

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL



High unexpected genetic diversity of a narrow endemic terrestrial mollusc

Pedro Madeira

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Professora Doutora Rita Castilho
Professora Doutora Maria Manuela Coelho

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL



High unexpected genetic diversity of a narrow endemic terrestrial mollusc

Pedro Madeira

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Professora Doutora Rita Castilho
Professora Doutora Maria Manuela Coelho

Preliminary Note

This Dissertation is presented including a scientific manuscript (chapter 2), enclosed by a general introduction (chapter 1) and final remarks (chapter 3), that has been submitted for publication in a scientific journal of the specialty, and follows the redaction rules of a standard journal.

Agradecimentos

A realização deste trabalho nunca teria sido possível sem o apoio das seguintes pessoas:

Professora Doutora Rita Castilho, que ao longo dos últimos anos se tem sempre disponibilizado para ser minha orientadora, contribuindo para o enriquecimento da minha formação pessoal e no mundo científico, inculcando sempre noções de responsabilidade e autonomia e guiando-me sempre que necessário na obtenção dos conhecimentos necessários à realização deste projeto. Um sincero agradecimento por exigir sempre o melhor possível de mim, pela compreensão e por nunca ter desistido de me empurrar para concluir o trabalho, e pelo apoio valioso que me deu para continuar a trabalhar depois de cada obstáculo. Nunca esquecerei todas as oportunidades que me proporcionou.

Professor Gonçalo Calado, que motivou este tema, recolheu e enviou as amostras que deram origem a este projeto e que me guiou pelo Vale da Couda, mostrando-me o local onde esta espécie habita e ensinando-me a fazer amostragem de caracóis terrestres.

Professor Francisco Moreira, pela cedência das coordenadas geográficas dos locais de amostragem e pelo que também me ensinou durante a saída de campo para amostragem deste caracol.

David e Geraldine Holyoak, por se disponibilizarem para auxiliar com a identificação morfológica de várias amostras utilizadas neste trabalho e por enviarem amostras de espécies que de outra forma não teria conseguido obter.

Doutora Regina Cunha, pela ajuda com o tratamento dos dados relativos à filogenia.

Doutora Rosa Chefaoui pela construção do modelo de distribuição de *C. coudensis* que permitiu a criação do modelo biogeográfico.

Professora Maria Manuela Coelho, por se disponibilizar para ser a minha orientadora vinculada à faculdade, pela contribuição preciosa para a conclusão e entrega desta tese e pelas palavras de conforto após alguns problemas.

Aos membros do laboratório de Biogeografia, Ecologia e Evolução do Centro de Ciências do Mar que estiveram sempre disponíveis para me ajudar durante este trabalho, partilhando conhecimento ou simplesmente guiando-me pelos recantos do laboratório.

Um último agradecimento à minha família, que acima de tudo sempre me deu a possibilidade de seguir o caminho académico que escolhi e sempre acreditaram que seria de completar fosse o que fosse com sucesso.

Resumo

Os processos geográficos e demográficos deixam marcas nos padrões de variabilidade genética das espécies. Sendo a diversidade genética essencial para garantir capacidade de adaptação das populações a alterações ambientais, a baixa diversidade genética representa um risco mais elevado para os endemismos, que exibem geralmente diversidade mais empobrecidas. *Candidula coudensis* é uma das 12 espécies de *Candidula*, um género de gastrópodes terrestres, recentemente identificada em Portugal. A espécie constitui um endemismo do Vale da Couda, no distrito de Leiria, com distribuição geográfica extremamente reduzida (13,5 km²). Dada a potencial vulnerabilidade de um endemismo tão localizado, foram usadas sequências do ADN mitocondrial (COI: citocromo oxidase subunidade I) e nuclear (ITS1: espaçador interno transcrito subunidade 1), e produzidos modelos de nicho ecológico, para avaliar a diversidade genética e a demografia histórica dos indivíduos do Vale da Couda e, ainda, mapear a sua potencial distribuição geográfica. De forma a determinar se os indivíduos de Vale da Couda constituem um grupo monofilético, foi construída, a partir das sequências de ADN mitocondrial, uma filogenia com as espécies do género presentes em Portugal. A população do Vale da Couda apresentou elevados valores de diversidade genética, existindo um elevado número de haplótipos que se dividem em quatro grupos não monofiléticos. Não parece existir nenhum padrão filogeográfico com os haplogrupos distribuídos de forma não uniforme nos diferentes locais de amostragem. A história demográfica dos indivíduos do Vale da Couda indica sinais de expansão demográfica, sendo necessária uma taxa de mutação de 15%MY para que esta date do Último Máximo Glacial (UMG). O modelo de nicho ecológico determinou que as características ecológicas mais importantes na restrição da distribuição de *C. coudensis* são a litologia, a isothermalidade e a precipitação anual. Durante o Último Máximo Glacial (UMG) a distribuição desta espécie localizar-se-ia mais a sul do que a distribuição atual, numa região onde provavelmente coexistiriam outras espécies de *Candidula*, como *C. setubalensis* e *C. arrabidensis*, que presentemente habitam aquela região. Com o final do UMG terá havido uma re-colonização para norte até ao local actualmente ocupado, onde as condições ambientais são apropriadas à sua persistência. Os padrões de elevada variabilidade genética e divergência entre haplogrupos no Vale da Couda podem ser tentativamente explicados por um cenário biogeográfico em que as condições do UMG tenham favorecido um isolamento de populações a um nível micro-geográfico, potenciando a diferenciação entre as mesmas e aumentando a variabilidade intraespecífica. O Vale da Couda terá então sido colonizado pelas várias linhagens que se formaram durante o UMG, o que explica a existência dos vários haplogrupos existentes actualmente nessa região. O facto de a análise demográfica demonstrar que existiu uma expansão da população corrobora esta hipótese. Actualmente, a elevada variabilidade genética desta população é mantida devido à grande dimensão da mesma.

Concluindo, a população do Vale da Couda contraria a ideia de que populações com distribuição limitada apresentam diversidades genéticas diminuídas, acrescentando este exemplo à lista de excepções a esta regra. Uma vez que, a elevada diversidade genética está associada a uma elevada população efectiva, e com tempo de geração curto, não há sinais de que a população de Vale da Couda precise de especial protecção do ponto de vista da conservação.

Palavras-chave: ADNmt; ADNn; *Candidula*; COI; Divergência genética; Filogeografia; ITS1; Modelos de nicho ecológico; UMG;

Abstract

Patterns of genetic variability are shaped by a combination of geographic, demographic and evolutionary processes. Genetic diversity is the raw material for adaptation. Populations and species with restricted geographic distributions usually display low levels of diversity, which represent an evolutionary risk. *Candidula coudensis* is an endemic species of land snails from Vale da Couda, Portugal, with an extremely restricted distribution (13.5 km²). Mitochondrial (cytochrome oxidase I) and nuclear (internal transcribed spacer 1) DNA sequences of species of the genus *Candidula* from Vale da Couda and other locations in Portugal were used to evaluate the monophyly of *C. coudensis*, patterns of genetic diversity and the species demographic history. Ecological niche models (ENM) for both present and Last Glacial Maximum (LGM) distributions of *C. coudensis* were constructed to assess possible shifts in the distribution that might have influenced the evolutionary history of this species. The population of Vale da Couda showed unexpected levels of genetic variability given its striking restricted geographic range. We found a high number of haplotypes which are part of four non-monophyletic haplogroups, unevenly distributed among sites. The putative past and contemporary models of geographic distribution of Vale da Couda lineages are compatible with a scenario of species co-existence in more southern locations during the last glacial maximum (LGM) followed by a post-LGM northern dispersal, tracking the species to optimal thermal, humidity and soil physical conditions. LGM conditions may have favoured population isolation at a micro-geographical level, promoting intra-specific diversification. This diversity is maintained presently due to a large effective population size. Mismatch analysis indicated a population expansion during the LGM, which corroborates our biogeographic scenario. The high divergence and non-monophyly between haplogroups prompts the need for future molecular taxonomy studies of the *Candidula* genus in Portugal to clarify the identity of species that seem to be non-monophyletic.

Keywords: *Candidula*; COI; Ecological niche models; Genetic divergence; ITS1; LGM; mtDNA; nDNA; Phylogeography

INDEX.....	VI
RESUMO	IV
ABSTRACT	V
INDEX OF FIGURES.....	VII
INDEX OF TABLES.....	VIII
LIST OF ABBREVIATIONS.....	IX
1. GENERAL INTRODUCTION.....	1
1.1. MITOCHONDRIAL AND NUCLEAR DNA.....	2
1.2. ESTIMATING GENETIC DIVERSITY IN POPULATIONS	3
1.3. ECOLOGICAL NICHE MODELLING	4
1.4. THE GENUS <i>CANDIDULA</i> IN PORTUGAL.....	5
1.5. OBJECTIVES.....	8
1.6. RESEARCH QUESTIONS AND HYPOTHESES	8
1.7. REFERENCES	9
2. HIGH UNEXPECTED GENETIC DIVERSITY OF A NARROW ENDEMIC TERRESTRIAL MOLLUSC*	12
2.1. ABSTRACT	12
2.2. INTRODUCTION.....	13
2.3. MATERIAL AND METHODS	14
2.3.1 <i>Taxon sampling</i>	14
2.3.2 <i>Laboratory procedures and sequence alignments</i>	15
2.3.3 <i>Population genetics</i>	16
2.3.4 <i>Taxonomic context</i>	16
2.3.5 <i>Environmental niche modelling</i>	16
2.4. RESULTS	17
2.4.1 <i>Population genetics</i>	17
2.4.2 <i>Phylogenetic estimation</i>	21
2.4.3 <i>Niche modelling</i>	21
2.5. DISCUSSION.....	24
2.5.1 <i>Population genetics</i>	24
2.5.2 <i>Environmental niche modelling</i>	24
2.5.3 <i>Biogeographic scenario</i>	25
2.5.4 <i>Taxonomic implications</i>	25
2.6. CONCLUSIONS	26
2.7. REFERENCES	26
3. FINAL REMARKS	31
3.1. REFERENCES	32
4. APPENDIX.....	33

Index of Figures

Figure 1.1 - Distribution of <i>Candidula</i> species in Portugal (Holyoak & Holyoak 2014).	7
Figure 2.1 - (A) Vale da Couda collection sites. (B) Distribution of <i>Candidula</i> sampling sites in mainland Portugal	15
Figure 2.2 - (A) MtDNA COI statistical parsimony haplotype network for Vale da Couda individuals. (B) Nuclear ITS1 statistical parsimony haplotype network for Vale da Couda individuals. Each branch represents one inferred mutational step; small black circles on branches represent additional inferred mutational steps; numbers in black squares denote more than twenty mutations.	19
Figure 2.3 - Distribution of mtDNA lineages in Vale da Couda sites. Numbers represent sites; numbers in <i>italic</i> represent distance in km between sites. Size of circles is proportional to the number of individuals. Colors depicting haplogroups are the same as in Figure 2.2	20
Figure 2.4 - Mismatch distribution of lineage D from Vale da Couda, indicating number of individuals in the analysis (n), Fu's FS test of selective neutrality and population expansion, evolutionary expansion age in mutational units (τ), effective population size before (θ_0) and after (θ_1) population expansion, and mean expansion time in units of thousand years (ka). Note that the range in expansion age corresponds to the 95% confidence interval of (s). SSD represents the sum of squares deviations, rg the raggedness statistics and R_2	21
Figure 2.5 - Predicted geographic distribution built with presences of Vale da Couda individuals and based on (A) current climate, and (B) Last Glacial Maximum (LGM) conditions. Colour scale represents high probability of occurrence in red and low levels in blue. Black dots represent the present-day known occurrences in the Vale da Couda. Ice sheet existing during LGM in the north of the Iberian Peninsula is not depicted.....	23
Figure 4.S1 - Estimates of net evolutionary divergence between haplogroups (axis on the left, dark grey bars \pm standard deviation) and within lineages (axis on the right, light grey bars \pm standard deviation), based on Tamura-Nei distances.....	35
Figure 4.S2 - Phylogenetic relationships between <i>Candidula</i> individuals from Vale da Couda (in red) and other locations in Portugal. Black squares represent individuals identified morphologically and anatomically. Colours of locations as in Figure 1.1. Outgroups removed from figure for illustrative purposes.....	35

Index of Tables

Table 2.1 - Vale da Couda lineages, sample sizes and summary statistics for COI and ITS1 sequence fragments (A). Vale da Couda site sample sizes, lineages present and summary statistics for COI and ITS1 (B)..... 18

Table 2.2 - Environmental Niche Factor Analysis (ENFA) results showing marginality and specialization factors scores. The three variables with higher marginality scores (in bold) were selected for subsequent analyses. 22

Table 2.3 - Mean AUC, sensitivity and specificity scores obtained from the 100 Maxent models according to the three thresholds used. (Spec_sens: threshold that maximizes the sum of the sensitivity and specificity). 22

Table 4.S1 - Sample location and summary statistics for the genus *Candidula*.....33

Table 4.S2 - Bioclimatic variables for current conditions retrieved from WorldClim dataset (Hijmans *et al* 2005)..... 34

List of Abbreviations

AUC – Area under curve

Bioclim – Bioclimatic variables

CI – Confidence interval

COI – Cytochrome oxidase I

DNA – Desoxyribonucleic acid

ENFA – Environmental niche factor analysis

ENM – Ecological niche model

GCM – General circulation model

H_i – Mean observed heterozygosity per individual within subpopulation

H_S – Mean heterozygosity in subpopulations

H_T – Heterozygosity in all populations

ITS1 – Internal transcribed spacer 1

LGM – Last glacial maximum

ML – Maximum Likelihood

mtDNA – mitochondrial DNA

nDNA – nuclear DNA

PCR – Polymerization chain reaction

RNA – Ribonucleic acid

rRNA – Ribosomal RNA

R_g – Raggedness index

ROC – Receiver operating curve

SDM – Species distribution model

SSD- Sum of squared deviations

1. General Introduction

Phylogeography originally combined evidence from both population genetics and geographic distribution, to understand the evolutionary processes that shape population structure (Avice 2000). Phylogeography was developed as a way to connect the microevolutionary processes shaping the differences between populations to the macroevolutionary differences between species and higher taxa (Avice *et al.* 1987). New analytical tools are being developed that render phylogeography into a more integrative science, combining phylogeographic estimates with data relative to species distribution, natural selection and ecological speciation (Hickerson *et al.* 2010).

