which qualifies it as one of the 25 biodiversity hotspots for conservation priority (Myers *et al.*, 2000). Only 4.7% of the original 2 362 000 km^2 of primary vegetation of this region remain and only 42 123 km^2 (38.3%) are considered protected. The Circum-Mediterranean region comprises a total of 770 species of terrestrial vertebrates, 235 of which are endemic (0.9% of all terrestrial vertebrates species), and 25.000 plant species of which 13 000 are endemic (4.3% of all plants species) (Myers *et al.*, 2000).

The Iberian Peninsula, isolated from the rest of Europe by the Pyrenees, has a total area of 581 000 km^2 providing habitat for a total of 64 different native fish species (forty of which are primary species, five secondary species and 13 peripheral species, i.e. saltwater tolerant) across eleven biogeographical provinces (Filipe *et al.*, 2009), with its southern range, integrant part of the Circum-Mediterranean hotspot, harboring the majority of its endemic species, adding special scientific and conservationist relevance to this region.

1.1 Cyprinidae: Origin and diversity in the Iberian Peninsula

Cyprinids represent the largest freshwater fish family and the second largest of all fishes in the world, containing approximately 2010 species (Helfman *et al.*, 1997). The Asian southeast presents the largest diversity in number of Cyprinid species, followed by Africa, North America and Europe (Helfman *et al.*, 1997).

There are currently three biogeographical hypotheses to explain the origin and diversity of the cyprinid fauna in the Iberian Peninsula, all of them assuming an Asiatic origin (Sanjur *et al.*, 2002): i) the dispersion of cyprinids into the Iberian Peninsula might have occurred across central Europe through river connections which was possible until the Pliocene, i.e. prior to the isolation of the Iberian Peninsula by the Pyrenees (Barnarescu *in* Sanjur *et al.*, 2002); ii) the dispersal of the cyprinid species occurred via Lago-Mare phase of the Mediterranean during the Messian crisis of the

Pliocene (5.5 MYA), when the Mediterranean basin was separated from the Atlantic Ocean and filled with freshwater from the Sarmatic Sea, thus allowing the dispersal of cyprinids around the Circum-Mediterranean region (Bianco, 1990) and; iii) the dispersal would have occurred through continental land-bridges during the formation of the actual North African coast (Doadrio, 1990). Speciation that subsequently occurred within the Iberian Peninsula has been explained by two separate geographical events (Sanjur *et al.*, 2003; Robalo *et al.*, 2006): a first stage of differentiation in the Miocene, due to the internal configuration on the Iberian endorheic drainage phase, followed by a second stage during the Pliocene-Pleistocene characterized by the re-arrange of the Iberian drainage system to an exoreic one and consequently by barriers associated with the current drainage configuration (Calvo *et al.*, 1993).

The internal complexity and importance of the Iberian Peninsula is well supported by several case studies in different animal species (Gómez & Lunt, 2006). Current data indicate a general trend in the increase of species diversity and number of endemic species from North to South, which raises questions about the colonization paths and biogeographical history of the Iberian Peninsula (Mesquita *et al.*, 2007). The *Squalius* genus is one of such case studies, as it presents a clear division between northern and southern drainages. Only one species is present in the northern region (*S. carolitertii*, Doadrio, 1987) and three species can be found in the southern range of the Iberian Peninsula: *S. pyrenaicus* (Günther, 1868); *S. torgalensis* (Coelho, Bogutskaya, Rodrigues & Collares-Pereira, 1998) and *S. aradensis* (Coelho, Bogutskaya, Rodrigues & Collares-Pereira, 1998). *S. aradensis* and *S. torgalensis* are basal to *S. pyrenaicus* and *S. carolitertii* with a probable differentiation around 6 MYA, when many basins almost dried up and became isolated. The separation between *S. carolitertii* and *S. pyrenaicus* is more recent and is attributed to the earlier split of the Douro basin in relation to the remaining Iberian drainages. Also *S. pyrenaicus* presents higher levels of genetic variability than *S. carolitertii* revealing the influence of ancient endorheic lagoons that originated some of the main rivers of the Iberian Peninsula (Sanjur *et al.*, 2003).

1.2 The *Iberochondrostoma* genus

Until recently, the Cyprinidae family was only represented by the genera *Barbus* (sub-family Cyprininae), *Squalius*, *Anaecypris* and *Chondrostoma* (subfamily Leuciscinae), but Robalo *et al*. (2006) proposed the division of *Chondrostoma* in six new genera: *Chondrostoma*, *Achondrostoma*, *Parachondrostoma*, *Pseudochondrostoma*, *Iberochondrostoma*, all represented in the Iberian Peninsula and also the *Protochondrostoma* genus that can only be found in Italy and Slovenia.

More particularly, the *Iberochondrostoma* genus includes *I. lemmingii* (Steindachner, 1866), *I. Lusitanicum,* (Collares-Pereira, 1980), *I. oretanum* (Doadrio & Carmona, 2003), *I. almacai* (Coelho, Mesquita & Collares-Pereira, 2005) and is morphologically characterized by: arched mouth without horny layer on the lower lip; 46–60 canaliculated scales on the lateral line; 11–12 scales above the lateral line; 5–6 scales below the lateral line; 7–8 branched rays in the ventral fin; 6–7 branched rays in the dorsal fin; 6–7 branched rays in the anal fin; 6–5/5 pharyngeal teeth; 24–27 gill rakers on the first branchial arch and the upper branch of the fifth ceratobranchial enlarged (Robalo *et al.*, 2006).

1.2.1 *Iberochondrostoma lemmingii*

The *Iberochondrostoma lemmingii*, one of the several endemic species of the Iberian Peninsula, can be found in permanent or intermittent river streams with moderate current and abundant aquatic vegetation (Cabral *et al.*, 2005). Its diet is mainly composed by algae and zooplankton and its reproductive season is between April and May (Kottletat & Freyhof, 2007). The average length is 14.4 cm for females and 11.4 cm for males, with sexual maturity being attained by the age of two (Fernandez-Delgado & Herrera, 1995) and no dispersal reproductive pattern (Doadrio & Carmona, 2003). This species occurs in Tagus, Guadiana and Guadalquivir Iberian drainages as well as in the small basins of Quarteira, Gilão, Almargem and Odiel (Figure 2 *in* Methods) (Doadrio *et al*., 2001; Cabral *et al*., 2005).

In Tejo, it is alopatric to the sister-species *I. lusitanicum*, occupying only the upper region of the basin, and sympatric with *Pseudochondrostoma polylepis*

(Steindachner, 1865) and *S. pyrenaicus*. In the Guadiana basin is sympatric with *Pseudochondrostoma willkommii* (Steindachner, 1866), *S. alburnoides* Steindachner, 1866, *Anaecypris hispanica* (Steindachner, 1866) and *S. pyrenaicus*. In smaller river basins is sympatric to *S. pyrenaicus* and *S. alburnoides* in Almargem and to *S. aradensis* in Quarteira (Cabral *et al*., 2005).

Figure 1. *Iberochondrostoma lemmingii*, photo by MAAboim.

Although its range distribution includes several different basins, in Portugal it only comprises an area of 150 km^2 and is currently classified as "Endangered" (Cabral *et al*., 2005). As for the remaining populations present in Spain, they are currently classified as "Vulnerable" (Doadrio *et al*., 2001).

A study on the genetic variability of Spanish populations of *I. lemmingii* has already been undertaken by Carmona *et al.* (2000), using cytochrome *b* gene and allozymes. Results obtained with both molecular markers were congruent and demonstrated a high level of differentiation among hydrographical basins, supporting the hypothesis of fragmentation of an ancestral widespread population in several regions that correspond to the actual configuration of the hydrological basins present in the Iberian Peninsula. However, this study did not contemplate Portuguese populations. Thus, present work will focus on the Portuguese population, as they were not included

in any previous study at the population level and will use mitochondrial DNA cytochrome *b* gene and microsatellites, a more informative nuclear marker than allozymes, already used in the Spanish populations.

1.3 Conservation genetic studies

1.3.1 Molecular markers in Conservation Genetics

Since mitochondrial DNA and microsatellites are considered as not being affected by selection, their diversity will result from the interaction of mutation and genetic drift (Moran, 2002), being the most commonly used molecular marker in conservation studies based on population genetics. Furthermore, they allow the quantification of genetic flow, providing specific and comparable measurements of processes that affect endangered populations providing an adequate description of the processes that gave rise to the current endangered state of a population or species (De Salle & Amato, 2004).

Conservation Genetics primary goal is "to maintain each species as a genetic dynamic unity capable of evolving with the environment" (Frankham *et al.*, 2004).

Both mitochondrial DNA and microsatellites are obtained through Polymerase Chain Reactions (PCR) requiring small amounts of biological samples that can be collected through non-invasive sampling, permitting studies on endangered species to be conducted without harm to the specimens.

Mitochondrial DNA is a widely used marker that consists in a single circular molecule inherited only by the maternal lineage. This type of inheritance and its high mutational rate, superior to the one found in nuclear DNA, and the absence of recombination are some of the advantages of this molecular marker in the study of the structure and demographic history of populations (Frankham *et al.*, 2004).

Although estimating recent demographic parameters based only in mitochondrial DNA is a controversial procedure, comparative studies also using this molecular marker provide qualitative signals of population changes, allowing for a more precise allocation of conservation efforts. So, its best results are obtained when combined with another molecular marker, preferably of nuclear origin (Moritz, 1994), like microsatellites.

Microsatellites are single locus co-dominant markers characterized by high mutational rates providing a finer scale to address ecological questions (Selkoe *et al.*, 2006). Technically, their small size makes them more resilient and easily amplified from degraded samples and can be combined in the genotyping process to provide fast and inexpensive replicated samples of the genome (Selkoe *et al.*, 2006). Like that, several microsatellite *loci*, provide a wider sample from the genome, diminishing sampling error while increasing statistical power (Selkoe *et al.*, 2006). Although the process to obtain microsatellite primers is complex and species-specific, cross-species amplification is possible with careful PCR conditions optimization (Salgueiro *et al.*, 2003; Mesquita *et al.*, 2005; Sousa *et al.*, 2007; Henriques *et al.*, 2007; Sousa *et al.*, in press).

In Conservation Genetics, one of the most frequent uses of both markers described above, mitochondrial DNA and microsatellites, is the definition of "Evolutionary Significant Units" (ESUs; *sensu* Moritz) and "Management Units" (MUs; *sensu* Moritz), respectively.

1.3.2 Conservation Units

An area in which Conservation Genetics has a great influence is the delineation of appropriate units for conservation, as conservation decisions often rely on the determination of species boundaries, a subject of Evolutionary and Systematic Biology (De Salle & Amato, 2004).

Two types of genetic units have been devised in conservation genetics: "Evolutionary Significant Units" (ESUs; *sensu* Moritz) and "Management Units" (MUs; *sensu* Moritz) (Moritz, 1994). ESU's separate reciprocally monophyletic populations for mtDNA and exhibit a significant divergence in nuclear *loci*, definition that can be applied to a population with speciation potential which represent the most important elements of intra-specific diversity and should be preserved as independent units (Moritz, 1994). On another hand, MUs represent different populations with a more recent ancestral between which some degree of genetic flow might still occur and that reveal significant divergence in allelic frequencies. The definition of these units is a matter of great importance as it is an extremely useful tool to define priority populations

for conservation. Nevertheless, its use has been criticized, in particular, when ESUs are defined solely on neutral molecular markers that ignore adaptive differences (Frankham *et al.*, 2004), meaning that populations with high degree of genetic flow, but with significant adaptive differences, would be treated as a single unit when they might require a separate management. The opposite might also occur: populations with low levels of genetic flow and high genetic differentiation, but with no adaptive differences, would be considered as distinct units when they might benefit from a single management strategy (Moritz, 1994). Another problem with Moritz definition of ESU resides in its dichotomous nature of "ESU or not". In order to circumvent this, Crandall *et al.*, (2000) proposed a system of discerning populations units based on eight categories of population distinctiveness. Depending on the magnitude of distinctiveness, each population or group of populations is assigned to a particular category that has specific management recommendations based on recent or historical exchangeability. Despite its limitations, in this work, we will use Moritz definition of ESU as it is the most commonly applied and the one that as been used to address conservation studies in Iberian cyprinids.

1.3.3 Conservation studies in endangered Iberian cyprinids

Several conservation genetic studies have already been made in threatened Iberian Cyprinids from the South of the Peninsula revealing genetic signatures of low number of effectives and/or low genetic variability levels caused by population fragmentation and annual collapses due to heterogeneous hydrological conditions (Mesquita *et al.*, 2005; Henriques, 2007; Sousa *et al.*, 2007; Sousa *et al.*, in press).

A study on the *S. aradensis* has clearly revealed low levels of genetic variability, evidences of demographic fluctuations and genetic vulnerability trough inbreeding in several populations (Mesquita *et al.*, 2005). The obtained results suggested the definition of four ESU's and the recommendation of conservation measures focused in the Arade populations (Arade and Boina) that presented the highest levels of genetic variability (Mesquita *et al.*, 2005).