Population genetics aim is to study the genetic composition of populations, analysing the distribution and frequency changes of both genotypes and phenotypes, driven by the basic evolutionary mechanisms. The four basic mechanisms behind evolutionary processes are: 1) genetic drift; 2) mutation; 3) migration; and 4) natural selection. Genetic drift is the change in allele frequencies resulting from the random sampling of the parental alleles, within the gene pool of a population (Masel 2011). Genetic drift effects are dependent on population size, resulting in the loss of diversity in smaller populations (Masel 2011) and the maintenance of high diversity in larger ones (Mitchell-Olds *et al.* 2007). Mutations are responsible for creating new alleles, and these new alleles can be either neutral, deleterious or advantageous, and their frequency on the population will change depending on that effect (Mitchell-Olds *et al.* 2007). Migration affects genetic diversity by creating gene flow between populations, and can result in the introduction of new alleles in the populations receiving immigrants (Mitchell-Olds *et al.* 2007). Finally, natural selection acts upon the genetic variation created by all these mechanisms and can be directional, balancing or disruptive, leading to a change in the allele frequencies (Mitchell-Olds *et al.* 2007). The evolutionary drivers translate themselves into divergence between gene pools, demographic changes in populations, migrations between metapopulations, that are generally promoted by geological and/or climate events. If genetic isolation is in place by whatever mechanism (e.g. allopatry, sexual selection), it is possible that, in time, local variants of a species turn into endemic species.

Endemic species are found by definition in relatively restricted areas (Gaston 1994). Species with limited geographic distributions that occupy few and specialized habitats, with small population sizes, are more susceptible to local extinctions. The endangerment of an endemic species can have different origins (e.g., loss of its habitat or loss of habitat quality), and it has an immediate impact on the biodiversity of the area. Even if the species does not disappear, a simple reduction of its habitat can represent a major drawback, since the decrease of an already limited geographic area of distribution will greatly influence the survivability of an endemism (Bonn *et al.* 2002). Endemic species constitute a model to explore population genetics in what effectively can be seen as an island setting. The geographic and demographic components interact with genetic dynamics of the species, often determining viability. Genetic diversity is essential to ensure the populations can withstand environmental fluctuations, during short timeframes, and also serves as the basis for selection and capacity to adapt to changes in the environment in the long run (e.g. Frankham 2005; Laikre *et al.* 2009). It is therefore useful to establish the genetic properties of those species, such as genetic diversity and connectivity, as well as historical demography, to better monitor these species and contribute to a sounder conservation management. In the next sections of this introduction, I will briefly address the most relevant components of this thesis: (1) Mitochondrial and nuclear DNA; (2) Estimating genetic diversity in populations; (3) Ecological Niche modelling; (4) The genus *Candidula* in Portugal.

1.1. Mitochondrial and nuclear DNA

The mitochondrial DNA (mtDNA) is one of the most widely used molecules in phylogeography and population genetics, due to its characteristics: (1) is relatively easy to amplify since it has multiple copies in each cell; (2) mtDNA genes are flanked by highly conserved regions, making it easier to create universal primers for each gene; (3) is inherited clonally from the mother, and in most animals is generally free of recombination (Birky 2001); (4) the genome is haploid. Despite some highly conserved mtDNA sections, regions such as the control region or D-loop display a high mutation rate, varying greatly in natural populations.

Cytochrome oxidase I is a mitochondrial gene that has been used for a long time in phylogeny and population genetics studies (Avice 1994). This gene evolves at a pace which enables the discrimination between closely related species (Hebert *et al.* 2004b; Hajibabaei *et al.* 2006; Pfenninger & Schwenk 2007) or different phylogeographic groups within the same species (Wares & Cunningham 2001; Cox & Hebert 2001). In fact, COI was used in combination with other DNA fragments to solve the systematics of gastropods, such as the Order *Anaspiidea* (Medina & Walsh 2000), the phylogeny and relationships of the *Pleurotomariidae* family (Harasewych *et al.* 1997), as well as in the phylogenetic analysis of *Biomphalaria tenagophila* (Jannotti-Passos *et al.* 2010).

To get a more complete picture on the genetic structure of populations and their evolutionary history one should include information on nuclear genes as well. However, in nuclear markers, heterozygous variants often result in a double peak base-calling fluorescence chromatograms generated from the direct sequencing of genomic PCR product because of the diploid condition. These peaks are difficult to disentangle into genetic phases without cloning. There is also some problems while using nuclear DNA, such as the difficulty to isolate DNA haplotypes; choosing a genomic region with high mutation rates; having a greater coalescence time when compared to mitochondrial genes; and the fact that intragenic recombination might generate patterns of reticulate evolution (Presa *et al.* 2002). Reticulate evolution occurs when different evolutionary lineages are combined. This can be due to hybridization between different species, horizontal gene transfer or genetic recombination (Linder *et al.* 2004). In many instances, mito-nuclear discordances are observed, as mitochondrial genes are expected to experience lineage sorting faster than nuclear genes due to the sorting rate being inverse to effective size (Funk & Omland 2003; Zink & Barrowclough 2008) and therefore may yield signals of population discrimination when nuclear DNA does not. Additionally, single loci can be affected by random variation of mutational and coalescent events, which might create some discrepant patterns among independent loci, making it important to utilize more than one loci in phylogeography (Jennings *et al.* 2013).

The internal transcribed spacer 1 (ITS1) is one of the DNA segments that separates the regions responsible for coding the ribosomal RNA, situated between the genes coding for 18 S and 5.8 S (Castalanelli *et al.* 2013). While the ribosomal genes are highly constrained, due to their importance in protein production, the internal transcribed spacers are non-coding, being free from functional or selective constraints, making them evolve at a faster rate (Weider *et al.* 2005). Being part of the rRNA multigene family, there can be multiple copies of ITS1 within one individual, existing as a part of tandem repeats (Elderkin 2009). Even though multiple copies may exist, intragenomic variation tends to not be significant, as these tandem repeats are homogenized by processes of concerted evolution. Nonetheless, some cases of intragenomic variation have been found, so caution is advised when using this gene for phylogenetic inference. Cloning can be used to assess if there is intragenomic variability and if it overlaps with intra and inter-specific variability (Harris 2000). ITS1 is commonly used for phylogeographic studies, with its high mutation rate being useful for inferring population structure and gene flow (Castalanelli *et al.* 2013), as well as being able to infer phylogenetic relationships from populations to families and even higher taxonomic levels (Cheng *et al.* 2006).

1.2. Estimating genetic diversity in populations

A primary goal of empirical population genetic studies is the identification, measurement, and evaluation of genetic diversity and differentiation among and within populations. There is an array of parameters to quantify genetic diversity and differentiation, but their choice is somewhat dependent on the type of genetic marker used. Here, we outline the two most common measures of genetic variation used in population-level genetic studies on DNA sequences, i.e., nucleotide and haplotype diversity. Because these metrics are dependable on population size and age, connectivity between populations, mutation rates and the impact of selection (e.g. Bazin *et al.* 2006; Wares, 2010), they are relevant measures in population genetics. Nucleotide diversity is a summary statistic used to represent molecular diversity within a sample of sequences, a measure of polymorphism of gene sequences within a population or group of individuals. This statistic is defined as being the average number of nucleotide differences per site between any two randomly chosen sequences from a sample population (Nei & Li 1979, Nei 1987). The mitochondrial genome is haploid and within this context, a haplotype is defined as a sequence or group of identical sequences. Haplotype diversity is the probability that two randomly chosen haplotypes are different (haplotype diversity, h ; Nei, 1987).

Species will most naturally become subdivided in more than one population often due to habitat discontinuities and fragmentation. This will promote the existence of smaller groups that will be more inbred relatively to the whole species. Wright (1943) understood the link between the extent of population differentiation and the inbreeding coefficient due to population subdivision, and proposed the most commonly used metric of genetic differentiation Wright's (1943, 1951, 1965) fixation index, F_{ST} , which was developed as part of a set of hierarchical parameters (F_{ST} , F_{IS} , and F_{IT}) to assess the way in which genetic variation is hierarchically partitioned in natural populations. Three measures of heterozygosity compose the F-statistics indexes of Wright (1943, 1951, 1965): H_T the heterozygosity in all populations; H_S the mean heterozygosity in the subpopulations and H_I the mean observed heterozygosity per individual within subpopulation, whereby F_{ST} , the fixation index is defined as $(H_T - H_S)/H_T$; F_{IS} , the inbreeding coefficient index is defined as $(H_S - H_I) / H_S$; and F_{IT} , the overall fixation index is defined as $(H_T - H_I)/H_T$.

In endemic species, typically with small population sizes, the F_{ST} index is particularly important. The populations of endemic species are often smaller in size, and one may expect them to display more subdivision than more numerous species. In general, the differentiation between pairs of populations within a larger sample is calculated as pairwise F_{ST} . In this case, H_T is calculated for the combined sample of the two populations compared. The F-statistics of Wright (1943, 1951, 1965) were devised for a bi-allelic loci scenario, and the F_{ST} -values ranges from 0.0 to 1.0, where 0.0 indicates identical allele frequencies in a pair of populations (no differentiation) and 1.0 indicates alternate fixation for a single unique allele in each population. This index has a major drawback, as it decreases with increasing allelic diversity (Wright 1978; Charlesworth 1998; Hedrick 1999, 2005; Jost 2008; Meirmans & Hedrick 2011). If there is more than one allele and in the extreme case where samples do not share alleles, the maximum value of 1.0 is never reached for fixation indices (Wright 1978; Hedrick 1999; Jost 2008). Thus, fixation indices will systematically underestimate genetic differentiation, especially when using highly polymorphic markers (Hedrick 1999). In the last few decades, the number of loci used in population genetic studies have increased dramatically, and a number of modern analogues to F_{ST} including G_{ST} (Nei 1973, 1987), θ_{ST} (Cockerham 1969, 1973; Weir & Cockerham 1984), and Φ_{ST} (Excoffier *et al.* 1992) are in use along with their standardized versions. The standardized versions are estimated by scaling the observed index-value relative to its maximum achievable value.

More recently, Jost (2008) proposed an alternative measure of differentiation called D : $((H_T - H_S)/(1 - H_S)) \times (n/(n - 1))$, where H_T and H_S are the same as above, and n is the number of

populations. Contrary to F_{ST} measures, D is not affected by within-population allelic variation, and when two population samples do not share any alleles, $D = 1$.

1.3. Ecological niche modelling

Considering the biogeography of a species, its distribution in space and time, we can infer which processes might have given rise to the patterns of variation found in nature (Avice *et al.* 1987). Phylogeography uses information on the distribution of organisms and combines patterns of genetic diversity to understand the evolutionary history of taxa in light of geographical and environmental past and present-day conditions, tying the distribution of a species or populations with the genetic data (Avice *et al.* 1987). Incorporating ecological factors into the analysis give us the possibility to check for correlations between a species distribution, or of certain populations, and their genetic similarities or differences. Ecological niche models serve as one of the best tools to link biogeographical and population genetics and phylogenetic data (Alvarado-Serrano & Knowles 2014).

Ecological niche models (ENMs), also known as Species Distribution Models (SDMs) (Elith & Leathwick 2009) are mathematical tools used to predict a species geographical distribution past, present or future, based on their known ecological niche. ENMs are very popular tools and have been used to understand dynamic changes in species distribution and/or extinctions due to climate changes, to study patterns of speciation, migration and to test phylogeographic hypothesis (Lima-Ribeiro & Diniz-Filho 2013). When applied in a phylogeographic framework, ENMs give us a biological insight into the observed patterns of genetic variation, providing a background that can be used to test hypothesis or even generate new ones. ENMs can serve multiple purposes in phylogeographic studies, being able to identify possible locations of past populations, characterizing the preferences and tolerances of a certain species, evaluate the degree of adaptation to abiotic factors between different populations, check if a niche divergence between populations is connected to speciation and evaluating how distribution shifts can lead to species-specific genetic variation patterns. In the last few years there has been an increased use of ENMs in phylogeographic studies, mostly in a *post hoc* fashion, to try and corroborate previous hypothesis. ENMs can be used as: 1) a visual assessment, to compare patterns of genetic variation with distribution data; 2) to evaluate the possible effects of landscape in the genetic structure of populations or in the gene flow between them; 3) to compare niche differentiation with possible patterns of speciation; 4) to evaluate the environmental stability of a given area; and 5) to generate predictions about the future distribution a species, given the possibility of environmental changes (Alvarado-Serrano & Knowles 2014). With the usage of ENMs we can study more hypothesis than before, as it enables us to uncover the processes shaping genetic diversity by giving a biological value to otherwise generic genetic statistics.