Another study, on *S. torgalensis*, revealed that although this species does not incur in immediate extinction, it does present low levels of genetic variability, turning it

vulnerable to stochastic events and genetic erosion. In this case, results permitted the delineation of two insipient MU's within the Mira drainage (Henriques, 2007).

Anaecypris hispanica, the most threatened cyprinid species in the Iberian Peninsula, is just distributed in the Guadiana basin and its populations are severely fragmented. Through the analysis of the mitochondrial cytochrome *b* gene sampled populations were grouped in three ESUs (Alves *et al.*, 2001). Later on, Salgueiro *et al*. (2003) proposed, based on microsatellite data, that all studied populations corresponded to different MUs and, although populations were in decline, they're not as genetically impoverished as previously though, an example that not all endangered populations have low levels of genetic variability.

Concerning the *Iberochondrostoma* genus, a study based on the mitochondrial cytochrome *b* gene and the nuclear beta-actin gene of *I. lusitanicum*, revealed the existence of 3 ESUs: i) Sado drainage, ii) Tagus drainage and iii) Lagoa de Albufeira (Robalo *et al*., 2007). Additionally, Sousa *et al*. (2007) in a more recent study based on cytochrome *b* gene and microsatellite *loci* defined two ESUs and showed that *I. lusitanicum* suffered a population collapse and a subsequent increase in extinction risk for the next few decades. A similar work in *I. almacai*, also supported the existence of a population collapse with the division of the species in two MU's (one comprising the Mira drainage and other comprising the Arade and Bensafrim drainages) being recommended that actions should be taken to increase the amount of gene flow between populations in the same MU (Sousa *et al.*, in press).

Of the twenty cyprinid species that can be found in Portugal, five are classified as "Critically Endangered". From the remaining fifteen species, six are classified as "Threatened", two as "Vulnerable", two as "Nearly Threatened" and five as "Least Concern" (Cabral *et al*., 2005). Even though there are other cyprinid species that are more endangered than *I. lemmingii*, it current status of "Threatened" is still reason for some concern. Also, as no studies have yet focused on Portuguese populations it is important to understand if the same patterns of genetic impoverishment and population structure found in other cyprinid species are also present in *I. lemmingii*.

All species considered in the previously mentioned conservation studies share the same ecological constraints as *I. lemmingii* and some are even sympatric in some drainagens. This is of great importance because the understanding of the genetic condition of sympatric species permits a wider perspective and interpretation of the genetic patterns for the species under study. Also, studies of comparative

phylogeography allow for the determination of areas of high endemism and evolutionary potential permitting a better conservation management thus, an improved protection of the targeted species.

1.4 Objectives

As currently there is no genetic data at the population level on *I. lemmingii*, the primary goal of this thesis is to determine its genetic variability and population structure in order to create an adequate conservation strategy for this species, currently classified as "Vulnerable". For the purpose, we aim to (1) determine the genetic variability of each sample, (2) describe patterns of genetic variability between rivers and basins; (3) detect and date potential demographic events, namely bottlenecks and expansions; (4) obtain results that can be used as a base for future works of temporal comparisons and (5) discuss the implications of the results on the conservation of *I. lemmingii* and suggest conservation measures for the most threatened populations.

2. METHODS

2.1 Sampling

A total of 187 individuals were captured by electrofishing in six different locations: 51 in Caia River and 45 in Foupana River (Guadiana Basin), 42 in Quarteira River (Quarteira Basin), 34 in Aravil River (Tejo Basin), and 15 in Almargem River (Almargem Basin) (Figure 2). Individuals were anesthetized with MS222, fin clipped, photographed and returned to the water after recovery. Clips from pelvic fins were preserved in 100% ethanol at 4ºC and genomic DNA was extracted following a standard phenol-chloroform protocol adapted from Sambrook *et al*. (1989) and stored at -20 ºC. DNA concentration was measured with a NanoDrop Spectrophotometer ND-1000 v3.2.1 and standardized to 100 ng/µl per sample.

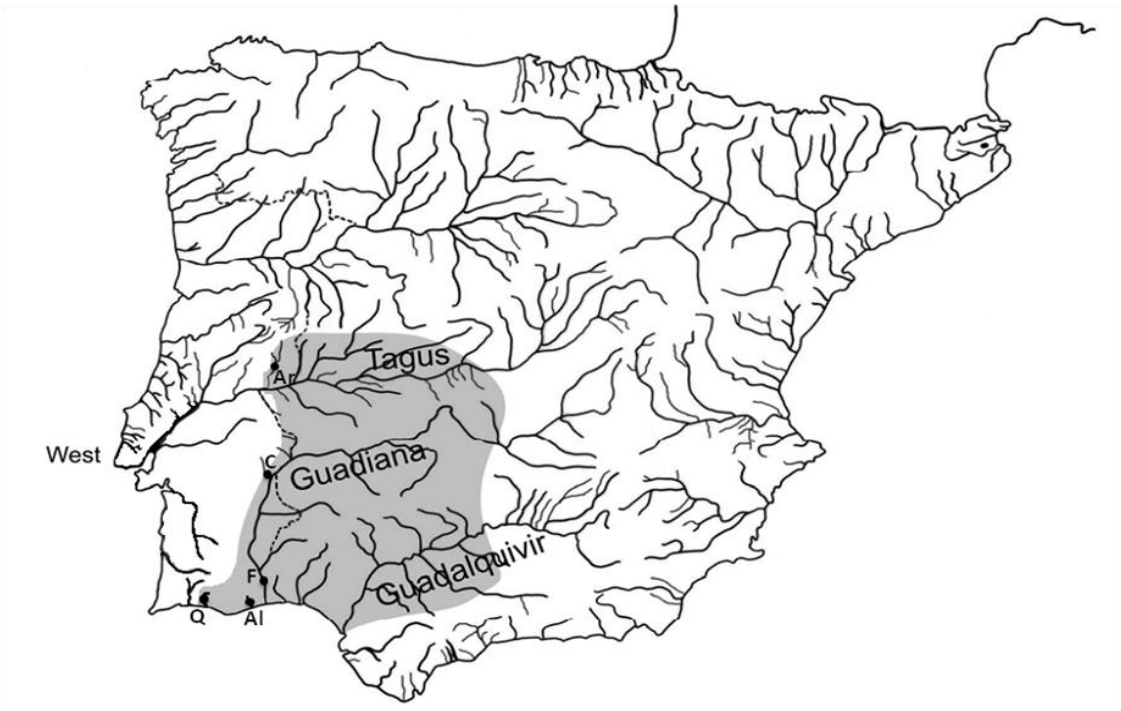

Figure 2. Iberian Peninsula map with sampling locations and *I.lemmingii* distribution. Q: Quarteira, A: Almargem, C: Caia, Ar: Aravil. Adapted from Robalo *et al.* (2008).

2.2 Laboratory Procedures

2.2.1 Mitochondrial DNA

A total of 63 individuals were sequenced for the cytochrome *b* mitochondrial gene: 14 individuals from Foupana, 13 from Caia, 12 from Quarteira, 13 from Aravil and 11 from Almargem.

The cytochrome *b* mitochondrial gene was amplified using the primers LCB1–5'-AATGACTTGAAGAACCACCGT-3' (Brito *et al.*, 1997) and HA–5'-CAAC GATCTCCGGTTTACAAGAC-3' (Schmidt & Gold, 1993).

Each 25-µL PCR reaction contained 150 ng DNA, 200 µM of each primer, 400 µM dNTP's, 2,50 mM $MgCl^2$ and 0,1µl of Taq Polymerase in 1×NH4 reaction buffer. The PCR reactions were conducted on a Bio-Rad® thermal cycler using the profile 94 °C for 1 min, 30 cycles of 94 °C for 1 min 54 °C for 1 min, 72 °C for 2 min and a final extension step of 72 °C for 3 min.

Successful amplification of PCR reactions was confirmed through electrophoresis in a 1% (agarose) gel stained with ethidium bromide and posterior visualization under ultra-violet light with a DC290 Kodak digital camera.

Purification of PCR Products was accomplished through an enzymatic clean-up process under the following conditions: 0,5µl of Exonuclease *I* (EXO I), 1µl Shrimp Alkaline Phosphatase (SAP) and 2,5µl of SAP Buffer 10× for each 20µl of PCR product at 37ºC for 30min, 80ºC for 15min and 12ºC for 5min.

Purified products were confirmed through electrophoresis as described before and then sequenced in both directions (forward and reverse in an ABI-PRISM genetic analyzer, Macrogen®). All obtained sequences were aligned with SEQUENCHER version 4.1 software (Gene Codes Corporation) and its homology was verified trough BLAST in the GENBANK database.

2.2.2 Microsatellites

No species-specific microsatellite primers for *I. lemmingii* exist in the literature, therefore, it was necessary to perform a cross-species amplification strategy with primers available for other cyprinid species, such as: *Luxilus cornutus* (Turner *et al.*, 2004); *Squalius aradensis* (Mesquita *et al.*, 2003); *Squalius alburnoides* (Pala &

Coelho, 2005); *Leucisus cephalus* (Larno *et al.*, 2005; Vyskocilová *et al.*, 2007); *Anaecypris hispanica* (Salgueiro *et al.*, 2003) and *Rhodeus sericesus* (Dawson *et al.*, 2003).

Optimization of PCR reactions was conducted in 12-µL PCR reactions containing 100 ng DNA, 200 µM of each primer (forward and reverse), 400 µM dNTP's, 2,50 mM $MgCl^2$ and 0,1µl of Taq Polymerase in 1×NH4 reaction buffer on a gradient Bio-Rad® thermal cycler. All primer pairs were tested under the following PCR conditions: 95 °C for 15 min, 94 °C for 30 s, annealing temperatures (Ta) from 50 °C to 60 °C in 2 °C intervals for 1 min 30 s were tested, 72 °C for 1 min for 30 cycles and 72 °C for 10 min (Mesquita *et al*., 2003). Some of the primers were also tested using alternative PCR conditions: an initial denaturation at 95 ºC for 30 s, followed by three series: 95 ºC for 30 s, 60 ºC (Ta) for 40 s, 72 ºC for 30 s (5 cycles); 95 ºC for 30s, 55 ºC (Ta) for 40 s, 72 ºC for 35 s (10 cycles); 95 ºC for 30 s, 50 ºC (Ta) for 40 s, 72 ºC for 35 s (25 cycles) and a final extension step at 72 ºC for 5 min (Salgueiro *et al*., 2003) (see Table I).

Successful amplification of PCR reactions was confirmed through electrophoresis in a 2% (agarose) gel stained with ethidium bromide and posterior visualization under ultra-violet light with a DC290 Kodak digital camera.

All tested primers that resulted in successful amplifications in *I. lemmingii* and correspondent annealing temperatures, are presented in Table I.

Table I: List of primers that amplified for *I.lemingii*, respective source and successful annealing temperatures

Microsatellites	Anealing Temperature (ºC)	Target Species	Authours
N2F11a	58	*Squalius aradensis*	Mesquita *et al.* (2003)
N4C8	60	*Squalius aradensis*	Mesquita *et al.* (2003)
N7G5	58	*Squalius aradensis*	Mesquita *et al.* (2003)
N7J4	58	*Squalius aradensis*	Mesquita *et al.* (2003)
N7K4	58	*Squalius aradensis*	Mesquita *et al.* (2003)
N7M11	56; 58	*Squalius aradensis*	Mesquita *et al.* (2003)
E1G6	58	*Squalius alburnoides*	Pala & Coelho (2005)
E2F8	50; 52; 54; 56	*Squalius alburnoides*	Pala & Coelho (2005)
LCO1	58	*Luxilus cornutus*	Turner *et al.* (2004)
LCO3	60	*Luxilus cornutus*	Turner *et al.* (2004)
LCO4	58	*Luxilus cornutus*	Turner *et al.* (2004)
LCO5	58	*Luxilus cornutus*	Turner *et al.* (2004)
IV04	T; 54	*Anaecypris hispanica*	Salgueiro *et al.*(2003)
X44	T	*Anaecypris hispanica*	Salgueiro *et al.*(2003)
XIII40	T	*Anaecypris hispanica*	Salgueiro *et al.*(2003)
Lce A149	T; 54	*Leuciscus cephalus*	Vyskocilová *et al.*(2007)
Lce DT	T; 54	*Leuciscus cephalus*	Vyskocilová *et al.*(2007)
Lce 100	T; 54	*Leuciscus cephalus*	Vyskocilová *et al.*(2007)
LC128	T; 54	*Leuciscus cephalus*	Larno *et al.* (2005)
Rser7	T; 58; 54	*Rhodeus sericesus*	Dawson *et al.* (2003)

T: Touchdown protocol by Salgueiro *et al.* (2003). Numbers represent annealing temperatures that permitted a successful amplification of respective primers under PCR conditions according to Mesquita *et al.* (2003).

Based on their amplification conditions, allelic sizes and non-complementarity between different primer pairs, 7 microsatellite loci were successfully grouped in 2 multiplex sets. The first multiplex set contained LCO5-$_{PET}$, LCO4-$_{VIC}$ primers (*Luxilus cornutus*; Turner *et al.*, 2004) and E1G6-$_{NED}$ (*Squalius alburnoides*; Pala & Coelho, 2005), while for the second multiplex N7G5-$_{VIC}$, N4C8-$_{FAM}$, N7K4-$_{NED}$ (*Squalius aradensis*; Mesquita *et al.*, 2003) and LCO1-$_{VIC}$ primers (*Luxilus cornutus*; Turner *et al.*, 2004) were grouped (see Table II in Results).

The PCR conditions for both multiplex primer sets were those used by Mesquita *et al.* (2003) with an annealing temperature of 58ºC in order to obtain successful amplification of all primers in the same mix and, an inferior number of cycles (diminished to 23) for a better resolution and visualisation of the microsatellite peaks.

PCR products were then diluted 1:20 in ultra pure water and 1µl of the dilution was suspended in 15µl of Formamide Hi-Di and 0,15µl of LIZ 500®. Samples were then

genotyped in an ABI PRISM – 310 GENETIC ANALYZER® after being previously denaturated for 3min 30s at 95ºC. Allele sizes were determined using the software GeneMapper 3.1. (Applied Byosistems®).

2.3 Statistical Analysis

2.3.1 Genetic diversity

Mitochondrial DNA

In order to explore mitochondrial DNA genetic diversity, the following measures were taken into account:

1. Number of haplotypes and private haplotypes in the sample, inferred with ARLEQUIN 3.11 (Excoffier *et al.*, 2005).
2. Haplotypic diversity (*h)*: probability that two randomly sampled haplotypes are different, calculated with ARLEQUIN 3.11 (Excoffier *et al.*, 2005).
3. Nucleotide diversity (π): the probability that two randomly chosen homologous sites are different. It is equivalent to the gene diversity at the nucleotide level for DNA data with ARLEQUIN 3.11 (Excoffier *et al.*, 2005).

Microsatellites

All used microsatellite loci were tested for null alleles, stuttering and large allele drop out using the program MICRO-CHECKER version 2.2.03 (Van Oosterhout *et al.*, 2004). The application uses a Monte Carlo simulation (bootstrap) method to generate expected homozygote and heterozygote allele size difference frequencies.

Microsatellite loci were also checked for evidences of linkage disequilibrium among them using ARLEQUIN 3.11 (Excoffier *et al.*, 2005). As the gametic phase is unknown the procedure for testing the significance of the association between pairs of loci is based on a likelihood ratio test, where the likelihood of the sample evaluated under the hypothesis of no association between loci (linkage equilibrium) is compared to the likelihood of the sample when association is allowed (Slatkin & Excoffier, 1996).

The significance of the observed likelihood ratio is found by computing the null distribution of this ratio under the hypothesis of linkage equilibrium, using a permutation procedure. Sequential Bonferroni correction was applied when multiple tests where performed.

After the definite choice of microsatellites suitable for the study, several standard genetic diversity measures were inferred:

1. Number of alleles by locus (A) with ARLEQUIN 3.11 (Excoffier *et al.*, 2005).
2. Allelic richness: average number of alleles by locus (A_r) corrected to take into account population size, through a rarefaction procedure as implemented in the software HP-RARE 1.0 (Kalinoswki, 2005).
3. Number of private alleles (PA): the number of alleles that can only be found in a single population, also implemented in HP-RARE 1.0 (Kalinoswki, 2005).
4. Observed heterozygosity: the proportion of heterozygous individuals that can be found in a single population with ARLEQUIN 3.11 (Excoffier *et al.*, 2005).
5. Expected heterozygosity: prediction of the proportion of heterozygous individuals in a single population based on the known allele frequencies of a sample. The observed allelic frequencies can be tested against the expected frequencies to see if the populations are in Hardy-Weinberg equilibrium, which states that allelic frequencies will remain constant trough time in a randomly mating large population with no immigration, mutation or natural selection.
6. Wright's fixation index (1951), or inbreeding coefficient (F_{IS}): measures the reduction in heterozygosity based on the relation between the observed heterozygous with the proportion of heterozygous that would occur in Hardy-Weinberg equilibrium. The direct consequence of inbreeding is the increase of homozygosity and therefore positive values of F_{IS}. FSTAT 2.9.3.2 (Goudet, 2001) was used to infer this parameter that varies between -1 (heterozygous excess) and +1 (homozygous excess).

2.3.2 Population Structure

Mitochondrial DNA

Phylogeographic relationships between individuals were first explored by a haplotypic network based on the cytochrome *b* sequences. These relationships are usually described as a tree, though true evolutionary relationships are reticulate rather than strictly dichotomous tree-like. For example, at the population level, biologists have always thought of relationships among individuals as being reticulate, simply because the primary historical pattern within a sexually reproducing species is inter-breeding (Morrison, 2005). Therefore the multitude of plausible trees to express relationships among individuals might be best expressed by a network, which displays alternative potential evolutionary paths (Bandelt *et al.*, 1999). NETWORK 4.5 software (Bandelt *et al.*, 1999) groups individuals based on their genetic distance. The software uses a Maximum Parsimony (MP) approach creating several trees and choosing the least complex. The Median-Joining algorithm was used with the default parameters as recommended for multiple state data (Bandelt *et al.*, 1999). This method creates the minimum spanning trees all combined in a single reticulate network and subsequently consensus sequences are added in order to achieve parsimony. These added sequences are median vectors and can be biologically interpreted as unsampled sequences or extinct ancestral sequences (Bandelt *et al.*, 1999).

Population differentiation was computed using pairwise Φ_{ST} (Nei, 1973) as implemented in ARLEQUIN 3.11 (Excoffier *et al.*, 2005) which is similar to classical F_{ST} (Wrigth, 1951) but takes into account the genetic distance among haplotypes and not only haplotipic frequencies (Excoffier *et al.*, 2005). For that, it was necessary to find the most adequate mutation model for the data using MODELTEST 3.7 software (Posada & Crandall, 1998). The TrN+I (Tamura-Nei; with I representing invariable sites) model was chosen to compute pairwise Φ_{ST} with ARLEQUIN 3.11 (Excoffier *et al.*, 2005). The Δ_i is inferior to 2 which means that the model still has substantial support and from all models that were available in the ARLEQUIN 3.11 software (Excoffier *et al.*, 2005), was the highest ranked. Pairwise Φ_{ST} were calculated taking into consideration the best fitted model, 10 000 permutations and significance levels of 0,05 were posteriorly adjusted by Bonferroni correction.

In order to access the population structure without prior knowledge of their sampling site origin the BAPS 4.1 software (Corander & Tang, 2007) was used. To run this programme the user must provide the number of clusters that wants to test and the software uses a stochastic optimization procedure to find the clustering solution with the highest "marginal likelihood" (Corander & Tang, 2007). The number of K was set between 1 and 6 and the analysis was performed 10 times. The partition that presented the highest value of marginal likelihood was recorded and analyzed.

An Analysis of Molecular Variance (AMOVA) as implemented in ARLEQUIN 3.11 (Excoffier *et al.,* 2005) was also performed to test the significance of the different structures presented in the previous analyses, i.e., the structure evidenced by both geographic sampling as well as the structure highlighted by BAPS. By defining groups of populations, the user defines a particular hierarchical genetic structure that will be tested. The haplotype differences were used as genetic distance, and a total of 10 000 permutations were used and the significance level was set to 0,05.

Microsatellites

In a preliminary approach, the relationship between individuals based on microsatellite loci data was explored through a Correspondence Factor Analysis (FCA) implemented in GENETIX 4.2 (Belkhir *et al.,* 2000). This analysis allows to observe the relationship between all individual genotypes in 2D/3D scatter plots and infer possible closely related groups, i.e., samples. This methodology also allows for the identification of "outliers", i.e., individuals that do not seem to be related to any cloud of points in the chart. These individuals should then be target of careful investigation in order to understand why they seem to be so far related with all the other sampled individuals. One of the reasons for this behavior might bee the misidentification of individuals from other species or the presence of hybrids.

The software POPULATIONS 1.2.30 (Langella, 2002) was used to create distance trees for individuals with the genetic distance D_{AS} (Jin & Chakraborty, 1994) which calculates shared allele distances, and for samples with D_A (Nei *et al.*, 1983) and D_C (Cavalli-Sforza & Edwards, 1967) genetic distances. All distances were used in

conjunction with the "Neighbour Joining" algorithm (Takezaki & Nei, 1996) and all produced trees were manipulated with MEGA 3.1 software (Kumar *et al.*, 2004).

The degree of differentiation between samples was then explored using Wrigth's F statistics (1951) using the algorithm proposed by Weir & Cockerham (1984) as implemented in ARLEQUIN 3.11 (Excoffier *et al.,* 2005). It was also calculated the value for R_{ST}, a derivation of F_{ST} specific for microsatellite data that takes into account the size of the alleles and the SMM (stepwise mutation model) according to Rousset (1996).

Pairwise F_{ST} and R_{ST} were calculated with ARLEQUIN 3.11 (Excoffier *et al.*, 2005) with a significance level of 0.05 after 10 000 permutations. Significance levels were then adjusted with Bonferroni correction.

The software STRUCTURE 2.2 (Pritchard *et al.,* 2000) was used to infer on population structure without any *a priori* information on the origin of the individuals. The program implements a model-based clustering method for inferring population structure using genotype data consisting of unlinked markers. First, is assumed a model in which there are K populations (where K may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned (probabilistically) to populations, or jointly to two or more populations if their genotypes indicate that they are admixed. It is assumed that within populations, the loci are at Hardy-Weinberg equilibrium, and linkage equilibrium, and individuals are assigned in to different groups in order to maintain the deviations to these equilibriums to a minimum (Pritchard *et al.*, 2000). The software was used under the "Admixture" model with allelic frequencies independent, and with values of K set between 1 and 10. For each K, 20 simulations were performed with a burning period of 50 000, followed by 100 000 Markov steps (MCMC). STRUCTURE 2.2 (Pritchard *et al.*, 2000) software outputs the average log likelihood of the data for each value of K at each step of the Markov Chain Monte Carlo (MCMC). In order to choose which value of K best suited the data, the protocol defined by Evanno *et al*. (2005) was followed. This method is based on an ad hoc statistic based on the rate of change of the log likelihood of the data between successive values of K (Evanno *et al.,* 2005).

A hierarchical analysis of population subdivision was performed using the Analysis of Molecular Variance (AMOVA; Excoffier *et al.*, 1992) implemented in ARLEQUIN 3.11 (Excoffier *et al.*, 2005). This analysis was performed to test for the

significance of the different structures defined by geographic sampling and STRUCTURE. The allelic frequencies were used as genetic distance in each sample and a total of 10 000 permutations were used and the significance level was set to 0.05.

2.3.3 Demographic History

Mitochondrial DNA

Mitochondrial DNA was also used to access demographic history with the use of Tajima's *D* test (Tajima, 1989), Fu's *Fs* (Fu, 1997) and mismatch distribution as implemented in ARLEQUIN 3.11 (Excoffier *et al.,* 2005). Although Tajima's *D* test and Fu's *Fs* were created to test for the presence of selection they are sensitive to demographic patterns and can be used to search for signatures of bottlenecks and expansion events. These tests are based in the infinite-site model, without recombination (Excoffier *et al.,* 2005). Tajima's *D* and Fu's *Fs* statistics tends to present negative values when there is an excess of recent mutations and therefore of rare alleles, making it possible to interpret large negative values as an evidence of expansion events. Tajima's *D* test looks at the expectation of the average number of nucleotide differences among DNA sequences randomly sampled from a population and can also be sensitive to bottleneck effects (Ramirez-Soriano *et al.*, 2008). Finally, in order to detect expansion events, mismatch distributions are a graphic evaluation of the observed number of nucleotide differences between each pair of sample haplotypes. This will result in a multimodal distribution for a population in demographic equilibrium and in a unimodal distribution for populations that incurred in expansion demographic events. ARLEQUIN's output of mismatch distribution analysis also includes the parameters of the expansion events with τ representing the expansion time, which is the product of the mutation rate for the haplotype (u) with the time since the expansion event in generations (t). The mutation rate for the haplotype is given by $2\mu k$ with μ representing the mutation rate per million years (MA) and k the number of nucleotides in the sequence. It was considered a mutation rate of 1.05% per MA as defined by Doadrio & Carmona (2004) and a generation time of two years (Fernandez-Delgado & Herrera, 1995).