There are various types of ENMs that can be used, but all of them require us to accept three general assumptions, the assumption that the species we are studying maintains the same niche through time, that there is an equilibrium between species and climate and the need of a taxonomic and taphonomic control, when using fossil records (Lima-Ribeiro & Diniz-Filho 2013). There are two main types of algorithms, the mechanistic and the correlative. The mechanistic algorithms are very data intensive, using specific data about the species in study, such as values relative to physiological and physical characteristics, connecting them to their capacity to inhabit different habitats and determine its distribution. On the other hand, correlative algorithms use a statistical approach based on the presence and, in some cases, absence of the species on certain areas to determine its environmental needs. The correlative methods are further divided between the ones that use only data relative to the presence of the species in a certain area and the ones that use both presence and absence (or pseudo-absence in some cases) data. Since mechanistic approaches need a much greater amount of data, correlative methods are usually preferred, being easier to apply while not losing that much precision. Correlative algorithms are

dependent on two types of data, the distribution of the species and the environmental layers. Distribution data should be carefully and thoroughly obtained through multiple surveys, to ensure that the species range is well represented. After this data is organized it is then plotted into a geographical map containing the environmental data. The environmental data is usually available in public databases and needs to be transformed so it can be included in the geographic map. With the data collected, one of the algorithms is chosen and the data should be treated accordingly, as to include absence data or not if the algorithm requires it (Alvarado-Serrano & Knowles 2014). We can then generate the model to be used and evaluate the performance, so we can fine-tune it to ensure the best results possible. To do so, there's a specific test that evaluates the prediction capabilities of ENMs, the Receiver operating characteristic (ROC), from which the value of Area under curve (AUC) is retrieved. The AUC varies between 0.5 and 1 and is a measure of the predictive capacities of the model. An AUC with a 0.75 or higher value is needed for the model to be considered useful, and it indicates that 75% of the times the model predicts the presence of the species correctly (Lima-Ribeiro & Diniz-Filho 2013). With the model validated, we can then project it for different regions, climates and even time periods (Alvarado-Serrano & Knowles 2014).

1.4. The genus *Candidula* in Portugal

The genus *Candidula* is represented by 24 putative species distributed in Europe, from eastern Canary Islands to the Balkans and northwards to Scotland and southernmost Sweden. In Portugal, 12 of those species can be found, with eight being endemic (Figure 1.1). Of the 12 species, 10 have round shells and two, *C. coudensis* and *C. setubalensis*, have a sharply kneeled shell. Most species are hard to distinguish using conchological characters, with *C. intersecta* and *C. olisippensis* being even considered as “cryptic species complexes”, since both present a great variation of morphological characters, depending on the type of habitat they are found and other lineages may exist. Morphological classification is usually dependent on the size analysis of the penial flagellum, as this differs greatly between species. For example, *C. coudensis* and *C. setubalensis* are very hard to distinguish with a conchological analysis, because both have a sharply kneeled shell, in contrast with all the other species that have rounded shells, but internal morphology allows a clear distinction, since they have very different penial flagellum sizes (Holyoak & Holyoak 2014).

Most of *Candidula* species prefer open and dry habitats, usually with calcareous substrate. In Portugal, *Candidula* species are found in a variety of habitats, ranging from rocky limestone grasslands to sand dunes. There are very few records of *Candidula* species coexisting in the same habitat in Portugal, the exceptions being *C. coudensis* with *C. olisippensis* in Vale da Couda, and *C. setubalensis* with *C. arrabidensis*, in Serra da Arrábida, where the kneeled shell species coexist with a round shelled neighbour, and still *C. belemensis* with *C. olisippensis*, in some locations in Beira Litoral, such as Serra do Sicó, and *C. gigaxi* with *C. ponsulensis*, in eastern Baixo Alentejo (Holyoak & Holyoak 2014).

Candidula coudensis (Holyoak & Holyoak 2010) is a recently found endemism, in the region of Leiria, Portugal, with a highly restricted geographic distribution, in an area of 13.5 km² of Vale da Couda (Moreira *et al.* 2015). There are few instances where an endemic species has such a small geographic distribution. The species can be found in open rocky limestone substrata, in disturbed areas with low shrubs, grasses and herbs, such as areas occupied by agriculture with areas of natural vegetation and olive tree grounds, in roadside areas or even in stone walls in nearby villages (Moreira *et al.* 2015). Although land snails are known to be able to survive and even take advantage of some habitat disturbance, there are some factors that can endanger *C. coudensis* in this region: (1) wildfires can spread quickly and widely in this type of habitat and may lead to its destruction or fragmentation, while also killing the snails by immolation or dehydration; (2) herbicides and pesticides used to treat the olive grounds lead to a decrease in herbaceous vegetation, which means less food and refuge for these animals;

(3) quarrying of stones and gravel lead to habitat degradation, while releasing great amounts of dust that have negative impacts on wildlife; (4) roads that lead to a decrease of connectivity between populations that have low mobility by nature; (5) and climate change, leading to bigger drought periods that can affect land snails directly or augment the potential for wildfires (Moreira *et al.* 2015). If the genetic make-up of *C. coudensis* reveals low diversities, which can affect the survival of individuals, then this is another relevant factor affecting the species that needs to be taken into consideration.

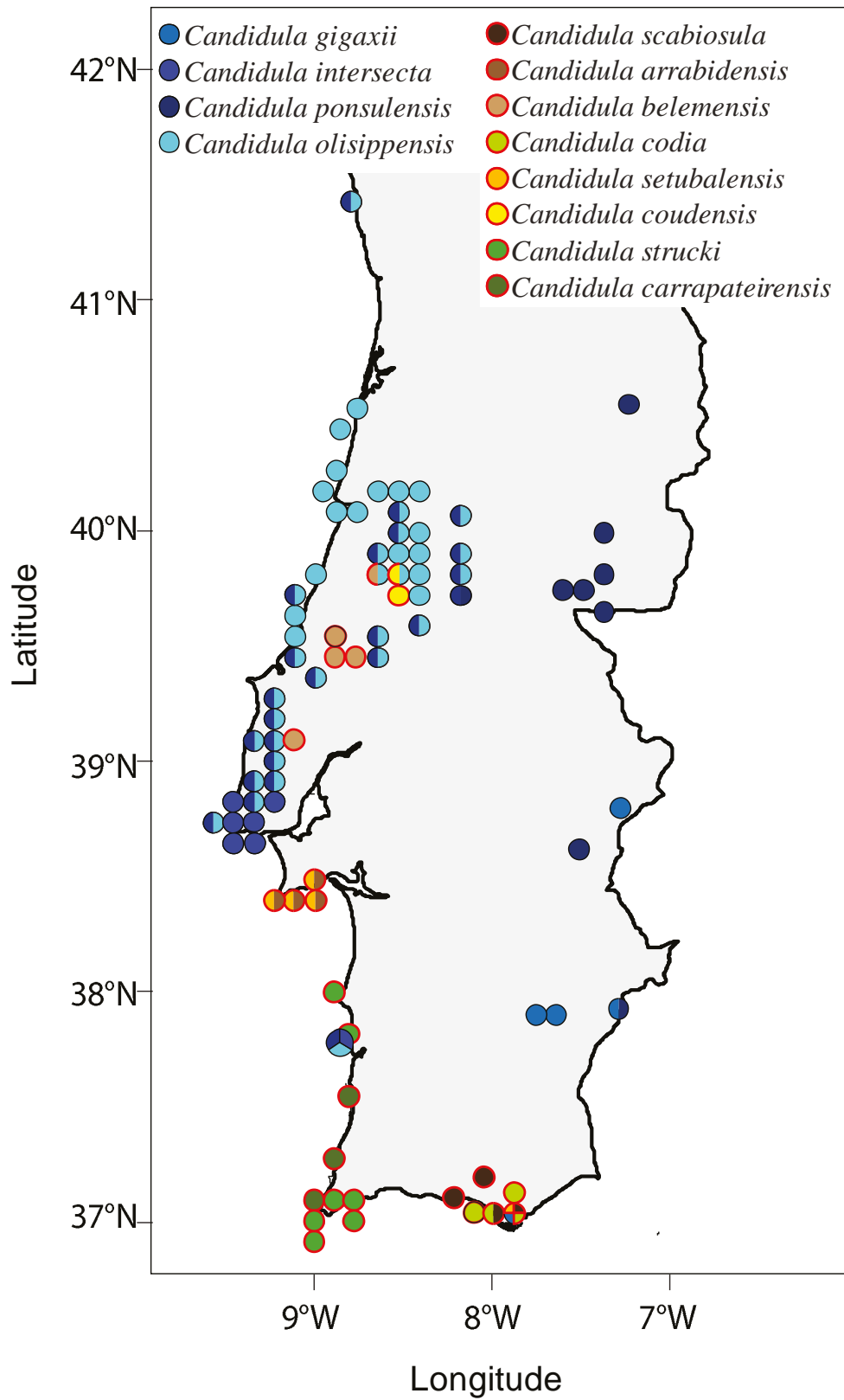


Figure 1.1 - Distribution of *Candidula* species in Portugal (Holyoak & Holyoak 2014).

1.5. Objectives

Land snails are ideal model organisms to study phylogeographic patterns because their reduced dispersal abilities tend to preserve genetic signatures. The main goal of the present study was to describe small-scale spatial patterns of genetic structure in *Candidula coudensis* shedding light on the evolutionary history of a species exhibiting extremely reduced geographic range. We also aimed at producing potential distribution maps based on ecological-niche modeling. To that end, we used molecular sequences from both mitochondrial and nuclear genes from *C. coudensis* and other *Candidula* species in Portugal, and presence/absence records.

1.6. Research questions and hypotheses

Is *Candidula coudensis* a monophyletic clade within the *Candidula* genus in Portugal?

How much diversity is there in the *Candidula* populations of Vale da Couda?

What is the demographic history of Vale da Couda individuals?

Is there a range shift in the putative distribution of *Candidula* during the LGM in comparison to present-day distributions?

We have posited the following hypotheses for *C. coudensis* based on *a priori* theoretical premises that are likely to shape the phylogeographic structure of terrestrial species with low dispersal abilities and restricted geographical distributions and on *a priori* knowledge about the morphology and distribution of *Candidula* species in Portugal:

- (1) Vale da Couda individuals may form a monophyletic clade, indicative of a single population on a restricted area in the absence of major phylogeographic breaks (rivers, large mountains, etc.).
- (2) The population from Vale da Couda may show reduced levels of haplotype and nucleotide diversities, consistent with isolated population on a limited geographical area.
- (3) The population from Vale da Couda may have undergone demographic expansion after the LGM, similarly to many other terrestrial species.

1.7. References

- Alvarado-Serrano DF, and Knowles LL. 2014. Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular ecology resources* 14:233-248. 10.1111/1755-0998.12184
- Avise JC. 1994. *Molecular markers, natural history and evolution*. New York: Chapman and Hall.
- Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, Massachussets: Harvard University Press.
- Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, and Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*:489-522.
- Bazin E, Glémin S, and Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312:570-572.
- Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* 35:125-148. 10.1146/annurev.genet.35.102401.090231
- Bonn A, Rodrigues ASL, and Gaston KJ. 2002. Threatened and endemic species: are they good indicators of patterns of biodiversity on a national scale? *Ecology Letters* 5:733-741. 10.1046/j.1461-0248.2002.00376.x
- Castalanelli MA, Cunningham RJ, Davis MB, Groth DM, and Grimm M. 2013. When genes go wild: highly variable internal transcribed spacer1 and conserved mitochondrial DNA haplotypes used to examine the genetic diversity and dispersal pathways of invasive *Hylotrupes bajulus* in Western Australia. *Agricultural and Forest Entomology* 15:236-244. 10.1111/afe.12010
- Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol* 15:538-543.
- Cheng H-L, Xia D-Q, Wu T-T, Meng X-P, Ji H-J, and Dong Z-G. 2006. Study on sequences of ribosomal DNA internal transcribed spacers of clams belonging to the *Veneridae* family (*Mollusca: Bivalvia*). *Acta Genetica Sinica* 33:702-710. 10.1016/S0379-4172(06)60102-9
- Cockerham CC. 1969. Variance of gene frequencies. *Evolution*:72-84.
- Cockerham CC. 1973. Analyses of gene frequencies. *Genetics* 74:679-700.
- Cox AJ, and Hebert PDN. 2001. Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. *Molecular ecology* 10:371-386. 10.1046/j.1365-294X.2001.01188.x
- Elderkin CL. 2009. Intragenomic variation in the rDNA internal transcribed spacer (ITS1) in the freshwater mussel *Cumberlandia monodonta* (Say, 1828). *Journal of Molluscan Studies* 75:419-421. 10.1093/mollus/eyp043
- Elith J, and Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40:677-697. 10.1146/annurev.ecolsys.110308.120159
- Excoffier L, Smouse PE, and Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Frankham R. 2005. Genetics and extinction. *Biological Conservation* 126:131-140. 10.1016/j.biocon.2005.05.002
- Funk DJ, and Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 34:397-423. 10.1146/annurev.ecolsys.34.011802.132421
- Gaston KJ. 1994. What is rarity? *Rarity*: Springer Netherlands, 1-21.