Microsatellites

Microsatellite data was processed with BOTTLENECK software (Piry *et al.*, 1999) in order to explore the demographic history of *I. lemmingii*. This software allows determining if there were bottleneck effects in the studied populations by calculating deviations in the number of expected heterozygous based in the allelic frequencies of the used *loci* (Piry *et al.*, 1999). A population that went trough a bottleneck from which resulted a diminished effective population size will experience a decrease in the number of heterozygotic individuals and also in the number of alleles since the rare alleles are easily lost. As the rare alleles have little contribution to the total diversity, the number of alleles will decrease at a faster rate when compared to the amount of heterozygosity (Piry *et al.*, 1999). By comparing the observed number of heterozygous individuals with the expected number of heterozygous individuals taking into account the number of alleles in each locus, sample size and assuming mutation-drift equilibrium it is possible for the software to detect signs of the occurrence of bottlenecks (Piry *et al.*, 1999). All possible models implemented in BOTTLENECK were tested (IAM, SMM and TPM). Two TPM models were tested: one with 90% SMM and 10% IAM and other with 70% SMM and 30% IAM.

3. RESULTS

3.1 Genetic diversity

3.1.1 Mitochondrial DNA

A total of 63 individuals, including representatives of all 5 samples, were sequenced for the cytochrome *b* mitochondrial gene. The length of the obtained fragment was of 1013pb with a mean composition of 26.8% A, 16.4% G, 29.5% T, 27.4% C.
A total of 18 different haplotypes were found in the 63 individuals surveyed. (Table II)

Table II. Variability measures for the mitochondrial cytochrome *b* gene.

	Caia	Foupana	Quarteira	Almargem	Aravil	Total
N	13	14	12	11	13	63
N_H	5	7	1	5	4	18
N_P	3	5	0	3	3	14
N_S	10	9	0	6	11	23
h	0.698	0.816	0	0.562	0.675	0.877
π	0.0032	0.0026	0	0.0014	0.0046	0.0039

N, number of individuals; N_H, number of haplotypes found; N_P, number of private haplotypes; N_S, number of substitutions; *h*, haplotidic diversity; π: nucleotidic diversity.

Most haplotypes were shared between samples but 18 were found in specific localities. A total haplotidic diversity of 0.877 indicates an overall high level of diversity in *I. lemmingii*. However, observed haplotipic diversity was very distinctive across samples ranging between 0 in Quarteira (only one haplotype was found) and 0,816 in Foupana, which has to be taken into consideration when interpreting overall values. In contrast, nucleotide diversity within each sample was more homogeneous and generally low (overall π=0.0039), ranging from 0 in Quarteira to 0.0046 in Aravil.

Excluding Quarteira (the sample with less genetic diversity, only one haplotype), Almargem presented the lowest values of both haplotipic and nucleotidic diversity and Foupana the higher value of haplotypic diversity but a low value of nucleotide diversity.

3.1.2 Microsatellites

No evidence for the presence of null alleles, stuttering or large allele drop out was found in the 7 loci analysed. Also, no consistent evidences of linkage disequilibrium were found after Bonferroni corrections (P<0.0006) as only two different pairs of *loci* presented significant values in punctual populations (Table IV).

All loci were polymorphic with a total of 100 alleles detected across all loci and samples. LCO5 was the less polymorphic loci with 4 alleles ranging from 152 to 166bp while LCO1 presented the highest rate of polymorphism with 39 alleles ranging from 184 to 280bp (Table III).

Table III. Microsatellites loci characterization: repeat motif, primer sequences, flourochrome labeling information, number of observed alleles and respective allelic range.

Microsatellite loci	Repeat motif	Primer Sequence	Nº alleles	Allelic size range (bp)
LCO5 (M1)	$(CAGA)_n(CA)_n$	F - 5' TTACACAGCCAAGACTATGT 3' R - 5' CAAGTGATTTTGCTTACTGC 3' *PET	5	152 – 166
E1G6 (M1)	$(GT)_n$	F - 5' CACAGGCATTCATACATATCAAGC 3' R - 5' GTCCGTGTGAGACTGAACCCTTAC 3' NED	11	148 – 200
LCO4 (M1)	$(GT)_nATTTT(GT)_n(GA)_n$	F - 5' ATCAGGTCAGGGGTGTCACG 3' R - 5' TGTTTATTTGGGGTCTGTGT 3' VIC	18	232 – 262
SarN7G5 (M2)	$(TG)_nCG(TG)_nCGTGCG$ $(TG)_n(TG)_n$	F - 5' GAGCTTCAGCACCGAGGAC 3' R - 5' CTACATGACAGGCATCTGCAGTA 3' VIC	10	90 – 110
SarN4C8 (M2)	$(GT)_n$	F - 5' GAACAAAGATCAGTGAAGCACC 3' R - 5' ACGTCAGACTTCAGGCATCC 3' 6-FAM	3	101 – 117
SarN7K4 (M2)	$(TG)_n$	F - 5' CATGTTTCCACATCTGAGCTAAAA 3' R - 5' ACGAGCATCAGTATCCAGAGACAC 3' NED	14	150 – 180
LCO1 (M2)	$(GATA)_nGGCTA(GATA)_n$	F - 5' CACGGGACAATTTGGATGTTTTAT 3' R - 5' AGGGGGCAGCATACAAGAGACAAC 3' VIC	39	184 – 280

M1: Multiplex 1; M2: Multiplex 2

All loci presented polymorphism in all samples with the exception of N4C8, LCO5 and E1G6 that were respectively monomorphic for Caia/Almargem/Aravil, Almargem and Aravil (Table IV). The mean allelic richness across all loci was similar between populations; however, the calculated values per loci were quite divergent ranging from 1 in several samples to 16 for the LCO1 locus in the Aravil population. Quarteira was the population with more private alleles averaged across all *loci* (1.240) opposed to Caia with only 0.750. Averaged observed heterozygosity across loci ranged between 0.395 in Caia and 0.610 in Almargem and the averaged expected heterozygosity ranged between 0.410 for Caia and 0.610 for Almargem. The Aravil population was the only to present observed values of heterozygosity superior to those that were expected, although none of them was significant. Aravil was the only population to present a negative value of F_{IS} indicative of possible increase in heterozygous frequencies while Foupana population presented the highest value of F_{IS} (0.060)

Table IV. Microsatellite diversity measures.

		Caia	Foupana	Quarteira	Almargem	Aravil
LCO5	N	51	45	42	15	34
	A	3	2	3	1	2
	Ar	2.599	2.164	2.558	1.000	1.888
	P_A	0.275	1.164	0.213	0.000	0.000
	H_O	0.235	0.111	0.476	-	0.118
	H_E	0.246	0.106	0.403	-	0.112
	F_{IS}	0.042	-0.050	-0.185	-	-0.048
E1G6	N	51	46	45	19	34
	A	2	7	3	2	1
	Ar	1.275	4.742	2.117	2.000	1.000
	P_A	0.275	2.666	0.822	0.218	0.000
	H_O	0.020	0.267	0.048	0.143	-
	H_E	0.020	0.369*	0.093	0.138	-
	F_{IS}	0.000	0.312	0.491	-0.040	-
LCO4	N	51	46	45	19	34
	A	11	13	9	7	9
	Ar	8.873	10.348	6.930	7.000	6.633
	P_A	1.060	1.576	0.290	1.472	0.397
	H_O	0.902	0.800	0.524	0.643	0.882
	H_E	0.868*	0.847	0.704	0.778	0.777
	F_{IS}	-0.039	0.064	0.258	0.179	-0.138

SarN7G5	N	51	46	45	19	34
	A	4	7	8	5	6
	Ar	2.866	6.119	6.441	4.864	4.139
	P_A	0.162	0.159	1.025	0.430	0.300
	H_O	0.216	0.756	0.738	0.667	0.588
	H_E	0.201	0.738	0.742	0.582	0.517
	F_{IS}	-0.074	-0.020	0.006	-0.152	-0.141
SarN4C8	N	51	46	45	19	34
	A	1	2	5	1	1
	Ar	1.000	1.934	3.662	1.000	1.000
	P_A	0.000	0.003	1.731	0.000	0.000
	H_O	-	0.111	0.571	-	-
	H_E	-	0.145	0.506	-	-
	F_{IS}	1.000	0.236	-0.088	-	-
SarN7K4	N	51	46	45	19	34
	A	7	3	8	4	6
	Ar	5.054	2.996	5.818	3.933	4.820
	P_A	1.226	0.002	3.280	0.150	0.839
	H_O	0.510	0.568	0.595	0.800	0.559
	H_E	0.596	0.567	0.552	0.687	0.640
	F_{IS}	0.146	-0.002	-0.080	-0.171	0.128
LCO1	N	51	46	45	19	34
	A	17	21	9	12	22
	Ar	11.807	13.674	6.889	11.598	15.957
	P_A	2.269	4.246	1.310	3.133	6.409
	H_O	0.882	0.886	0.732	0.800	0.941
	H_E	0.904	0.922	0.789	0.899	0.951
	F_{IS}	0.024	0.040	0.092	0.113	0.011
ALL	N	51	46	45	19	34
	A	6.571	7.125	6.125	6.000	8.000
	Ar	4.780	6.000	4.920	4.490	5.060
	P_A	0.750	1.400	1.240	0.770	1.130
	H_O	0.395	0.454	0.508	0.610	0.544
	H_E	0.410	0.482	0.538	0.616	0.535
	F_{IS}	0.038	0.060	0.037	0.010	-0.031

N: number of individuals; A: Number of alleles; Ar: allelic diversity obtained through a rarefaction procedure; P_A: Private alleles; H_o: observed heterozigosity; H_e: expected heterozigosity; F_{is}: Fixation index. Significant deviations of H_O in relation to H_E are marked with * (*p*-value<0.0016 after Bonferroni corrections).

3.1 Population Structure

3.1.1 Mitochondrial DNA

The most common haplotype (H1) was found in a total of 19 individuals from Quarteira (N=12, which represent all sample) and Almargem (N=7) (Table V). The majority of the other haplotypes were unique to one sample except for one haplotype shared between Foupana and Almargem; another shared between Caia and Aravil and a third one shared between Caia and Foupana. The median-joining network (Figure 3) showed some degree of complexity with no clear geographical structure among haplotypes. A central group of haplotypes could be identified and two peripheral ones belonging to Aravil.

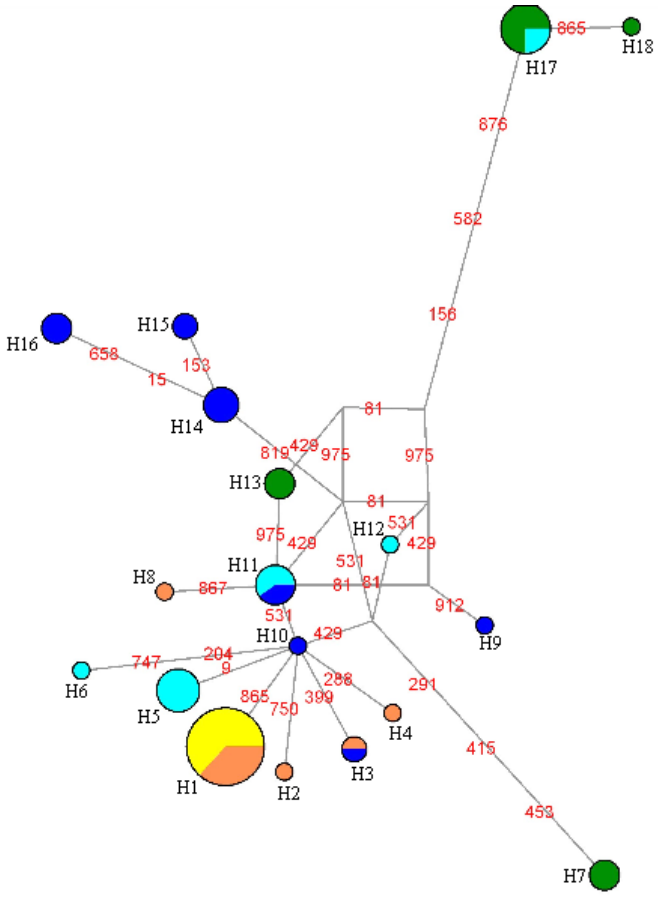

Figure 3. Median-joining Network of cytochrome *b* (mtDNA) sequences. The area of the circles is proportional to the frequencies of the haplotypes. Colours represent the geographic origin of haplotypes: Foupana (dark blue), Caia (light blue), Quarteira (yellow), Almargem (orange) and Aravil (green). Mutation positions are marked in red.

Table V. Haplotype nucleotide differences with respective sequence location and distribution of haplotypes in each sample.

	9	15	81	153	156	204	208	219	239	241	259	263	311	358	412	458	470	509	516	528	617	676	691	725	C	F	Ar	Al	Q
H1	G	C	G	A	A	A	A	A	A	G	G	A	A	G	G	A	A	A	A	C	T	C	A					7	12
H2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	G	-	-	-	-					1	
H3	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-		1			1	
H4	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-					1	
H5	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	6					
H6	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	G	-		G	-	-	-	-	1					
H7	-	-	-	-	-	-	G	-	A	A	G	-	-	-	-	-	-	-	G	-	-	-	-			3			
H8	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	G	T	-	-	-					1	
H9	-	-	A	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	G	-	-	G	-		1				
H10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-		1				
H11	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	G	-	-	-	-		3	2			
H12	-	-	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	G	-	-	-	-	1					
H13	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	G	-	-	-	G					3	
H14	-	-	-	-	-	-	-	-	-	A	-	G	-	-	-	-	G	G	-	-	-	-	-		4				
H15	-	-	-	G	-	-	-	-	-	A	-	G	-	-	-	-	G	G	-	-	-	-	-		2				
H16	-	T	-	-		-	-	-	-	A	-	G	-	A	-	-	G	G	-	-	-	-	-		3				
H17	-	-	A	-	G	-	-	-	-	A	-	G	A	-	-	-	-	G	-	C	-	G		2		6			
H18	-	-	A	-	G	-	-	-	-	A	-	G	A	-	-	-	-	C	-	C	-	G				1			

C: Caia; F: Foupana; Ar: Aravil; Al: Almargem; Q: Quarteira.

Φ_{ST} pairwise values ranged from 0.145 between Almargem and Quarteira to 0.642 between Foupana and Quarteira (Table VI). All values were highly significant with $P<0.05$ suggesting that each of the 5 samples analyzed represent genetically definable populations. Quarteira presented the highest pairwise Φ_{ST} values when compared with all other samples, which seems to point to this population as the most significantly divergent from the others. It is also worth noticing that the pairwise Φ_{ST} value between Caia (Guadiana basin) and Aravil (Tagus basin) (0.267) is among the lower ones and is very similar to that found between Caia and Foupana (0.264) which belong to the same river basin (Guadiana). All values remained significant after Bonferroni corrections ($P<0.005$) except for the comparison between Almargem and Quarteira.

From these results it is possible include the samples from Quarteira and Almargem in a single group and maintain all the remaining samples as differentiated populations.

Table VI. Φ_{ST} mtDNA pairwise comparisons. Φ_{ST} values below the diagonal and correspondent *P*-values above the diagonal.

	Caia	Foupana	Quarteira	Almargem	Aravil
Caia	–	0.0003*	0.0000*	0.0002*	0.007*
Foupana	0.264	–	0.0000*	0.0000*	0.0000*
Quarteira	0.457	0.642	–	0.0349	0.0000*
Almargem	0.236	0.459	0.145	–	0.0000*
Aravil	0.267	0.372	0.614	0.459	–

Significance of tests after Bonferroni correction, are marked with *.

The genetic structure of populations, without prior knowledge of their geographic origin, revealed 5 putative genetic groups (Figure 4) with a log(ml)=-298.2356. Three of the generated groups presented mixed constitutions with individuals from different geographic samples: one group included individuals from all sampling sites, except Quarteira; another was constituted by individuals from Foupana and Aravil; and a third one was composed by individuals from Quarteira and Almargem. The two other genetic groups showed no mixed origin of individuals: one included individuals merely from Foupana and the a fifth was solely constituted by three individuals from the Aravil sample. However, as the software tends to overestimate the amount of groups found (Latch *et al.,* 2006), this last group probably lacks biological meaning. Moreover, Quarteira sample presented individuals from a single genetic group. Despite a non-complementarity between genetic groups and geographic samples, this analysis revealed similarity between Almargem and Quarteira samples and a high differentiation of Foupana, Caia and Aravil as genetic populations despite some degree of admixture.

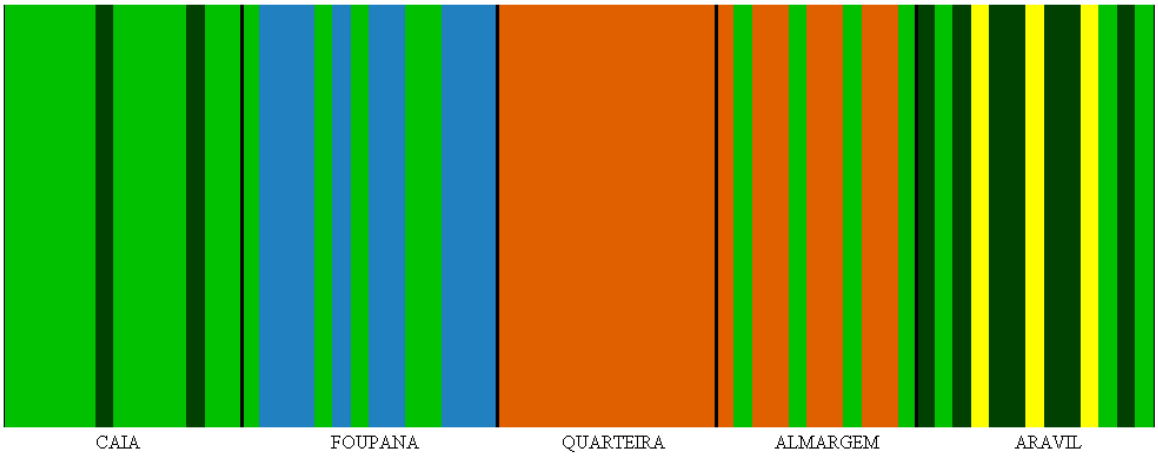

Figure 4. Results from BAPS 4.1 software. Geographic populations are separated with black lines and each vertical colored line corresponds to one individual and the group on which he was genetically placed. Each color corresponds to one genetic group.

The hierarchical gene diversity analysis (AMOVA) based on geographic basins (4 groups: Guadiana (Caia + Foupana) / Tagus (Aravil) / Quarteira / Almargem) and on the structure supported by BAPS and Φ_{ST} (4 groups: Quarteira + Almargem / Foupana / Aravil/Caia) failed to reveal a significant amount of variation explained among groups (P>0.05) (Table VII). As for the structure proposed with the results obtained with the Φ_{ST} and BAPS, the AMOVA analysis revealed the higher value for explained variance among groups (45.88%) and although this value was still non significant it presented the lowest P-value of both proposed structures (P=0.100). The non significance of the AMOVA analysis that was performed for the Φ_{ST} and BAPS might be related to the fact that the variation within populations was also very high and significant (55.07%, P<0.001).

Table VII. AMOVA results for mitochondrial DNA.

	BASINS				
Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Index
Among groups	3	43.226	0.431	19.300	F_{CT}=0.193
Among populations within groups	1	8.679	0.551***	24.710	F_{SC}=0.306
Within populations	58	72.441	1.249***	55.990	F_{ST}=0.44
Total	62	124.346	2.231	100	–
	BAPS				
Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Index
Among groups	2	39.405	0.646	28.060	F_{CT}=0.459
Among populations within groups	2	12.500	0.409***	17.730	F_{SC}=-0.017
Within populations	58	72.411	1.249***	54.210	F_{ST}=0.449
Total	62	124.346	2.304	100	–

On top: a basin like structure, on bottom: the previously proposed structure based on Φ_{ST} and BAPS 5.1 software (Pritchard *et al*, 2000). Significant values are marked with * (P<0.05), ** (P<0.01) or *** (P<0.001).

3.3.2 Microsatellites

The 3D Correspondence factor analysis performed using microsatellite data explained a total of 85.64% of all variation in the first three axes. The first axis alone explains 41.48% of the total variance and shows a spatial separation of the Quarteira population from the remaining populations (Figure 5). The second axis explains 28.20% of variation and allows for the distinction of two groups, the Foupana+Almagem group and the Caia+Aravil group. Finally, the third axis, explains 16.16% of the variation and provides some distinction of the populations inside the previously describe groups separating Foupana from Almargem and Caia from Aravil.

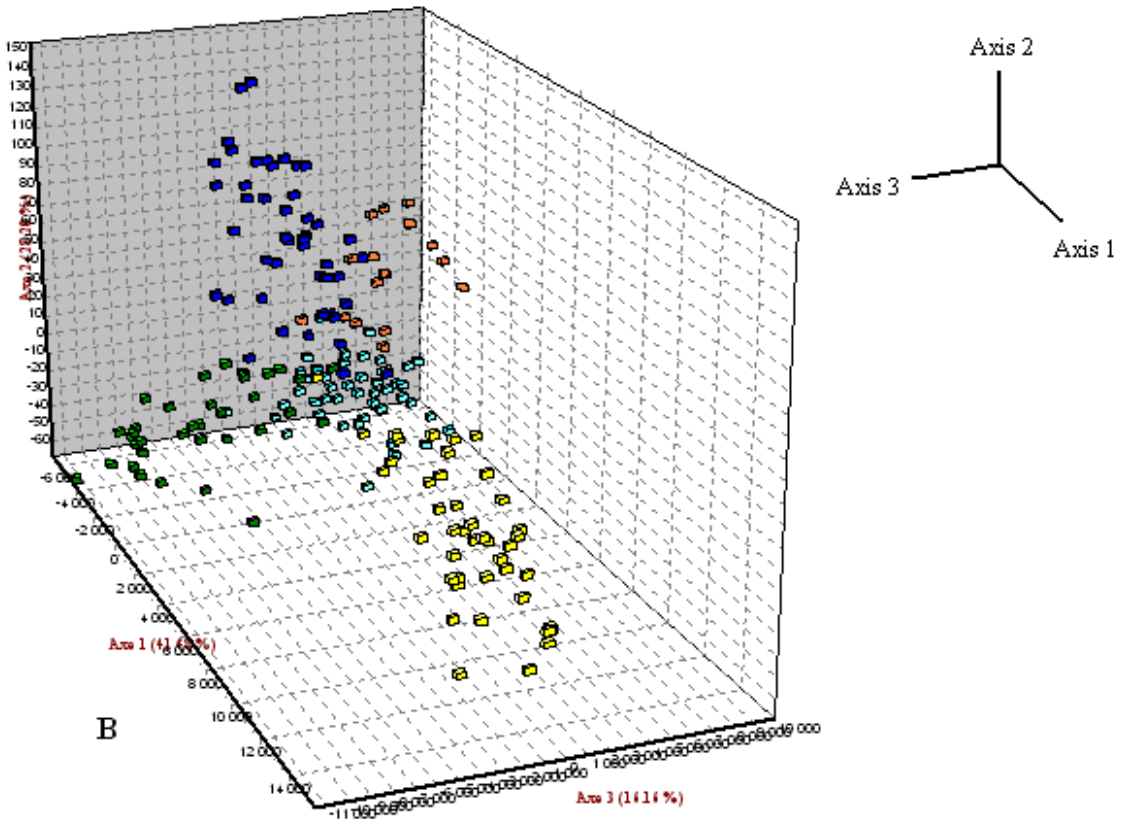

Figure 5. Correspondence factor analysis of *I. lemmingii*. Colors correspond to different sampling sites: Green (Aravil), Yellow (Quarteira), Light Blue (Caia), Dark Blue (Foupana), Orange (Almargem).

The reconstruction of the genetic distance tree for populations (Figure 6) held similar results with both D_A (Nei *et al.*, 1983) and D_C (Cavalli-Sforza & Edwards, 1967). Quarteira presents a more distant position from any other population reflected on its branch length, while Caia and Aravil populations seem to be the most closely related.

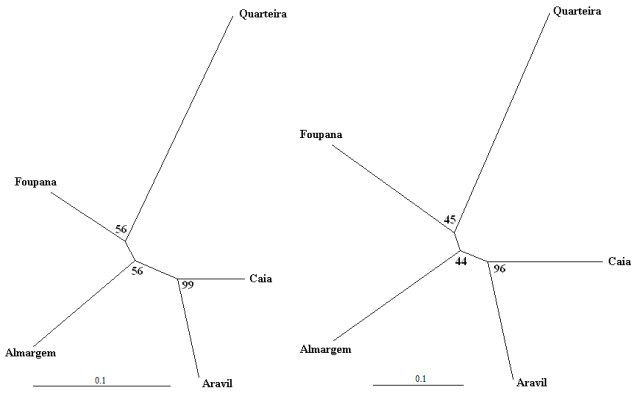

Figure 6. Reconstruction of the distance trees according to Nei *et al.* (1983) D_A distance on the left, and Cavalli-Sforza & Edwards (1967) D_C distance, on the right.

The reconstruction of the genetic distance tree for individuals with D_{AS} distance (Figure 7) revealed a clear clustering of Quarteira individuals in a distinct branch of the tree. Also, despite the inclusion of individuals from different geographic localities in the same branches, a slight tendency of individuals from Foupana and Caia to cluster in different branches of the tree could also be observed. Also, samples from Aravil seem to have more affinity to cluster with samples from Caia then with any other samples.