- Hajibabaei M, Janzen D, Burns J, Hallwachs W, and Hebert P. 2006. DNA barcodes distinguish species of tropical *Lepidoptera*. *Proceedings of the National Academy of Sciences of the United States of America* 103:968-1039. 10.1073/pnas.0510466103
- Harasewych M, Adamkewicz SL, Blake JA, Saudek D, Spriggs T, and Bult CJ. 1997. Phylogeny and relationships of *pleurotomariid* gastropods (*Mollusca: Gastropoda*): an assessment based on partial 18S rDNA and cytochrome c oxidase I sequences. *Molecular marine biology and biotechnology* 6:1-20.
- Harris DJ, and Crandall KA. 2000. Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (*Decapoda: Cambaridae*): implications for phylogenetic and microsatellite studies. *Mol Biol Evol* 17:284-291.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, and Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101:14812-14817. 10.1073/pnas.0406166101
- Hedrick PW. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313-318.
- Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633-1638.
- Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, and Yoder AD. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* 54:291-301. 10.1016/j.ympev.2009.09.016
- Holyoak DT, and Holyoak GA. 2010. A new species of *Candidula* (*Gastropoda, Hygromiidae*) from central Portugal. *Iberus* 28:67-72.
- Holyoak DT, and Holyoak GA. 2014. A review of the genus *Candidula* in Portugal with notes on other populations in Western Europe (*Gastropoda, Pulmonata, Hygromiidae*) *Journal of Conchology* 41:629-672.
- Jannotti-Passos LK, Ruiz JC, Caldeira RL, Murta SM, Coelho PMZ, and Carvalho OS. 2010. Phylogenetic analysis of *Biomphalaria tenagophila* (Orbigny, 1835) (*Mollusca: Gastropoda*). *Memórias do Instituto Oswaldo Cruz* 105:504-511.
- Jennings RM, Etter RJ, and Ficarra L. 2013. Population differentiation and species formation in the deep Sea: the potential role of environmental gradients and depth. *PLoS ONE* 8:e77594. 10.1371/journal.pone.0077594
- Jost L. 2008. GST and its relatives do not measure differentiation. *Molecular ecology* 17:4015-4026. 10.1111/j.1365-294X.2008.03887.x
- Laikre L, Nilsson T, Primmer CR, Ryman N, and Allendorf FW. 2009. Importance of genetics in the interpretation of favourable conservation status. *Conservation Biology* 23:1378-1381. 10.1111/j.1523-1739.2009.01360.x
- Lima-Ribeiro M, and Diniz-Filho J. 2013. Modelos ecológicos e a extinção da megafauna: clima e homem na América do Sul. *CUBO, São Carlos*.
- Linder CR, Moret BM, Nakhleh L, and Warnow T. 2004. Network (reticulate) evolution: biology, models, and algorithms. The Pacific Symposium on Biocomputing.
- Masel J. 2011. Genetic drift. *Current Biology* 21:R837-R838. 10.1016/j.cub.2011.08.007
- Medina M, and Walsh PJ. 2000. Molecular systematics of the Order *Anaspiidea* based on mitochondrial DNA sequence (12S, 16S, and COI). *Molecular Phylogenetics and Evolution* 15:41-58. 10.1006/mpev.1999.0736
- Meirmans PG, and Hedrick PW. 2011. Assessing population structure: F(ST) and related measures. *Molecular ecology resources* 11:5–18. Doi 10.1111/J.1755-0998.2010.02927.X

- Mitchell-Olds T, Willis JH, and Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* 8:845-856. 10.1038/nrg2207
- Moreira F, Calado G, and Dias S. 2015. Conservation status of a recently described endemic land snail, *Candidula coudensis*, from the Iberian Peninsula. *PLoS ONE* 10:e0138464. 10.1371/journal.pone.0138464
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70:3321-3323.
- Nei M. 1987. Genetic distance and molecular phylogeny. In: Ryman N, and Utter FW, eds. *Population Genetics & Fishery Management*. Seattle: Washington Sea Grant Program, University of Washington, 193–223.
- Nei M, and Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76:5269-5273.
- Pfenninger M, and Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC evolutionary biology* 7:121. 10.1186/1471-2148-7-121
- Presa P, Pardo BG, Martínez P, and Bernatchez L. 2002. Phylogeographic congruence between mtDNA and rDNA ITS markers in brown trout. *Mol Biol Evol* 19:2161-2175.
- Primack RB. 2006. *Essentials of conservation biology*. Sunderland, Massachusetts: Sinauer Associates.
- Wares JP. 2010. Natural distributions of mitochondrial sequence diversity support new null hypotheses. *Evolution* 64:1136-1142.
- Wares JP, and Cunningham CW. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 55:2455-2469.
- Weider LJ, Elser JJ, Crease TJ, Mateos M, Cotner JB, and Markow TA. 2005. The functional significance of ribosomal (r)DNA Variation: Impacts on the evolutionary ecology of organisms. *Annual Review of Ecology, Evolution, and Systematics* 36:219-242. 10.2307/30033803
- Weir BS, and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370. 10.2307/2408641
- Wright S. 1943. Isolation by distance. *Genetics* 28:114.
- Wright S. 1951. The genetical structure of populations. *Annals of eugenics* 15:323-354.
- Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*:395-420.
- Wright S. 1978. *Evolution and the genetics of populations. Variability within and among natural populations*. .
- Zink RM, and Barrowclough GF. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology* 17:2107-2121. 10.1111/j.1365-294X.2008.03737.x

2. High unexpected genetic diversity of a narrow endemic terrestrial mollusc*

Madeira, Pedro M.¹, Chefaoui, Rosa M.¹, Cunha, Regina L.¹, Moreira, Francisco^{2,3,4}, Dias, Susana^{2,4}, Calado, Gonalo² and Castilho, Rita^{1†}.

¹Centre for Marine Sciences, CCMAR-CIMAR Laborat3rio Associado, Campus de Gambelas, 8005-139 Faro, Portugal

²Departamento de Ci4ncias da Vida, Escola de Psicologia e Ci4ncias da Vida, Universidade Lus3fona, Campo Grande, 376, 1749-024 Lisboa – Portugal

³REN Biodiversity Chair, CIBIO/InBIO Associate Laboratory, Universidade do Porto, Campus Agr3rio de Vair3o, Vair3o, Portugal

⁴Centro de Ecologia Aplicada Prof. Baeta Neves/InBIO Associate Laboratory, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, Lisbon, Portugal

†Corresponding author:

Centro de Ci4ncias do Mar (CCMAR), Campus de Gambelas, 8005-139 Faro, Portugal. Tel: +351 918397282, Fax: +351 289 800 069.

Keywords: endemic species; terrestrial gastropods

2.1. Abstract

A large array of species and species complexes from the Iberian Peninsula display strong genetic subdivisions indicative of past population isolation in separate glacial refugia as a result of survival throughout the Pleistocene ice ages. We used mitochondrial and nuclear sequence data to analyze phylogeographic patterns in endemic land snails from a valley of central Portugal (Vale da Couda), putatively attributed to *C. coudensis*, and showing an exceptionally narrow distributional range. Considering the restricted geographic distribution, we found an unusually high number of haplotypes that are not

monophyletic and are unevenly distributed among the sampling sites. Our results show a departure from the expectation that species with restricted distributions have low genetic variability. The putative past and contemporary models of geographic distribution of Vale da Couda lineages are compatible with a scenario of species co-existence in more southern locations during the last glacial maximum (LGM) followed by a post-LGM northern dispersal tracking the species optimal thermal, humidity and soil physical conditions. Mismatch analysis indicated a population expansion during the LGM, which corroborates our biogeographic scenario.

2.2. Introduction

Phylogeography combines evidence from both population genetics and phylogenetics, to understand the evolutionary processes that shape geographic population structure (Avice 2000). These evolutionary processes include divergence among gene pools, demographic changes in populations, and migrations between metapopulations, generally promoted or constrained by geological and/or climate events. If genetic isolation is in place by whatever mechanism (e.g. allopatry or sexual selection), it is possible that, in time, local variants of a species turn into endemic species. Endemic species are usually found in relatively small areas (Gaston 1994), occupying specialized habitats with small population sizes that are more susceptible to local extinctions (Primack 2006). Endemic species constitute therefore a model to explore population genetics in what effectively can be seen as an island setting. The geographic and demographic components interact with the genetic dynamics of the species, often determining species viability. Genetic diversity is essential to ensure that populations can withstand environmental fluctuations during short timeframes and also serves as the basis for selection and capacity to adapt to changes in the environment in the long run (Frankham 2005; Laikre *et al.* 2009). It is therefore important to assess the genetic properties of the populations of those species, such as genetic diversity and connectivity, as well as historical demography.

Identifying the drivers of geographic distribution patterns is also essential to understand the population dynamics in space and time. Species distribution modeling (SDM) allows one to examine the relationship between the identified presence records of a species, lineage or related species, with the environmental characteristics of these locations. From the inferred relationship it is possible to estimate the response, function and contribution of environmental variables (Austin *et al.* 2006), and predict the potential geographical range (Elith & Leathwick 2009). Recently, there has been a growing trend towards the integration of SDM hindcasts with phylogeography as a useful approach to obtain consistent eco-evolutionary hypotheses. This combination allows insights into how the distribution of climatic refuges and postglacial colonization pathways may have influenced genetic diversity of current populations (see e.g. Hewitt 2004).

Land snails are good models for evolutionary studies, since phylogeographic patterns are often preserved due to their limited dispersal capabilities and specific habitat requirements (Pfenninger *et al.* 2007). Also, snails display an unusually high intraspecific genetic variation, ca. 10-30% in mtDNA sequences (Hayashi & Chiba 2000; Pinceel *et al.* 2005; Shimizu & Ueshima 2000), which renders the taxa appropriate to understand processes shaping the partitioning of genetic variation in space. Additionally, many land snail examples in the literature show the existence of cryptic species in sympatry (Köhler & Burghardt 2015).

The land snails of the genus *Candidula* present in Europe, from eastern Canary Islands to the Balkans and northwards to Scotland and southernmost Sweden are represented by 24 putative species. Portugal has 8 endemic species (*C. coudensis*, *C. setubalensis*, *C. scabiosula*, *C. arrabidensis*, *C. belemensis*, *C. carrapateirensis*, *C. codia* and *C. strucki*) from a total of 12 (*C. gigaxii*, *C. intersecta*, *C. ponsulensis*, *C. olisippensis*) (Holyoak & Holyoak 2014). Most species are hard to distinguish using conchological characters only and it takes a combination of morphological characters, such as the size of the penial

flagellum or shell shape, to classify the specimens (Holyoak & Holyoak 2014). Nevertheless, a clear, comprehensive, taxonomic assessment based on both morphological and molecular data has not been previously done. Most *Candidula* species prefer open and dry habitats, usually with calcareous substrate. In Portugal, species can be found in a variety of habitats, ranging from rocky limestone grasslands to sand dunes. There are records of coexisting *Candidula* species in Portugal: *C. coudensis* and *C. olisippensis* in Vale da Couda, and *C. setubalensis* and *C. arrabidensis* in Serra da Arrábida, *C. belemensis* and *C. olisippensis* in various locations of Beira Litoral, such as Serra do Sicó, and *C. gigaxii* and *C. ponsulensis* in eastern Baixo Alentejo (Holyoak & Holyoak 2014).

Candidula coudensis (Holyoak & Holyoak, 2010) is a recently described endemic species with a highly restricted geographic distribution in Vale da Couda, Leiria, Portugal. Moreira *et al.* (2015) found live *C. coudensis* individuals only in a small area of ca. 13.5 km². A broader survey of 100 km² in the region, revealed the presence of *C. coudensis* shell remains in only 25.2 km² and to the best of our knowledge, not a single specimen attributed to this species was to this date found outside these limits.

The species can be found in open rocky limestone substrata, olive tree grounds, areas of natural vegetation, in roadside areas or even in stone-walls in nearby houses (Moreira *et al.* 2015). The extremely constrained geographic distribution is somewhat rare and there are several possible non-exclusive reasons that would justify such circumscribed distribution: (i) active dispersal may be very small with individuals hardly moving; (ii) very strict environmental and ecological requirements; (iii) present-day individuals of the species are remnants of an older wide-spread haplogroup that range-contracted due to reduction of humidity levels after the Last Glacial Maximum (LGM, c. 20 ka), and/ or (iv) present-day habitat disturbance processes. Specifically, we tested the following hypotheses based on premises that are likely to shape the phylogeographic structure of the land snails from Vale da Couda, putatively attributed to *C. coudensis*: (1) Vale da Couda individuals may form a monophyletic clade, indicative of a single population on a restricted area in the absence of major phylogeographic breaks (e.g. rivers or large mountains); (2) Vale da Couda individuals may show reduced levels of haplotype and nucleotide diversities, consistent with an isolated population on a limited geographical area, and (3) Vale da Couda individuals may have undergone demographic expansion after the LGM, similarly to many other terrestrial species.

Using a combination of DNA sequences (fragment of the cytochrome oxidase subunit I (COI) mitochondrial gene and of the first nuclear intron - ITS1) and geo-referenced field records of the species we sought to address the above hypotheses by (1) revealing the genetic diversity and geographic structure of contemporary Vale da Couda individuals, and (2) reconstructing the local demographic history. Using Iberian environmental data relative to past and current conditions retrieved from public repositories, we inferred locations of the putative refugia during the LGM and provided estimates of relative environmental suitability of Vale da Couda individuals that can assist future fieldwork.

2.3. Material and Methods

2.3.1 Taxon sampling

Sampling in Vale da Couda resulted in 73 individuals collected from 4 different sites (Figure 2.1). We received field permit from the Nature and Forests Conservation Institute (ICNF), Portugal (identifier: 81S0/201S/DCNF-LVT/DPAP) for sampling in Arrábida Natural Parque. Immediately after collection, whole shells containing the individual were stored in ethanol 70%.

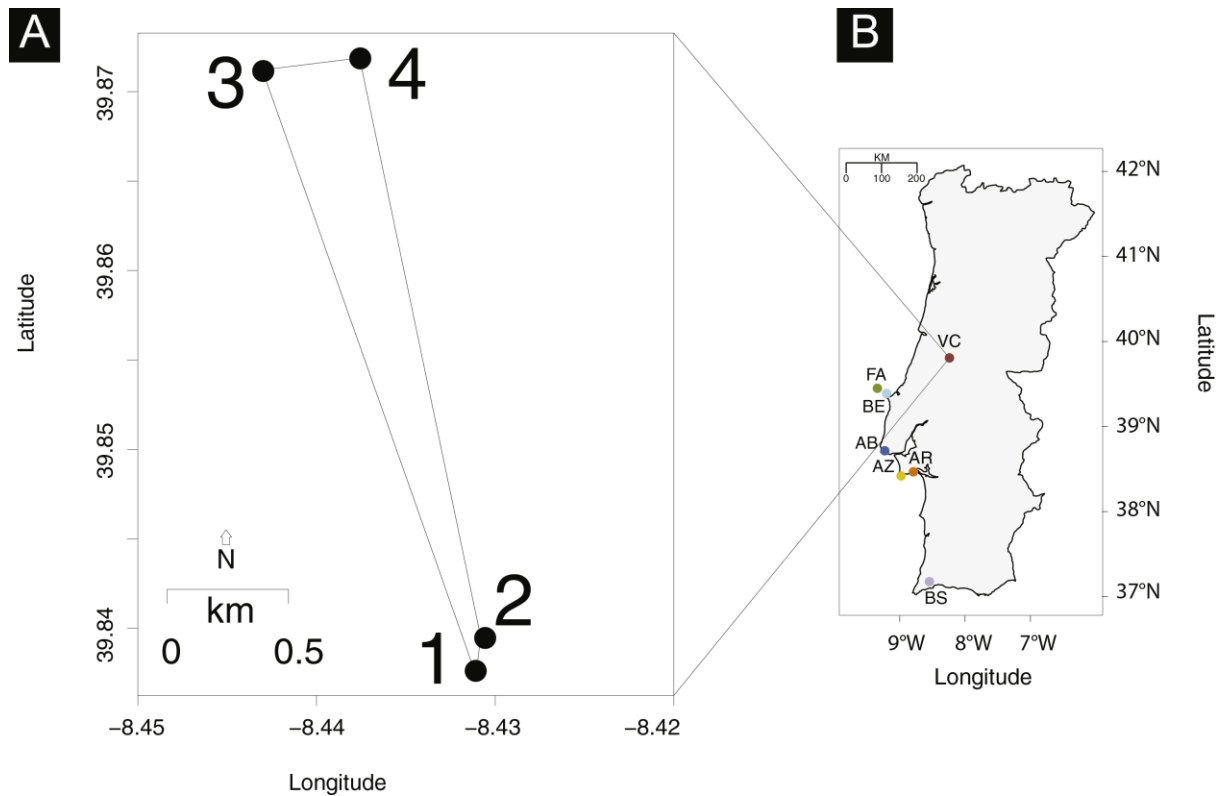


Figure 2.1 - (A) Vale da Couda collection sites. (B) Distribution of *Candidula* sampling sites in mainland Portugal

2.3.2 Laboratory procedures and sequence alignments

DNA was extracted from the samples using a CTAB protocol (Doyle & Doyle 1987). Universal primers (Folmer *et al.* 1994) were used in PCRs to amplify 600 bp of the COI gene. PCR amplifications were performed in 25 µl total volume, using 5 µl 5X PCR Colorless Buffer (pH 8.5), 2 mM (of a 1.5 µl 25 mM MgCl₂ solution), 0.2 mM (0.5 µl of a 20 mM dNTP stock), 0.2 µl 5u/µl 1U GoTaq DNA polymerase Promega (Madison, USA) and 0.2 µM (0.5 µl of a 10 µM stock) of each primer. The COI PCR profile consisted of 2 minutes at 95°C, 35 cycles of 30 seconds at 94°C, 30 seconds at 53°C followed by an extension for 1 minute at 72°C and a final one with 5 minutes. ITS1 gene was amplified by PCR with forward primer ITS1 - 5'-TCCGTAGGTGAACCTGCGGAAGGAT-3' (White *et al.* 1990) and reverse primer 5.8c - 5'-TGCGTTCAAGATATCGATGTTCAA-3' modified from (Hillis & Dixon 1991). PCR amplifications were performed in 25 µl total volume, using 5 µl 5X PCR Colorless Buffer (pH 8.5), 2 mM (of a 1.5 µl 25 mM MgCl₂ solution), 0.2 mM (0.5 µl of a 20 mM dNTP stock), 0.2 µl 5u/µl 1U GoTaq DNA polymerase Promega (Madison, USA) and 0.2 µM (0.5 µl of a 10 µM stock) of each primer. The ITS1 PCR profile consisted of 3 minutes at 97°C, 35 cycles of 1 minute at 95°C, 1 minute at 55°C and 2 minutes at 72°, followed by a final extension of 5 minutes at 72°C. The PCR results were purified by ethanol precipitation (Sambrook & Russell 2001). Sequencing was performed on an ABI 3130xl (Applied BIOSystems) automated sequencer at CCMAR facilities.

COI sequences were aligned using MUSCLE (Edgar 2004), implemented in Geneious version 7.0.4 (Kearse *et al.* 2012), and contained no gaps. Heterozygous ITS1 sequences were fed into Mixed Sequence Reader (MSR) (<http://msr.cs.nthu.edu.tw>), which separates the information from the chromatogram into a major and minor sequence, corresponding to each allele, while comparing the sequence information with a given reference sequence (Chang *et al.* 2012). Major and minor sequences for each sample were recovered and posteriorly aligned using MAFFT default options (Katoh & Standley 2013).

2.3.3 Population genetics

Molecular diversity indices, including nucleotide (π) (Nei 1987) and haplotype (H) (Nei & Tajima 1981) diversities, were estimated using DnaSP v5.10 (Librado & Rozas 2009). To evaluate the level of population differentiation among four Vale da Couda sites, we used F_{ST} genetic fixation (Weir & Cockerham 1984) and D_{est} genetic differentiation (Jost 2008) statistics were estimated with the modelling package *diveRsity* 1.9.5 (Keenan *et al.* 2013). The variance of each statistic was assessed through the calculation of 10 000 pairwise bootstrapped 95% confidence limits using a bias corrected method that basically re-centers the confidence interval (CI) around the initial parameter estimate. We employed both genetic estimators as they present advantages and drawbacks in quantifying population structure (for a discussion see Bird *et al.* 2011; Jost 2008; Meirmans & Hedrick 2011; Ryman & Leimar 2009; Whitlock 2011).

Phylogeographic relationships among haplotypes of COI and ITS1 alleles were represented using the Median Joining Network method (Bandelt *et al.* 1999) implemented in Network (version 4.6.1.0; fluxus-engineering.com) that infers the most parsimonious branch connections between sequences. Net divergences between and within mtDNA and nuclear DNA haplogroups were calculated using MEGA6 (Tamura *et al.* 2013) using the Tamura-Nei model (Tamura & Nei 1993) for both data sets.

Mismatch analysis was used to explore the demographic history of Vale da Couda haplogroups with sample size above 15, with the raggedness index (rg, Harpending 1994), the sum of squared deviations (SSD, Schneider & Excoffier 1999) and R2 (Ramos-Onsins & Rozas 2002) a statistic based on the difference between the number of singleton sites and the average number of nucleotide differences. Demographic mismatch analysis was based on the null hypothesis of expansion; thus, non-significant values mean non-rejection of population expansion. DnaSP (Librado & Rozas 2009) was used to obtain observed and expected distributions under the constant population model and the growth population model.

Initial and final θ estimates (before and after population growth or decline) and τ values were calculated with Arlequin v. 3.5.1.3 (Excoffier & Lischer 2010). Time of inferred population expansion were determined by $T_{exp} = \tau / (2 \mu n)$, where μ = COI substitution rate per base and per generation, and n = number of bases of the COI fragment (Rogers & Harpending 1992), assuming a generation time of one year (Pfenninger *et al.* 2003).

2.3.4 Taxonomic context

In order to place the Vale da Couda samples in a broader phylogenetic context and to ascertain the non-monophyly of the individuals from Vale da Couda (given the very distant haplogroups found – see Results section below), putative *Candidula spp.* individuals were collected in different locations (Supplementary information Table S1). It was not, however, our intention to produce a complete and thorough phylogeny for the genus *Candidula*. We followed Holyoak and Holyoak (2014) taxonomy to identify some specimens based on morphology. The partial sequences of the mitochondrial (mtDNA) COI gene including 73 *Candidula* from Vale da Couda, produced a data set of 464 nucleotide positions. The Akaike Information Criterion (Akaike 1974) implemented in MODELTEST selected the K81uf+I+G as the evolutionary model that best fits the data set. Since this model is not available in PHYML v.3.0 (Guindon *et al.* 2003), we selected the second best-fit model, the HKY+G. The selected model and model parameters were used in the ML analysis performed with PhyML. The robustness of the inferred trees was tested by nonparametric bootstrapping (BP) using 1000 pseudoreplicates. ML analysis was carried out at the Mobyle platform (<http://mobyle.pasteur.fr/cgi-bin/portal.py>).

2.3.5 Environmental niche modelling

The study area considered was the Iberian Peninsula. Bioclimatic variables for current conditions were retrieved from WorldClim dataset (Hijmans *et al.* 2005) in 30 arc seconds (~1 km), resolution used for

all modelling analyses (Supplementary information Table S2). In addition, because of the species preference for limestone soils, where is most frequently found (Moreira *et al.* 2015), we extracted the distribution of carbonate sedimentary rocks (e.g. limestone, dolomite and marl) from a global lithological map (Hartmann & Moosdorf 2012). The percentage of this lithological class was calculated for each grid cell of the Iberian Peninsula to be included as a quantitative variable in the models. Assuming that no significant change on the Iberian distribution of continental rock lithology was produced during the last 21k years, we used the same lithological variable for the LGM projections. LGM climatic variables were obtained from Schmatz *et al.* (2015) in 30 arc seconds resolution according to four general circulation models (GCMs) pertaining to the Coupled Model Intercomparison Project (CMIP5: <http://cmip-pcmdi.llnl.gov/cmip5/>): CCSM, CNRM, IPSL and MIROC3.2.

The model was built based on 89 presence records, identified by (Moreira *et al.* 2015), which fall in 33 different 1 km² cells. As the distribution of this recently discovered species is restricted (Moreira *et al.* 2015), the spatial autocorrelation of the variables is high, thus we limited the number of variables to a maximum of three to avoid over-parameterization. To select the variables, we firstly performed a Pearson correlation analysis using a threshold of $r = |\pm 0.7|$. Then, we performed an Ecological Niche Factor Analysis (ENFA, Hirzel *et al.* 2002) with the preselected uncorrelated variables. ENFA computes factors accounting for the position of the occurrence data in the multidimensional environmental space of the study area. These factors describe the environmental niche of the species by computing the distance between the mean habitat for the species in relation to the study area (marginality) and the variance of the species' niche (specialization). Thus, ENFA can be an exploratory analysis to select the most relevant variables describing the niche of the species (see e.g. Chefaoui *et al.* 2015; Lobo *et al.* 2010).

To model the distribution of Vale da Couda individuals under current and LGM conditions we used Maxent (Phillips *et al.* 2006), a maximum entropy algorithm which uses presence and background data. This technique allows a “clamping” process, which handle predictors outside the training range as if they were at the limit. We selected ten times more background points than presences at random in order to set a prevalence of 0.1, as this proportion was used before with good results (e.g., Chefaoui *et al.* 2015; Chefaoui & Lobo 2008). We split data into a training (80%) and a test set (20%) to perform a cross-validation during 100 iterations. To validate the models, we obtained the area under the receiver operating characteristic (ROC) curve (AUC), the sensitivity (presences correctly predicted) and the specificity (absences correctly predicted) scores using three different thresholds for validation: the prevalence (= 0.1), the value which maximizes the sum of the sensitivity and specificity, and the highest threshold at which there is no omission. An ensemble of predictions was obtained for current conditions by computing the average of the 100 iterations. For LGM projection, we produced a hindcast using the average of the four GCMs. All analyses were performed in R (R Development Core Team 2013) using “adehabitat” and “dismo” packages.