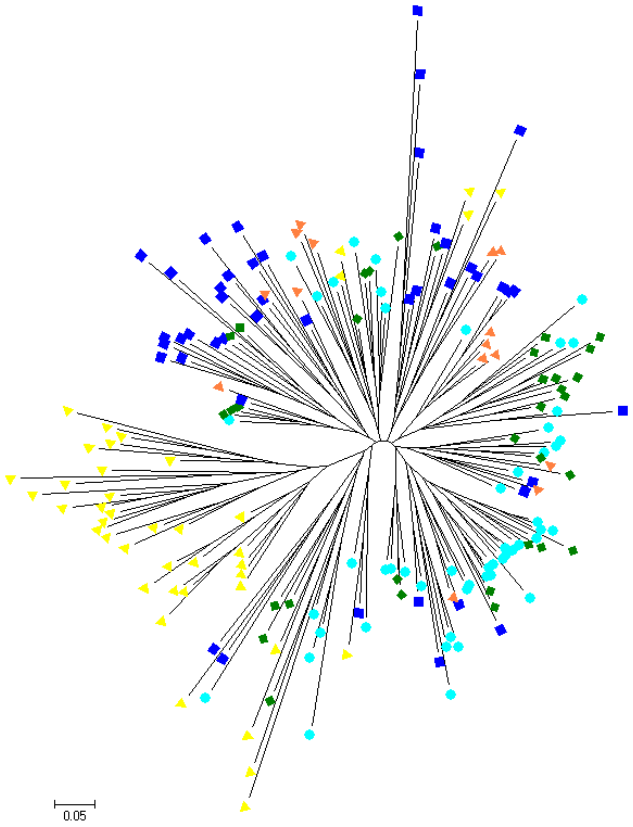

Figure 7. Tree of individuals calculated with D_{AS} genetic distance (Lin & Chakrabotti, 1993). ● Caia, ■ Foupana; ▼ Quarteira; ▲ Almargem; ◆ Aravil.

Wright's F statistic (Wright, 1951) and Rousset's R statistic (Rousset, 1996) seem to indicate an existence of a significant genetic structuring over all samples (average F_{ST}=0.102, P<0.001; average R_{ST}=0.224, P<0,001).

Pairwise F_{ST} values ranged from 0.030 between Foupana and Almargem to 0.187 between Caia and Quarteira (Table VIII). All values were significant, even after Bonferroni correction (P<0.005). The higher values of differentiation were found in pairwise comparisons between Quarteira and all other populations as obtained for comparisons based on mitochondrial gene. It is also interesting to notice that the value of F_{ST} between Foupana and Caia (from the same River Basin) is higher than the values

obtained between Foupana and Almargem and between Caia and Aravil (from different River basins).

Table VIII. F_{ST} microsatellite pairwise comparisons. F_{ST} values bellow diagonal and respective *p*-values above the diagonal.

	Caia	Foupana	Quarteira	Almargem	Aravil
Caia	–	0.000*	0.000*	0.000*	0.000*
Foupana	0.109	–	0.000*	0.001*	0.000*
Quarteira	0.187	0.106	–	0.000*	0.000*
Almargem	0.081	0.030	0.136	–	0.000*
Aravil	0.053	0.076	0.155	0.077	–

Significant values after Bonferroni corrections, are marked with *

Pairwise R_{ST} values ranged from 0.089 between Caia and Foupana to 0.383 between Quarteira and Aravil (Table IX). However, pairwise comparisons Almargem x Foupana; Aravil x Foupana and Aravil x Almargem were not significant and; after Bonferroni corrections (P<0.005), the comparison between Almargem and Caia also lost significance. Accordingly to these results Quarteira population is still the most divergent, presenting the greater values of pairwise R_{ST} but Foupana, Aravil and Almargem do not present significant population differentiation.

Table IX. R_{st} microsatellite pairwise comparisons. R_{st} values bellow diagonal and respective *P*-values above diagonal.

	Caia	Foupana	Quarteira	Almargem	Aravil
Caia	–	0.0000*	0.0001*	0.0069*	0.0002*
Foupana	0.089	–	0.0000*	0.4402	0.1628
Quarteira	0.104	0.247	–	0.0000*	0.0000*
Almargem	0.113	0.000	0.323	–	0.9992
Aravil	0.139	0.009	0.382	0.000	–

Significant values after Bonferroni corresctions are marked with *

The genetic structure of populations, without prior knowledge of their geographic origin, revealed a maximum value for the estimated likelihood at K=3 (Figure 8) accordingly to Evanno *et al*. (2005). Basically, for K=3, individuals from Quarteira were clearly separated from individuals of other populations while individuals from Caia and Aravil where considered as a second group and, Foupana and Almargem as a third one.

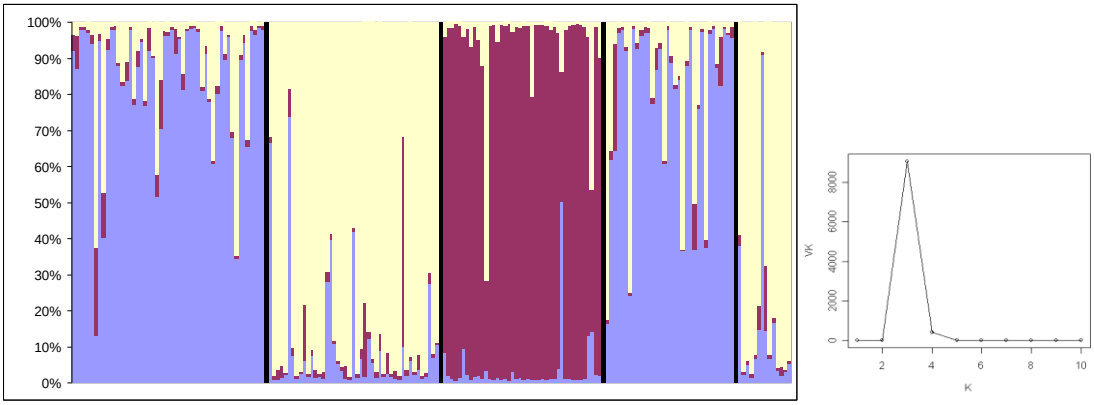

Figure 8. Left: Most likely population structure obtained with STRUCTURE 2.2 software (Pritchard *et al.*, 2000) under the admixture model. Populations are separated with black lines and are respectively: Caia, Foupana, Quarteira, Aravil and Almargem. Each vertical line corresponds to one individual. Each color represents one structure group. Right: Distribution of the rate of change in the log probability of data between successive values of K accordingly to Evanno *et al.* (2005). The modal value of this distribution represents the true value of K.

It is also worth noticing that the results obtained with K=2 (Appendix) already revealed a strong differentiation of the Quarteira population which formed one defined group with all other populations forming a second group. These results seem to support the highly divergent nature of the Quarteira population as it was already proposed by previous analyses like *F*-statistics and FCA analysis.

AMOVA analysis was performed for the 3 groups defined by STRUCTURE (Caia+ Aravil, Foupana + Almargem and Quarteira) as well as for samples structured in 4 different river basins (Almargem, Quarteira, Foupana + Caia and Aravil). However, both analyses did not show significant variance among groups (Table X).

Table X. AMOVA analysis results for microsatellite data testing population structure defined by river basins and STRUCTURE results

	BASINS				
Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Index
Among groups	3	47.038	0.015	0.820	F_{CT}=0.008
Among populations within groups	1	20.470	0.197***	10.610	F_{SC}=0.107
Within populations	369	606.596	1.64***	88.580	F_{ST}=0.115
Total	373	674.104	1.856	100	–
	STRUCTURE				
Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation Index
Among groups	2	58.206	0.168	8.920	F_{CT}=0.089
Among populations within groups	2	12.301	0.071***	3.780	F_{SC}=0.042
Within populations	369	606.596	1.644***	87.300	F_{ST}=0.127
Total	373	674.104	1.883	100	–

On top: a basin like structure and on bottom: the structure as defined in STRUCTURE 2.2 software (Pritchard *et al.* 2000). Significant values are marked with * (p<0,05), ** (p<0,01) or *** (p<0,001).

Nevertheless, for the three group's structure there is a greater percentage of variation explained by differentiation among groups then for basin groups. When the population structure proposed by STRUCTURE is taken into account it leads to a ten-fold increase in the percentage of variance explained "among groups" which adds some support to the already obtained data from STRUCTURE and pairwise F_{ST}.

3.3 Demographic history

3.3.1 Mitochondrial DNA

Tajima's *D* and Fu's *Fs* statistics were both negative for the majority of populations, but only *Fs* values were significantly negative (Table XI). These statistics could not be accurately calculated for the Quarteira sample due to the existence of one single haplotype in this sampling set. Negative *Fs* values indicate a rejection of the selective neutrality hypothesis due to natural selection or a large population expansion.

Table XI. Tajima's *D* and Fu's *Fs* neutrality test results.

	Tajima's *D*	*p-value*	Fu's *Fs*	*p-value*
Caia	0.042	0.544	-12.314	0.000
Foupana	-0.217	0.446	-15.589	0.000
Quarteira	0.000	1.000	—	1.000
Almargem	-1.218	0.119	-14.809	0.000
Aravil	1.221	0.913	-9.911	0.000

The mismatches distribution analyses produced to test the hypothesis of a sudden population expansion model (Figure 9) only showed a clear unimodal distribution of pairwise differences for the Foupana sample. Almargem and Caia presented what can be called the very smooth unimodal distributions, while the mismatch distribution pattern for Aravil was difficult to identify. However, none of the used goodness-of-fit tests were significant, thus, the null hypothesis of the equality of the observed and expected values was not rejected and the sudden expansion model not accepted. Although the Aravil population presented an almost significant *p*-value for the SSD test (Table XII).

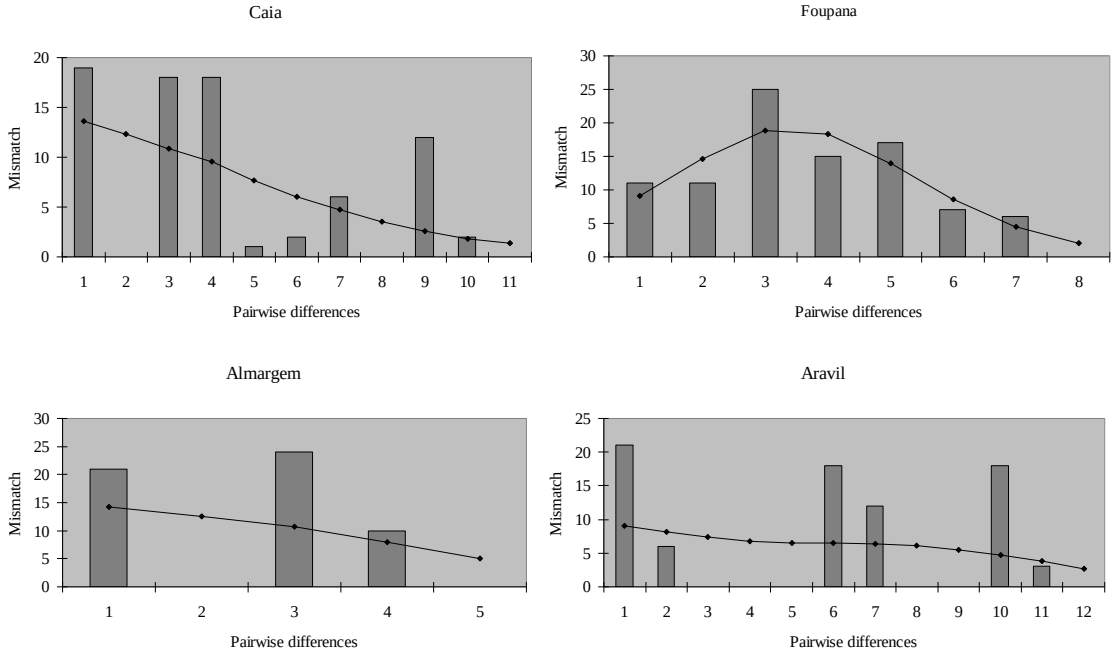

Figure 9. Mismatch distribution of the observed and expected values for each sample.

Table XII. Mismatch distribution parameters and goodness of fit tests.

	Caia	Foupana	Quarteira	Almargem	Aravil
M_{OM}	3.256	2.67	0	1.418	4.615
M_{OV}	8.037	2.846	0	1.396	12.785
T	2.287	3.15	-	2.895	8.586
τ (C.I.)	0 – 87.3	0.09 – 6.5		0 – 57.9	0 - 976
Exp.	107 507	148 075		136 087	403 610
Goodness-of-fit tests					
SSD	0.0764	0.008	-	0.127	0.114
P(SSD)	0.194	0.626	-	0.117	0.055
R	0.209	0.047	-	0.434	0.217
p(R)	0.141	0.661	-	0.087	0.07

M_{OM}: Mismatch observed mean; M_{OV}: Mismatch observed variance; τ: time in number of generations; τ (C.I.): confidence intervals of the number of generations; Exp: expansion time in years; SSD sum of square deviations; R: Reggedness index. Significant values are marked with * (P<0.05), ** (P<0.01) or *** (P<0.001).

Even tough none of the samples unequivocally fitted an sudden expansion model and despite the broad confidence intervals associated with this kind of calculations, expansion times for putative expansion events in *I. lemmingii* were calculated and seem to have occurred during the Ice Age. As a mere theoretical exercise, expansion time for the Aravil population was calculated in 0.4 million years ago while time since expansions for Caia, Foupana and Almargem point to 107 000, 148 000 and 136 000 years respectively (Table XII).