2.4. Results

2.4.1 Population genetics

MtDNA sequence data of 73 putative *C. coudensis* individuals from Vale da Couda generated a 560-bp fragment alignment with a total of 142 polymorphic sites, 124 of which were parsimony informative. These polymorphisms defined 42 haplotypes with an overall haplotype diversity and mean nucleotide diversity of $h = 0.964 \pm 0.011$ and $\pi = 0.084 \pm 0.004$, respectively (Table 1A). These haplotypes were organized into four main divergent haplogroups, with 22 to 63 mutation steps apart (Figure 2.2A). Net sequence divergence between haplogroups ranged from 11.8 to 47.5%, while within net sequence divergence ranged from 0.1-2.1% (Supplementary information Figure S1). A large proportion of individuals (45%) possess unique haplotypes. The majority of haplotypes (88%) is found in only one

location (i.e. ‘private’ haplotypes), and only five haplotypes are shared among sites (12%). Despite the existence of these distinct haplogroups, there is no obvious phylogeographic pattern and no evidence for closely related haplotypes (i.e., same haplogroup) to come from the same location (Figures 2.2 and 2.3). High pairwise F_{ST} values between the three sites (site 4 was not used because it only has two individuals) vary between 0.03 (site 1 vs. site 3) and 0.07 (site 1 vs site 2), while D_{est} (Jost 2008) values vary between 0.88 (site 1 vs. site 3) and 0.96 (site 1 vs. site 2).

Nuclear DNA sequence data was only obtained from 35 individuals from Vale da Couda, generating a 503-bp fragment alignment with a total of 39 polymorphic sites, 17 of which were parsimony informative. The sequences defined 13 haplotypes with an overall haplotype diversity and mean nucleotide diversity of $h = 0.822 \pm 0.050$ and $\pi = 0.009 \pm 0.002$ respectively (Table 1A). These haplotypes constitute two haplogroups separated by 12 mutation steps (Figure 2.2B). Net sequence divergence between haplogroups was 2.3%, while within net sequence divergence ranged from 0.6-1.6%. Only 26% of the individuals have a unique haplotype. Of the total 13 haplotypes, 10 were private and three (23%) were shared between locations.

Table 2.1 Vale da Couda lineages, sample sizes and summary statistics for COI and ITS1 sequence fragments (A). Vale da Couda site sample sizes, lineages present and summary statistics for COI and ITS1 (B).

(A)

Lineages (L)	COI						ITS1					
	N	Nh	Locations	$h \pm s.d.$	$\pi \pm s.d.$	L	N	Nh	Locations	$h \pm s.d.$	$\pi \pm s.d.$	
A	8	3	1, 3	0.464 \pm 0.040	0.001 \pm 0.000	S	2	2	1,2	1.000	0.500	0.016 \pm 0.008
B	13	9	1, 2, 3, 4	0.936 \pm 0.051	0.018 \pm 0.002	R	33	11	1,2	0.799	0.054	0.006 \pm 0.001
C	7	6	2, 3, 4	0.952 \pm 0.096	0.010 \pm 0.096							
D	36	17	1, 2, 3	0.889 \pm 0.001	0.009 \pm 0.001							
-	8	6	1, 2, 3	0.893 \pm 0.111	0.013 \pm 0.003							

(B)

Locations	COI						ITS1					
	N	Nh	Lineages	$h \pm s.d.$	$\pi \pm s.d.$	L	N	Nh	Lineages	$h \pm s.d.$	$\pi \pm s.d.$	
1	31	18	A, B, D, -	0.927 \pm 0.031	0.075 \pm 0.008		10	6	R, S	0.778 \pm 0.137	0.011 \pm 0.005	
2	27	15	B, C, D, - A, B, C, D,	0.863 \pm 0.062	0.081 \pm 0.007		25	10	R, S	0.763 \pm 0.007	0.007 \pm 0.002	
3	13	12	-	0.987 \pm 0.035	0.091 \pm 0.009							
4	2	2	B, C	1.000 \pm 0.500	0.083 \pm 0.042							

L, Lineages; N, Sample size; Nh, number of haplotypes; h, haplotype diversity; π , nucleotide diversity; s.d., standard deviation

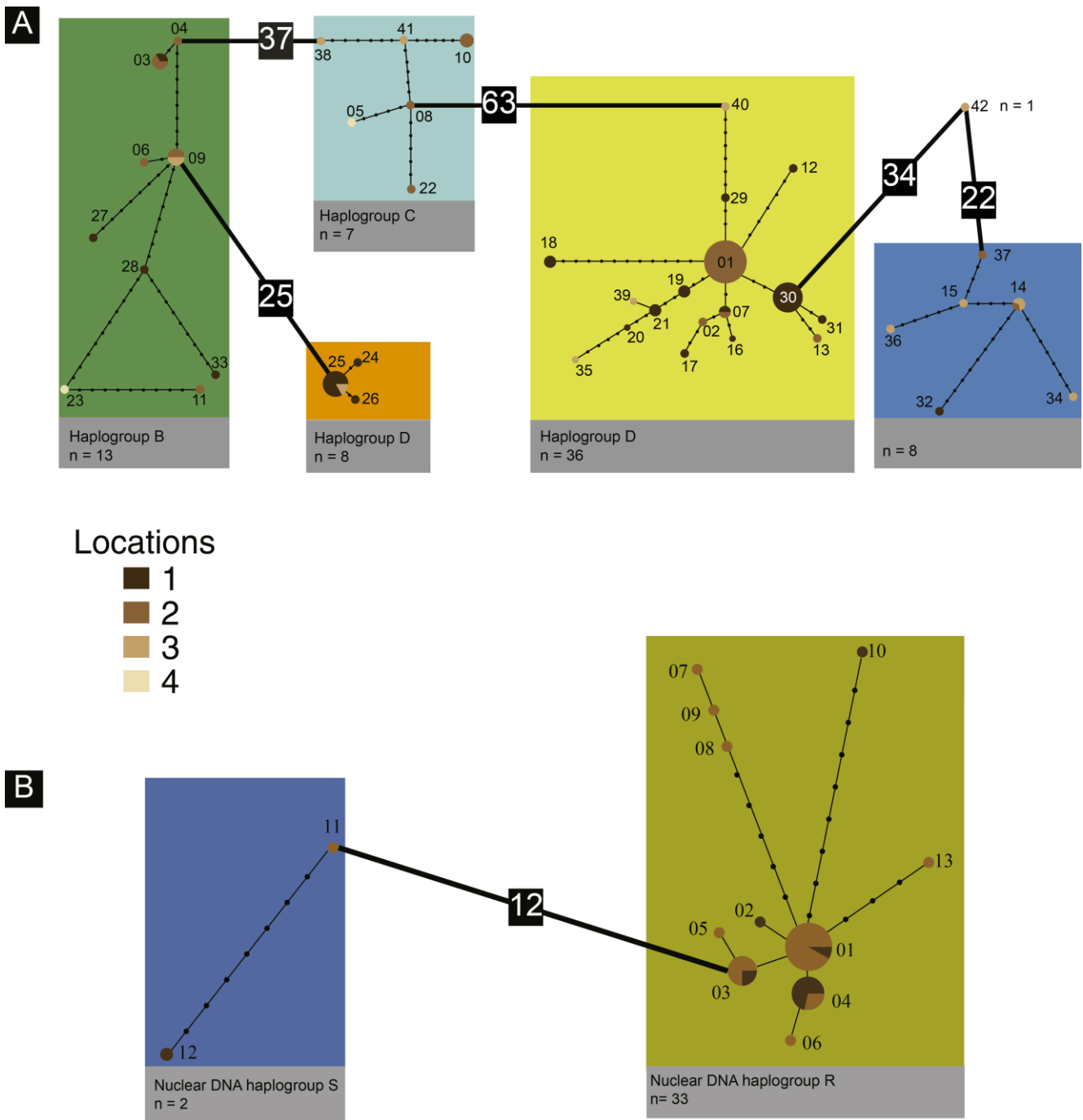


Figure 2.2 - (A) MtDNA COI statistical parsimony haplotype network for Vale da Couda individuals. (B) Nuclear ITS1 statistical parsimony haplotype network for Vale da Couda individuals. Each branch represents one inferred mutational step; small black circles on branches represent additional inferred mutational steps; numbers in black squares denote more than twenty mutations.

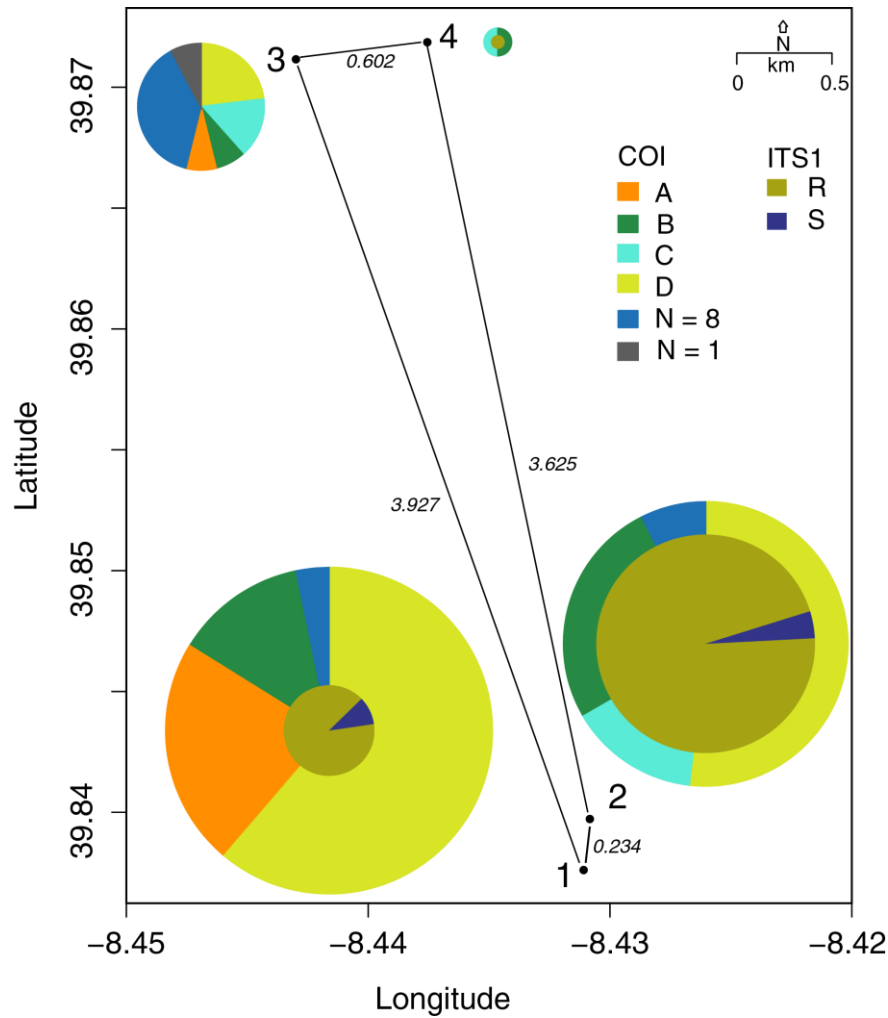


Figure 2.3 - Distribution of mtDNA lineages in Vale da Couda sites. Numbers represent sites; numbers in italic represent distance in km between sites. Size of circles is proportional to the number of individuals. Colors depicting haplogroups are the same as in Figure 2.2

MtDNA haplotypes are unevenly distributed among the four sampling sites. Sites 1 and 3 have representatives from all groups while site 2 has no representation of haplogroup C. In site 4 only haplogroups B and D are represented. The two ITS1 haplogroups are just present in two sampled sites, 1 and 2. We found no association between nuclear DNA and mtDNA haplogroups. ITS1 sequences are organized into two haplogroups (R and S), with disproportional representation among them (Figure 2.3). The mismatch analysis for Vale da Couda haplogroup D shows signs of historical demographic expansion (Figure 2.4). The frequency of pairwise differences between haplotypes of haplogroup D showed a distribution consisting of a unimodal curve, which combined with non-significant values of SSD ($p = 0.13$) and raggedness ($p = 0.08$) do not reject the null hypothesis of expansion. Adding to this, a low and significant value of R_2 ($p = 0.0035$) also supports the hypothesis of population growth for Vale da Couda haplogroup D. Since the population growth hypothesis is supported by the mismatch distribution we can use the estimated τ to calculate the time of the expansion, according to the formula $\tau = 2ut$, where t is time and u is the mutation rate for the gene in study. A population expansion of haplogroup D compatible with the LGM would require a mutation rate of ca. 15%/MY, which is consistent with the high-end mutation rates observed in some gastropods (Chiba 1999; Davison *et al.* 2009; Haase *et al.* 2003; Thomaz *et al.* 1996).

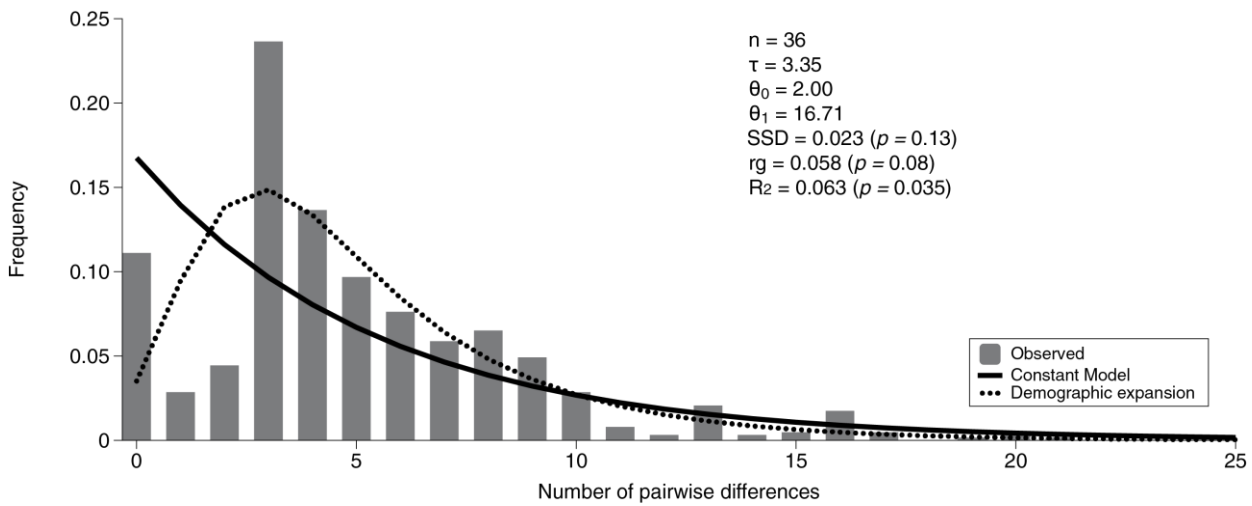


Figure 2.4 - Mismatch distribution of lineage D from Vale da Couda, indicating number of individuals in the analysis (n), Fu's FS test of selective neutrality and population expansion, evolutionary expansion age in mutational units (τ), effective population size before (θ_0) and after (θ_1) population expansion, and mean expansion time in units of thousand years (ka). Note that the range in expansion age corresponds to the 95% confidence interval of (s). SSD represents the sum of squares deviations, rg the raggedness statistics and R^2 .

2.4.2 Phylogenetic estimation

Results from the haplotype network seemed to indicate that the Vale da Couda individuals are not monophyletic, given the extreme genetic distance between haplogroups. The ML analysis ($-\ln L = -1622.11$) based on the COI data set yielded the topology depicted in Supplementary information Figure S2, and supports this idea. Specimens from Vale da Couda grouped into two main clades that did not cluster together. One clade included three haplogroups supported by high BP values (A, B, C). Haplogroup C grouped with specimens assigned to *C. olisippensis*. The other clade included haplogroup D and a group of eight specimens from Vale da Couda that showed an unresolved phylogenetic position. These specimens grouped with *C. setubalensis* from Arrábida (Supplementary information Figure S2).

2.4.3 Niche modelling

Eight uncorrelated climatic variables were used to perform ENFA analysis, which finally distinguished lithology, isothermality (BIO3), and the annual precipitation (BIO12) as the three most relevant variables defining the niche for Vale da Couda lineages (Table 2). ENFA marginality factor revealed that the lithology (grid cells with high percentage of carbonate sedimentary rocks) was the most relevant predictor of its distribution, an expectable result as the species has been found exclusively on limestone (Moreira et al 2015). Besides, ENFA showed that the species has a preference for locations where the isothermality and the annual precipitation are higher than the mean conditions of the Iberian Peninsula (Table 2).

Maxent models produced a strong discrimination between presence and background data regardless of the threshold used (Table 3). Overall validation scores of models calibrated under current conditions were: mean AUC = 0.981 ± 0.002 , mean sensitivity = 0.979 ± 0.017 , and mean specificity = 0.982 ± 0.012 (Table 3). The resulting ensemble for the current distribution showed two main areas with high probability of presence of Vale da Couda lineages: (1) one around the presently known distribution, and (2) different patches at the north of the Iberian Peninsula (Figure 2.5A). LGM projection indicates that past distribution of suitable habitats could have been wider, with also appropriate conditions in the Andalusian region and in a smaller area in the Central System (Figure 2.5B).

Table 2.2 - Environmental Niche Factor Analysis (ENFA) results showing marginality and specialization factors scores. The three variables with higher marginality scores (in bold) were selected for subsequent analyses.

Variable	Marginality	Specialization
bio1	0.19	0.40
bio3	0.47	0.00
bio7	-0.40	0.10
bio8	-0.10	0.05
bio12	0.45	0.07
bio9	0.07	-0.91
bio17	-0.10	0.03
lithology	0.59	0.00

Table 2.3 - Mean AUC, sensitivity and specificity scores obtained from the 100 Maxent models according to the three thresholds used. (Spec_sens: threshold that maximizes the sum of the sensitivity and specificity).

Model validation	Threshold	Mean \pm standard deviation
AUC	Prevalence	0.983 \pm 0.007
	No omission	0.980 \pm 0.019
	Spec_sens	0.980 \pm 0.018
Sensitivity	Prevalence	0.999 \pm 0.006
	No omission	0.971 \pm 0.037
	Spec_sens	0.969 \pm 0.038
Specificity	Prevalence	0.968 \pm 0.013
	No omission	0.989 \pm 0.020
	Spec_sens	0.990 \pm 0.010

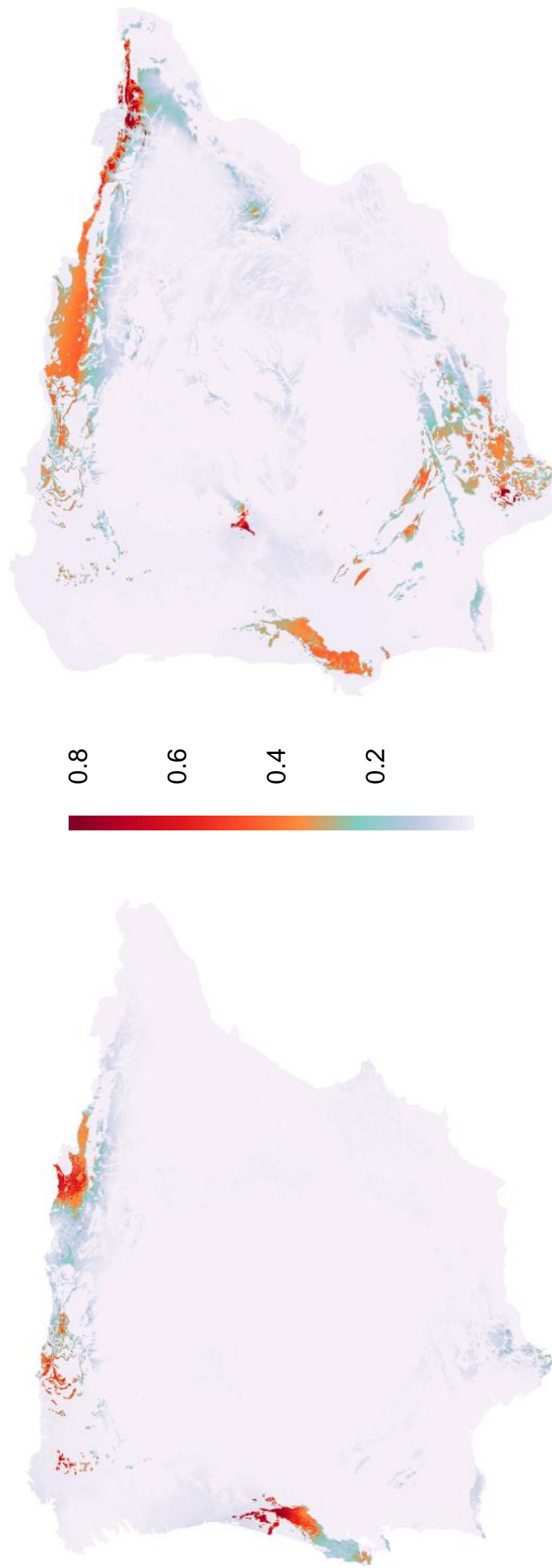


Figure 2.5 - Predicted geographic distribution built with presences of Vale da Couda individuals and based on (A) current climate, and (B) Last Glacial Maximum (LGM) conditions. Colour scale represents high probability of occurrence in red and low levels in blue. Black dots represent the present-day known occurrences in the Vale da Couda. Ice sheet existing during LGM in the north of the Iberian Peninsula is not depicted

2.5. Discussion

The mitochondrial marker used in this study produced a Vale da Couda complex phylogeny, with highly divergent and genetically diverse non-monophyletic lineages. Our results show a departure from the expectation that species with restricted distributions have low genetic variability, adding to a list of exceptions that keeps growing (e.g. Coates *et al.* 2006; Ellis *et al.* 2006; Gevaert *et al.* 2013; Young *et al.* 1996). The maintenance of diversity in rare species can be explained by the existence of a large effective population size (Ellstrand & Elam 1993), which is likely the case with *Candidula* from Vale da Couda that are usually locally abundant. The distribution models obtained for Vale da Couda lineages should be interpreted as regions with environmental conditions analogous to those where the species presently occurs (Pearson *et al.* 2007) and represent a probability of occurrence of the species.

2.5.1 Population genetics

Studies have shown populations with high genetic structure existing during the LGM in the Iberian Peninsula (Gómez & Lunt 2007). The particular geographical characteristics of this Peninsula (the existence of multiple mountain ranges with an east-west orientation creating a wide array of microclimatic changes, the influence of both the North Atlantic and the Mediterranean Sea, and existence of different climates ranging from Mediterranean, alpine, desert and Atlantic) foster the perfect conditions for the isolation of populations creating the “refugia within refugia” (Gómez & Lunt 2007). Even though the LGM distribution model suggests a larger distribution area for Vale da Couda lineages, it is possible that populations have endured geographical fragmentation at a micro-geographical level, compatible with the observed patterns of highly divergent lineages (Byrne *et al.* 2014). Due to effects of genetic drift in geographically limited species we would expect that our results showed lineages from Vale da Couda to be genetically depauperated but each sampled location displayed high levels of genetic diversity (Table 1B). The four highly divergent mtDNA clades found in Vale da Couda may have resulted from multiple colonizations by different populations of the same species that extended their distribution towards more southern locations during the LGM.

There are two main explanations for the maintenance of high genetic diversity in rare species: (1) large effective population sizes and/ or (2) gene flow between fragmented subpopulations. Large effective population sizes diminish the loss of genetic diversity due to genetic drift (Ellstrand & Elam 1993), and this is likely to be the case, as *Candidula* species tend to be locally abundant as suggested by recent surveys of Vale da Couda populations (Moreira *et al.* 2015). The existence of four highly divergent mtDNA lineages grouping in different clades suggests the absence of gene flow between populations.

2.5.2 Environmental niche modelling

Two distribution models were produced for specimens found in Vale da Couda: a present-day model and a LGM model (c. 20 ka). The LGM model shows a wider area that extends to the south with higher probability of occurrence compared with the present-day distribution of the individuals from Vale da Couda (Fig. 5B). This predicted distribution implies a co-occurrence between Vale da Couda lineages and other species of the genus (e.g. *C. setubalensis* and *C. olissipensis*) currently occupying these southern locations. The differences between the paleo-model and the contemporary model are somewhat unexpected considering that most of the northern hemisphere terrestrial organisms have contracted their geographic distributions to the south during the harsher glacial climate conditions, and have expanded their distribution by re-colonizing former northern territories after deglaciation (Hewitt 1999). However, mountainous regions of the north of the Iberian Peninsula (i.e. Pyrenees and Cantabrian Range) are known to have been covered by ice during Pleistocene glaciations, though the precise position of the ice sheet in the LGM remains uncertain (see e.g. Palacios *et al.* 2015). Thus, most of those northern regions

found suitable by our LGM model could not have been occupied by these terrestrial land snails because of the existent ice sheet before deglaciation.

According to ENFA results (Table 2) the present-day distribution of lineages from Vale da Couda is mainly driven by the presence of carbonate-dominated lithological units under rainy and isothermal climatic conditions. These specific requirements seem to be in agreement with those shown by other terrestrial mollusc species (Hermida *et al.* 2000; Kadmon & Heller 1998; Tattersfield *et al.* 2001; Tsoar *et al.* 2007).

Given the putative low dispersal capacity of this group, the most plausible hypothesis is that during Quaternary glaciations Vale da Couda lineages might have dispersed towards suitable habitat located in south-central Portugal (Lisbon and northeast of Lisbon, including Leiria), where LGM hindcast suggests appropriate conditions for these organisms. A postglacial change of climatic conditions towards lower precipitation in the Lisbon area may have caused its contraction to its actual distribution using the suitable Mesozoic calcareous rock as a corridor. Although we have addressed some common hindcasting uncertainties by using different GCMs and a clamping mask hindcast approach, we could not resolve the lack of accurate lithological data for emerged coastal land in the LGM. Thus, further appropriate habitats not depicted in our models could have existed in regions near the coast.