3.3.2 Microsatellites

Regardless of the mutation model taken into account none of the populations revealed signs for bottleneck departures from the mutation drift equilibrium under the different models as all *p*-values were superior to 0.05 (Table XIII). Actually, some of the results seem to be significant for population expansion (P>0.95).

Tab XIII. BOTTLENECK results based on microsatellite loci

	IAM	SMM	TPM (70%)	TPM (90%)
Foupana	0.71094	0.99609	0.97266	0.98828
Caia	0.32031	0.99023	0.84375	0.97266
Quarteira	0.47266	1.00000	0.98633	1.00000
Aravil	0.65625	0.99219	0.96094	0.97656
Almargem	0.31250	0.98438	0.92188	0.95313

IAM (Infinite Allele Model), SMM (Stepwise Mutation Model), TPM 70% (Two-Phase mutation Model with 70% of SMM), TMP 90% (Two-Phase mutation Model with 90% of SMM).

4. DISCUSSION AND FINAL REMARKS

4.1 Genetic Diversity

The overall values of genetic diversity in *I. lemmingii* were generally higher than those observed for other endangered *Iberochondrosoma* species present in the south region of Portugal, like *I. lusitanicum* (Sousa *et al.*, 2007) and *I. almacai* (Sousa *et al*., in press).

Levels of mitochondrial diversity were very heterogeneous among populations with Quarteira and Foupana as their extreme representatives, showing, respectively, one and seven haplotypes ($0.000 \leq h \leq 0.816$). The null haplotype diversity of Quarteira population contrasts with the obtained results for the sympatric populations of *S. aradensis* (h=0.537 and h=0.392) (Mesquita *et al*., 2005).. As the only haplotype present in Quarteira is shared with the Almargem sample, the referred low levels of haplotype diversity for *I. lemingii* might be a result of a common origin or colonization by a small number of individuals from the Almargem population, with consequent bottleneck effect (see below). However, the differences in results can also reflect differences in population size between the two species and sampling errors cannnot be excluded.

Concerning microsatellite data, *I. lemmingii* variability appeared to be more homogeneous among samples, but the values of expected heterozygosity obtained for these populations were higher than those obtained for other *Iberochondrostoma* species (Sousa *et al*., 2007, in press). However, comparisons among species must be addressed with careful considerations, as there might be a certain bias due to differences among the used markers. Thus, by only considering the common *loci* used for *I. lemmingii* and *I. lusitanicum* (LCO4, LCO5, E1G6, N7K4), the diversity levels observed were still higher in *I. lemmingii* for most *loci* (LCO4, E1G6, N7K4).

Robalo *et al*. (2008), explained this pattern of higher genetic variability in *I. lemmingii* as a consequence of species' differential distribution sizes: species with a broader distribution (like *I. lemmingii*) are more likely to maintain several ancestral polymorphisms while species with more restricted ranges (like *I. almacai* and *I. lusitanicum*) might incur in higher losses of genetic variability due to stochastic events.

However, *I. lemmingii* presented slightly lower values of genetic variability than *Anaecypris hispanica* (Alves *et al*., 2000; Salgueiro *et al*., 2003), a "Critically Endangered" cyprinid (Cabral *et al*., 2005), restricted to the Guadiana drainage, which showed unexpectedly high values of genetic diversity.

4.2 Genetic Structure

In general, freshwater fishes patterns of population structure are usually associated with geographical basins, as observed in *I. lusitanicum* and *I. almacai* (Sousa *et al*., 2007, in press). However, despite the significant levels of genetic differentiation among all samples of *I.lemingii*, a pattern of population structure based on inter-drainages similarities and strong intra-drainage differentiation emerged.

The considerably high levels of genetic differentiation found between all samples were supported by Φ_{ST} mitochondrial and F_{ST} microsatellites results, as expected in highly fragmented populations (Salgueiro *et al.*, 2003). But, the analyses performed by BAPS and STRUCTURE software, used, respectively, to determine the hidden population structure in mitochondrial and microsatellite data, showed a lower genetic differentiation between Caia and Aravil (Guadiana and Tagus drainages, respectively). This closer relationship had already been reported by other authors, some of whom reported that these drainages were connected until 80 000 years ago (Carmona *et al*., 2000; Robalo *et al*., 2008). Also, a high level of differentiation was observed among Guadiana samples (Φ_{ST}= 0.264 and F_{ST}=0.109 between Caia and Foupana).. This pattern of high intra-drainage differentiation had already been reported in another cyprinid species distributed in this drainage, namely, *A. hispanica* (Alves *et al*., 2001), and was associated with a decrease in gene flow between populations characteristic of highly fragmented and low dispersal ability species. Furthermore, this structure may also be a consequence of brackish water upstream in the main Guadiana River, isolating downstream freshwater tributaries like Foupana from upstream tributaries like Caia..

Although the analysis of both molecular markers revealed some congruent results (referred above), some incongruences were also found, more specifically, regarding Quarteira, Almargem and Foupana samples. While mitochondrial data revealed Foupana as the most differentiated population (Φ_{ST}=0.642 and Φ_{ST}=0.459 with Quarteira and Almargem, respectively), microsatellite data showed a higher level of

differentiation for Quarteira population (F_{ST}=0.106 and F_{ST}=0.136 with Foupana and Almargem, respectively). However, the analysis of the haplotype network revealed that haplotypes present in Almargem and Quarteira might have been originated in the ones found in Foupana and, thus, it is possible that the referred populations might indeed be more similar than suggested by Φ_{ST} values.

Once Foupana population presents the highest levels of mitochondrial diversity, it is likely that Quarteira and Almargem populations might have been colonized from a small number of individuals from that population, carrying an uncommon mitochondrial haplotype. The existence of a common origin for Quarteira, Almargem and Guadiana populations is not unlikely, since all these drainages have already been considered part of the same ichtyogeographic area by several authors, based on ecological, genetic and historical data (Mesquita *et al*., 2007; Filipe *et al*., 2009). Alternatively, the observed differentiation between Foupana and both Almargem and Quarteira might also have been a consequence of a strong genetic drift in the last two populations, due to smaller population sizes.

4.3 Demographic history

Demographic history of the studied samples was assessed with Fu's *Fs* and Tajima's D tests. Although Fu's *Fs* revealed significant deviations to neutrality in all populations except Quarteira (an expected result as it only had only haplotype sampled), these results were not corroborated by Tajima's D test. Significant deviations to neutrality may be a by-product of demographic events, once mitochondrial DNA is usually considered not to be under selection. Thus, the occurrence of demographic events was also tested by mismatch distribution analyses performed for Aravil, Caia, Foupana and Almargem samples, revealing that: i) Aravil sample appeared to be a stable population (multimodal mismatch distribution); ii) Almargem and Caia showed very weak signs of expansions; and iii) Foupana population presented the strongest signals of expansion with a clear unimodal distribution. However, the weak and non-significant signals observed might be due to the intermittent hydrological conditions that southern drainages are exposed to, causing yearly cycles of bottlenecks and expansions. Moreover, this hypothesis is congruent with the fact that the Aravil population is probably not under extreme droughts as the southern ones, showing an

apparent higher stability. Despite the general weak signals of expansion events, expansion dates for Foupana, Caia and Almargem were, respectively, calculated in 148 075, 107 507 and 136 087 years ago, which correspond to the beginning of the Eemian interglacial period in the late Pleistocene (Van Kolfschoten *et al*., 2002). Although the southern Iberian Peninsula was not affected by the glaciations like the northern and central regions of Europe, it is possible that, as the weather grew warmer, new refugia became available for some species, which, in turn, would have allowed population expansions.

Regarding the detection of demographic events with microsatellite data, no evidences of bottleneck departure from mutation drift equilibrium under the different mutation models were found. The lack of evidences of population collapse contrasts with what has been described for other endangered cyprinids, like *I. almacai* and *I. lusitanicum* (Sousa *et al*., 2007, in press). Nevertheless, it is possible that, in the present case, there was not enough statistical power to detect recent demographic events, since, although the used method can produce results with only 4 *loci*, a higher statistical power is achieved with 10 to 15 *loci* (Cornuet & Luikart, 1996). Moreover, Sousa *et al*. (in press) were not able to detect strong or consistent results with BOTTLENECK software for *I. almacai*. However, in the same study the authors (Sousa *et al.*, in press) could attain signs of small effective population sizes, resulting from a decline beginning in the last few centuries, using the method implemented in MSVAR 1.3 (Storz & Beaumont, 2002).

4.4 Implications for Conservation

One of the goals of this study was the delineation of a conservation plan for *I. lemmingii*, currently classified as "Threatened" (Cabral *et al*., 2005).

Although significant differentiation was found in the analysis of the mitochondrial DNA, there are no evidences to delineate different ESUs, but all samples could be considered different MUs.

The general level of genetic diversity for *I. lemmingii* was higher than the ones found in other cyprinid species from southern Portugal. Nevertheless, this species should also be a target of concern due to its population fragmentation and small effective sizes. Furthermore, due to the intermittent nature of the southern drainages, its

populations are more vulnerable to stochastic events and, thus, to sudden population collapses and subsequent losses of genetic diversity, increasing their extinction risk. Therefore, it is still recommended that special attention should be given to these populations.

Among possible intervention conservation measures, the translocation of individuals could be suggested (e.g. between Caia and Aravil) in order to maintain/ increase levels of genetic variability. However, translocation of individuals between southern populations is a matter that deserves more moderation, once the relationships among these populations are not yet fully resolved. Furthermore, once the differences in allelic frequencies were significant, further studies should be performed to determine if populations are adapted to different environments (i.e. if populations are exchangeable *sensu* Templeton, 1989), in order to ensure the success of the restoration actions.

From the conservation point of view, continental aquatic resources have received far less attention than marine or terrestrial, although they present a greater biodiversity density (Prenda *et al.*, 2006). Impacts in water quality, habitat degradation and construction of dams, that significantly alter the physical structure of the ecosystem and the ecological functioning of flowing waters, are some of the main threats to conservation of freshwater fishes. Species introductions (e.g. in the last century several exotic species have been introduced; Almaça, 1995), translocations and illegal water captures are also other threats that might be responsible for species decline and extinction (Collares-Pereira *et al*., 2000; Collares-Pereira & Cowx, 2004). Thus, an efficient management strategy should also be focused on the habitat, once there is no sense in recovering a species that cannot survive on its own environment due to anthropogenic influences. Therefore, the creation of a management program with river vigilance would allow the quicker detection and response to the previously described impacts, being not only a great benefit for *I. lemmingii*, but also for all the remaining species that share the same habitat.

Finally, the obtained results show that the population structure and demographic history of *I. lemmingii* is rather complex and a good conservation programme could only profit from a deeper knowledge on the phylogeography of this species. A great effort is already being made to include two more samples, one from the Guadalquivir drainage and another located between the Caia and Foupana populations in the Guadiana drainage and, to increase the number of utilised microsatellite *loci* to obtain a

greater statistical power. These measures will complement the already obtained results and will contribute for a better understanding of the genetic diversity and population structure of *I. lemmingii*.

5. REFERENCES

Almaça C, (1995). Freshwater fish and their conservation in Portugal. *Biological Conservation*, **72**:125-127.

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

Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**:37-48.

Belkhir K, Borsa P, Chiki L, Refaust N, Bonhomme F (2000). GENETIX 4.01; logiciel sous Windows pour la génétique des populations. Laboratoire Génome, Populations, Interactions, Université de Montpellier, Montpellier. Available in: http://www.univmontp2.fr/~GENETIX 4.2/GENETIX 4.2/GENETIX 4.2.html.

Bianco P, (1990). Potential role of the paleohistory of the Mediterranean and Paratethys basins on the early dispersal of Euro-Mediterranean freshwater fishes. *Ichthyological Exploration of Freshwaters*, **1**:167–184.

Brito RM, Briolay J, Galtier N, Bouvet Y, Coelho MM (1997). Phylogenetic relationships within the genus Leuciscus (Pisces, Cyprinidae) in Portuguese Fresh Waters, based on mitochondrial DNA cytocrome b sequences. *Molecular Phylogenetics and Evolution*, **3**:435-442.

Cabral MJ (Coord.), Almeida J, Almeida PR, Dellinger T, Ferrand de Ameida, N, Oliveira ME, Palmeirim, JM, Queiroz AI, Rogado L, Santos-Reis M (eds) (2005). Livro Vermelho dos Vertebrados de Portugal. Instituto da Conservação da Natureza. Lisboa. 660pp.

Calvo JP, Daams R, Morales J, Lopez-Martinez N, Augusti J, Anadon P, Armenteros I, Cabrera L, Civis J, Corrochano A, Diaz-Molina M, Muñoz A, Perez-Garcia A, Perez-Gonzalez A, Portero JM, Robles F, Santisteban C, Torres T, Van Der Meulen AJ, Vera

JA & Mein P (1993) Up-to-date Spanish continental Neogene synthesis and paleoclimatic interpretation. *Revista de la Sociedad Geológica de España,* **6:** 29–40.

Carmona JA, Domínguez J, Doadrio I (2000). Congruence between allozyme and cytochrome b gene sequence data in assessing genetic differentiation within the Iberian endemic Chondrostoma lemmingii (Pisces: Cyprinidae). *Hereditary,* **84**:721-731.

Cavalli-Sforza LL & Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics,* **19**:233-257.

Collares-Pereira MJ & Cowx IG (2004). The role of catchment scale environmental management in freshwater fish conservation. *Fisheries Management and Ecology*, **11**:303-312.

Collares-Pereira MJ, Cowx IG, Ribeiro F, Rodrigues JA, Rogado L (2000). Threats imposed by water resource development schemes on the conservation of endangered fish species in the Guadiana River Basin in Portugal. *Fisheries Management and Ecology*, **7**:167-178.

Corander J & Tang J (2007). Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences*, **205**:19-31.

Cornuet JM & Luikart G (1996). Description and power analysis of two test for detecting recent population bottlenecks from allele frequency date. *Genetics*, **144**:2001-2014

Crandall KA, Binida-Emonds ORP, Mace GM, Wayne RK (2000). Considering evolutionary processes in conservation biology. *Tree,* **15**(7):290-295.

Dawson DA, Burland TM, Douglas A, Le Comber SC. Bradshaw M (2003). Isolation of microsatellite loci in the freshwater fish, the bitterling Rhodeus sericesus (Teleostei: Cyprinidae). *Molecular Ecology Notes,* **3**:199-202.

DeSalle R & Amato G (2004). The expansion of Conservation Genetics. *Nature Reviews Genetics*, **5**:702-711.

Doadrio I, (1990). Phylogenetic relationships and classification of west Paleartic species of the genus Barbus (Osteichthyes: Cyprinidae). *Aquatic Living Resources*, **3**:265–282.

Doadrio I (editor). (2001). Atlas y libro rojo de los peces continentales de españa. Dirección General de Conservación de la Naturaleza. Museu Nacional de Ciencias Naturales. Madrid. 364pp.

Doadrio I & Carmona JA (2003). A new species of the Genus *Chondrostoma* Agassiz, 1832 (Actinopterygii, Cyprinidae) from the Iberian Peninsula. *Graellsia*, **58**:29-36.

Doadrio I Carmona JA (2004). Phylogenetic relationships and biogeography of the genus Chondrostoma inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**:802-815.

Evanno E, Regnaut S & Goudet E (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 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**:47-50.

Excoffier L, Smouse P, Quattro J (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**:479-491.

Filipe AF, Araújo MB, Doadrio I, Angermeier PL, Collares-Pereira MJ (2009). Biogeography of Iberian freshwater fishes revisited: the roles of historical versus contemporary constraints. *Journal of Biogeography*, **11**, 2096-2110.

Fernandez-Delgado C & Herrera M (1995). Age structure, growth and reproduction of *Rutilus lemmingii* in an intermittent stream of the Guadalquivir river basin, southern Spain. *Hydrobiologia*, **299**:207-213.

Frankham R, Ballou JD & Briscoe DA (2004). A Primer of Conservation Genetics. Cambridge University Press. 220pp.

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

Gómez A & Lunt DH (2006). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *In*: Phylogeography of Southern European Refugia ,Weiss S. & Ferrand, N. (eds.), Springer, Netherlands, pp: 155-188.

Goudet J (2001). FSTAT: a programe to estimate and test gene diversities and fixation indices (version 2.9.3). Available in: http://www.unil.ch/izea/softwares/fstat.html.

Helfman GS, Collette BB, Facey DE (1997). *Fishes of the world*. Blackwell Publishing, U.S.A. 544pp.

Henriques R (2007). Estrutura genética de *Squalius torgalensis* (Pisces: Cyprinidae), uma espécie endémica ameaçada de peixes dulciaquícolas. Tese de Mestrado, Faculdade de Ciências, Universidade de Lisboa. 84pp.

Jin L, Chakraborty R (1994) Estimation of genetic distance and coefficient of gene diversity from single probe multilocus DNA fingerprinting data. *Molecular Biology and Evolution*, **11**(1):120-127.

Kimura M & Otha T (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the USA*, **75**:2868-2872.

Kumar S, Tamura K, Nei M (2004). MEGA 3: integrated software for molecular evolutionary genetic analysis and sequence alignment. *Briefings in Bioinformatics*, **5**:150-163.

Larno V, Launey S, Devaux Alain, Laroche J (2005). Isolation and characterization of microsatellite loci from chub Leuciscus cephalus (Pisces: Cyprinidae). *Molecular Ecology Notes*, **5**:752-754.

Kalinowski ST (2005). HP-RARE 1.0: a computer program for perfoming rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**:187-189.

Kottelat M & Freyhof J (2007). Handbook of European freshwater fishes. Kottelat, Cornol, Swirzterland and Freyhof, Berlin, Germany. 646pp.

Langella O (2002). Populations 1.2.28. Logiciel de génétique des populations. Laboratoire Populations, génétique et évolution, CNRS UPR 9034, Gif-sur-Yvette. Available in: http://bioinformatics.org/~tryphon/populations/#anne_bibliographie.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006). Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, 7:298-302.

Mesquita N, Cunha C, Carvalho GR, Coelho MM (2007). Comparative phylogeography of endemic cyprinids in the south-west Iberian Peninsula: evidence ofr a new ichthyogeographic area. *Journal of Fish Biology* **71**:45-75.

Mesquita N, Cunha C, Hänfling B, Carvalho GR, Zé-Zé L, Tenreiro R, Coelho MM (2003). Isolation and characterization of polymorphic microsatellite loci in the endangered freshwater fish Squalius aradensis (Cyprinidae). *Molecular Ecology Notes*, **3**:572-574.

Mesquita N, Hänfling 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.

Moran P (2002). Current conservation genetics: building an ecological approach to the synthesis of molecular and quantitative genetic methods. *Ecology of Freshwater Fish*, **11**:30–55.

Moritz C (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology and Evolution*, **9**, 373-375.

Morrison DA (2005). Networks in phylogenetic analysis: New tolls for population biology. *International Journal for Parasitology*, **35**:567-582.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature*, **403**:135-139.

Nei M (1973) Analysis of Gene diversity in subdivided populations. *Procedings of the National Academy of Sciences USA*, **70**(12):3321-3323

Nei M, Tajima F, Tateno Y (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**:153-170.

Pala I, Coelho MM (2005). Contrasting views over a hybrid compex: Between speciation and evolutionary "dead-end". *Gene*, **347**:283-294.

Piry S, Luickart G, Cornuet JM (1999). BOTTLENECK: a computer programe for determining recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**:248-249.

Posada D & Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**:817-818. Available in: http//darwin.uvigo.es/.

Prenda J, Clavero M, Blanco-Garrido F, Menor A, Hermoso V (2006). Threats to the conservation of biotic integrity in Iberian fluvial ecosystems. *Limnetica*, **25**:377-388.

Pritchard JK, Stephens M; Donnely P (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**:945-959.

Ramirez-Soriano A, Ramos-Onsins S, Rozas J, Calafell F, Navarro A (2008) Statistical Power Analysis of Neutrality Tests Under Demographic Expansions, Contractions and bottlenecks With Recombination. *Genetics*, **179**:555-567.

Robalo JI, Almada VC, Levy A, Doadrio I (2006). Re-examinaton and phylogeny of the genus *Chondrostoma* based on mitochondrial and nuclear data and the definition of 5 new genera. *Molecular Phylogenetics and Evolution*, **42**:362-372.

Robalo JI, Almada VC, Levy A, Doadrio I (2006). Re-examinaton and phylogeny of the genus *Chondrostoma* based on mitochondrial and nuclear data and the definition of 5 new genera. *Molecular Phylogenetics and Evolution*, **42**:362-372.

Robalo JI, Doadrio I, Valente A, Almada VC (2007). Identification of ESUs in the critically endangered Portuguese minnow Chondrostoma lusitanicum Collares-Pereira 1980, based on a phylogeographical analysis. *Conservation Genetics*, **8**:1225-1229.

Rousset F (1996). Equilibrium values for measures of population subdivision for setpwise mutation processes. *Genetics*, **145**:1219-1228.

Salgueiro P, Carvalho G, Collares-Pereira MJ, Coelho MM (2003). Microsatellite analysis of genetic population structure of the endangered cyprinid Anaecypris hispanica in Portugal: implications for conservation. *Biological Conservation*, **109**:47–56.

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: a laboratory manual. New York, NY: Cold Spring Harbor Laboratory Press.

Sanjur OI, Carmona JA, Doadrio I (2003). Evolutionary and Biogeographical patterns within Iberian populations of the genus Squalius inferred from molecular data. *Molecular Phylogenetics and Evolution*, **29**:20-30.

Schmidt TR & Gold JR (1993). Complete sequence of the mitochondrial cytochrome *b* gene in the cherryfin shiner, Lythrurus roseipinnis (Teleostei: Cyprinidae). *Copeia*, **3**:883-885

Selkoe KA & Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Slatkin M & Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity*, **76**:377-383.

Sousa V, Penha F, Collares-Pereira MJ, Chikhi L, Coelho MM (2007). Genetic structure and signature of population decrease in the critically endangered freshwater cyprinid *Chondrostoma lusitanicum*. *Conservation Genetics*, **9**:791-805.

Sousa V, Penha F, Pala I, Chikhi L, Coelho MM (in press). Conservation genetics of a critically endangered Iberian minnow: evidence of population decline and extirpations. *Animal Conservation*, doi:10.1111/j.1469-1795.2009.00317.x

Storz JF & Beaumont, MA (2002). Testing for genetic evidence of population contraction and expansion: an empirical analysis of microsatellite DNA variation using ahierarchical Bayesian model. *Evolution*, **56**:154-166

Tajima F (1989). The effect of change in population size on DNA polymorphisms. *Genetics*, **123**:597-601.

Takezaki N & Nei M (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**:389-399.

Templeton AR (1989). The meaning of species and speciation: a genetic perspective. In *Speciation and its consequences*: 3–27. Otte D & Endler JA (Eds). Sunderland, MA: Sinnauer Associates Inc.

Turner TF, Dowling TE, Broughton RE, Gold JR (2004). Variable microsatellite markers amplify across divergent lineages of cyprinid fishes (subfamily Leuscinae). *Conservation Genetics*, **5**:279-281.

Van Oosterhout, C., Hutchinson W. F., Willis O. M., Shipley P. F. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**:535-538.

Vyskocilová M, Simková A, Martin JF (2007). Isolation and characterization of microsatellites in Leuciscus cephalus (Cypriniformes, Cyprinidae) and cross-species amplification within the family Cyprinidae. *Molecular Ecology Notes*, **7**:1150-1154.

Weir BS & Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, **38**:1358-1370.

Wright S (1951). The genetical structure of populations. *Annals of Eugenetics*, **15**:323-354.

6. APPENDIX

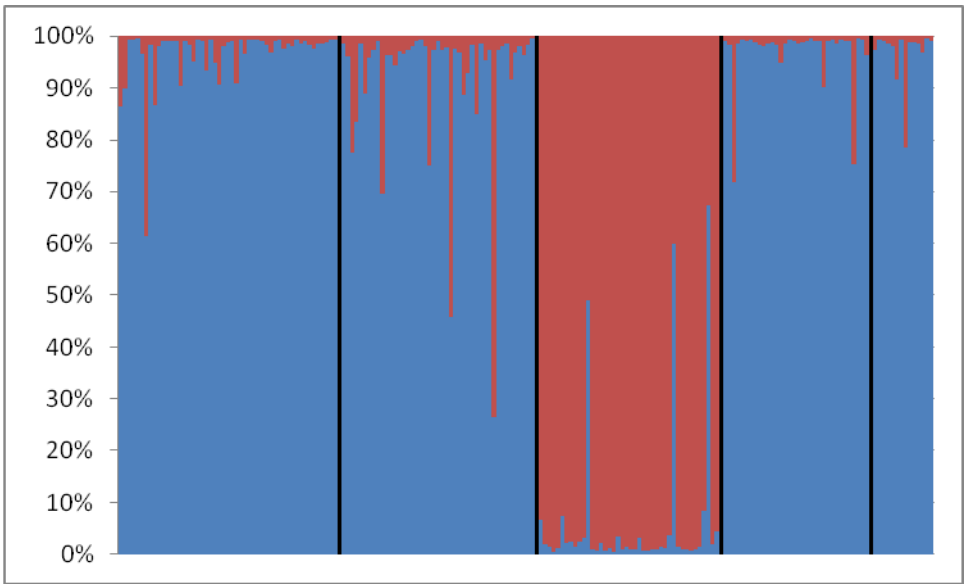

Figure 6.1. Population structure obtained with STRUCTURE 2.2 software (Pritchard *et al.*, 2000) under the admixture model with K=2. Populations are separated with black lines and are respectively: Caia, Foupana, Quarteira, Aravil and Almargem. Each vertical line corresponds to one individual. Each color represents one structure group.