2.5.3 Biogeographic scenario

The putative past and contemporary models of distribution of Vale da Couda lineages are compatible with a biogeographic LGM scenario of species co-existence in more southern locations followed by a northern dispersal tracking the species optimal thermal, humidity and soil physical conditions. This co-existence is plausible given the fact that *C. setubalensis* and *C. arrabidensis* occur in sympatry, as well as *C. olisippensis* and *C. coudensis* (Holyoak & Holyoak 2014) and share habitat requirements with Vale da Couda lineages. Moreover, Roucoux (2001) shows low but fluctuating tree pollen through the LGM, along with abundant grass and some herb pollen, indicating likely widespread suitability of the grassy habitats for *Candidula* species throughout the LGM. Similar events have already been detected for other land snail species (Harl *et al.* 2014; Sauer & Hausdorf 2010; Shimizu & Ueshima 2000). After the LGM, environmental conditions during deglaciation were such that promoted northward dispersal of land snails and the establishment of populations in locations of suitable isothermality and precipitation like Vale da Couda. *C. setubalensis* and *C. arrabidensis* maintained a southern distribution, in the Setubal Peninsula. Specifically, we hypothesize that Pleistocene conditions may have isolated populations into pockets of suitable habitats in more southern locations, which promoted population differentiation and intra-specific diversification without apparent geological barriers.

2.5.4 Taxonomic implications

Divergence between COI sequences has been used as a tool to identify new species based on the premise that there is a “DNA barcoding gap”, meaning that the genetic differences between intra- and inter-specific individuals do not overlap, with the latter being bigger than the former (Bucklin *et al.* 2011). Several studies have shown that high levels of mtDNA divergence are extremely common within land snail species, reaching values as high as 30% (Chiba 1999; Davison *et al.* 2009; Haase *et al.* 2003; Thomaz *et al.* 1996), and values between species going as low as 1% (Davison *et al.* 2009). These findings preclude the application of the DNA barcoding approach to land snails because there is no known “barcoding gap” and as such species cannot be discriminated solely based on a threshold value of sequence divergence. Nevertheless, a conservative approach to our results imply the existence in Vale da Couda of at least two species, represented by the two most divergent lineages (Supplementary information Figure S2). These lineages are likely to display the highest genetic variability of the two species that according to Holyoak (2014) coexist in that region, *C. coudensis* and *C. olisippensis*. As

discussed before, patterns of high mitochondrial divergence can be attributed to the maintenance of ancient polymorphisms from past isolations (Dillon Jr & Robinson 2009) combined with the effects of present large overall population size and gene flow between fragmented sub-populations. It was not our goal to produce a full molecular taxonomy of the genus *Candidula*. Our intent was to stress the high genetic diversity present in a narrow geographic range and propose a plausible biogeographic scenario for the haplogroups encountered. Our results prompt for an in-depth investigation on the molecular taxonomy of the genus on the basis of multiple mitochondrial and nuclear markers that will enable to disentangle the complexity of the genus *Candidula* in Portugal but which is out of the scope of the present study.

2.6. Conclusions

The genetic survey presented here revealed the existence of four main mitochondrial lineages in Vale da Couda (previously attributed to a single species) with independent evolutionary histories and exhibiting extremely narrow geographic ranges. These results do not corroborate previous morphological studies that considered the existence of a single species, *Candidula coudensis* in the studied area. LGM hindcasts revealed the existence of putative glacial refugia south of the current distribution of the lineages of Vale da Couda. Although unexpected in organisms with such a restricted distribution, the high genetic diversity found in Vale da Couda lineages adds to the growing list of exceptions where low-dispersal species show high levels of genetic variability. These findings have implications for the understanding of the genetic characteristics of rare and endemic species. From a conservation perspective, Vale da Couda lineages do not seem to be endangered, with high genetic diversity within and between lineages maintained by putative large effective population sizes.

2.7. References

- Akaike H. 1974. A new look at the statistical model identifications. *IEEE Transactions on automatic control* 19:716–723.
- Austin MP, Belbin L, Meyers JA, Doherty MD, and Luoto M. 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. *Ecological Modelling* 199:197-216. 10.1016/j.ecolmodel.2006.05.023
- Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, Massachussets: Harvard University Press.
- Bandelt HJ, Forster P, and Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
- Bird CE, Karl SA, Smouse PE, and Toonen RJ. 2011. Detecting and measuring genetic differentiation. In: Held C, Koenemann S, and Schubart C, eds. *Phylogeography and population genetics in Crustacea*. Boca Raton, Florida, USA: CRC Press, Taylor & Francis Group, 31-55.
- Bucklin A, Steinke D, and Blanco-Bercial L. 2011. DNA Barcoding of marine metazoa. *Annual Review of Marine Science* 3:471-508. doi:10.1146/annurev-marine-120308-080950
- Byrne M, Coates DJ, Forest F, Hopper SD, Krauss SL, Sniderman JK, and Thiele KR. 2014. A diverse flora—species and genetic relationships. *Plant Life on the Sandplains in Southwest Australia, a Global Biodiversity Hotspot*. Crawley, Australia: UWA Publishing, 81-99.
- Chang C-T, Tsai C-N, Tang CY, Chen C-H, Lian J-H, Hu C-Y, Tsai C-L, Chao A, Lai C-H, and Wang T-H. 2012. Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. *The Scientific World Journal* 2012.

- Chefaoui RM, Assis J, Duarte CM, and Serrão EA. 2015. Large-scale prediction of seagrass distribution integrating landscape metrics and environmental factors: The case of *Cymodocea nodosa* (Mediterranean–Atlantic). *Estuaries and Coasts* 39:123-137. 10.1007/s12237-015-9966-y
- Chefaoui RM, and Lobo JM. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 210:478-486. 10.1016/j.ecolmodel.2007.08.010
- Chiba S. 1999. Accelerated evolution of land snails *Mandarina* in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. *Evolution*:460-471. 10.2307/2640782
- Coates DJ, Tischler G, and McComb JA. 2006. Genetic variation and the mating system in the rare *Acacia sciophanes* compared with its common sister species *Acacia anfractuosa* (Mimosaceae). *Conservation Genetics* 7:931-944. 10.1007/s10592-006-9136-7
- Davison A, Blackie RL, and Scothern GP. 2009. DNA barcoding of stylommatophoran land snails: a test of existing sequences. *Molecular ecology resources* 9:1092-1101. 10.1111/j.1755-0998.2009.02559.x
- Dillon Jr RT, and Robinson JD. 2009. The snails the dinosaurs saw: are the pleurocerid populations of the Older Appalachians a relict of the Paleozoic Era? *Journal of the North American Benthological Society* 28:1-11. 10.1899/08-034.1
- Doyle J, and Doyle JL. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* 19:11-15.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797. 10.1093/nar/gkh340
- Elith J, and Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40:677-697. 10.1146/annurev.ecolsys.110308.120159
- Ellis J, Pashley C, and McCauley D. 2006. High genetic diversity in a rare and endangered sunflower as compared to a common congener. *Molecular ecology* 15:2345-2355. 10.1111/j.1365-294X.2006.02937.x
- Ellstrand NC, and Elam DR. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* 24:217-242.
- Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* 10:564-567. 10.1111/j.1755-0998.2010.02847.x
- Folmer OM, Black W, Hoeh R, Lutz R, and Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology & Biotechnology* 3:294–299.
- Frankham R. 2005. Genetics and extinction. *Biological Conservation* 126:131-140. 10.1016/j.biocon.2005.05.002
- Gaston KJ. 1994. Causes of rarity. *Rarity*. Amsterdam: Kluwer Academic Publishers, 115–135.
- Gevaert SD, Mandel JR, Burke JM, and Donovan LA. 2013. High genetic diversity and low population structure in Porter's sunflower (*Helianthus porteri*). *Journal of Heredity* 104:407-415. 10.1093/jhered/est009
- Gómez A, and Lunt DH. 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia: Evolutionary perspectives on the origins and conservation of European biodiversity*. Dordrecht: Springer Netherlands, 155-188.
- Guindon S, eacute, phane, and Gascuel O. 2003. A simple, fast and accurate algorithm to estimate large phylogenies by Maximum Likelihood. *Systematic Biology* 52:696 – 704. 10.1080/10635150390235520

- Haase M, Misof B, Wirth T, Baminger H, and Baur B. 2003. Mitochondrial differentiation in a polymorphic land snail: evidence for Pleistocene survival within the boundaries of permafrost. *Journal of Evolutionary Biology* 16:415-428. 10.1046/j.1420-9101.2003.00542.x
- Harl J, Páll-Gergely B, Kirchner S, Sattmann H, Duda M, Kruckenhauser L, and Haring E. 2014. Phylogeography of the land snail genus *Orcula* (*Orculidae*, *Stylommatophora*) with emphasis on the Eastern Alpine taxa: speciation, hybridization and morphological variation. *BMC evolutionary biology* 14:1-26. 10.1186/s12862-014-0223-y
- Harpending H. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66:591-600.
- Hartmann J, and Moosdorf N. 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems* 13. 10.1029/2012GC004370
- Hayashi M, and Chiba S. 2000. Intraspecific diversity of mitochondrial DNA in the land snail *Euhadra peliomphala* (Bradybaenidae). *Biological Journal of the Linnean Society* 70:391-401. 10.1111/j.1095-8312.2000.tb01230.x
- Hermida J, Ondina M, and Rodriguez T. 2000. The relative importance of edaphic factors on the distribution of some terrestrial gastropod species: autecological and synecological approaches. *Acta Zoologica Academiae Scientiarum Hungaricae* 46:265-274.
- Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68:87-112. 10.1111/j.1095-8312.1999.tb01160.x
- Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 359:183-195. 10.1098/rstb.2003.1388
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, and Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology* 25:1965-1978. 10.1002/joc.1276
- Hillis DM, and Dixon MT. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology* 66:411-453.
- Hirzel AH, Hausser J, Chessel D, and Perrin N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83:2027-2036. 10.2307/3071784
- Holyoak DT, and Holyoak GA. 2014. A review of the genus *Candidula* in Portugal with notes on other populations in Western Europe (*Gastropoda*, *Pulmonata*, *Hygromiidae*) *Journal of Conchology* 41:629-672.
- Jost L. 2008. GST and its relatives do not measure differentiation. *Molecular ecology* 17:4015-4026. 10.1111/j.1365-294X.2008.03887.x
- Kadmon R, and Heller J. 1998. Modelling faunal responses to climatic gradients with GIS: land snails as a case study. *Journal of Biogeography* 25:527-539. 10.1046/j.1365-2699.1998.2530527.x
- Katoh K, and Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30:772-780. 10.1093/molbev/mst010
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, and Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-1649. 10.1093/bioinformatics/bts199
- Keenan K, McGinnity P, Cross TF, Crozier WW, and Prodöhl PA. 2013. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* 4:782-788. 10.1111/2041-210X.12067

- Köhler F, and Burghardt I. 2015. Cryptic diversity in a widespread land snail: revision of the genus *Xanthomelon* Martens, 1860 from the Australian Monsoon Tropics (Pulmonata, Camaenidae). *Zoologica Scripta* 45:127-144. 10.1111/zsc.12144
- Laikre L, Nilsson T, Primmer CR, Ryman N, and Allendorf FW. 2009. Importance of genetics in the interpretation of favourable conservation status. *Conservation Biology* 23:1378-1381. 10.1111/j.1523-1739.2009.01360.x
- Librado P, and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452. 10.1093/bioinformatics/btp187
- Lobo JM, Jiménez-Valverde A, and Hortal J. 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103-114. 10.1111/j.1600-0587.2009.06039.x
- Meirmans PG, and Hedrick PW. 2011. Assessing population structure: F(ST) and related measures. *Molecular ecology resources* 11:5–18. Doi 10.1111/J.1755-0998.2010.02927.X
- Moreira F, Calado G, and Dias S. 2015. Conservation status of a recently described endemic land snail, *Candidula coudensis*, from the Iberian Peninsula. *PLoS ONE* 10:e0138464. 10.1371/journal.pone.0138464
- Nei M. 1987. Genetic distance and molecular phylogeny. In: Ryman N, and Utter FW, eds. *Population Genetics & Fishery Management*. Seattle: Washington Sea Grant Program, University of Washington, 193–223.
- Nei M, and Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97:145–163.
- Palacios D, de Andrés N, López-Moreno JJ, and García-Ruiz JM. 2015. Late Pleistocene deglaciation in the upper Gállego Valley, central Pyrenees. *Quaternary Research* 83:397-414. 10.1016/j.yqres.2015.01.010
- Pearson RG, Raxworthy CJ, Nakamura M, and Townsend Peterson A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34:102-117. 10.1111/j.1365-2699.2006.01594.x
- Pfenninger M, Nowak C, and Magnin F. 2007. Intraspecific range dynamics and niche evolution in *Candidula* land snail species. *Biological Journal of the Linnean Society* 90:303-317. 10.1111/j.1095-8312.2007.00724.x
- Pfenninger M, Posada D, and Magnin F. 2003. Evidence for survival of Pleistocene climatic changes in northern refugia by the land snail *Trochoidea geyeri* (Soós 1926) (*Helicellinae*, *Stylommatophora*). *BMC evolutionary biology* 3:1-13. 10.1186/1471-2148-3-8
- Phillips S, Anderson R, and Schapire R. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259. 10.1016/j.ecolmodel.2005.03.026
- Pinceel J, Jordaens K, Pfenninger M, and Backeljau T. 2005. Rangewide phylogeography of a terrestrial slug in Europe: evidence for Alpine refugia and rapid colonization after the Pleistocene glaciations. *Molecular ecology* 14:1133–1150. 10.1111/j.1365-294X.2005.02479.x
- Primack RB. 2006. *Essentials of conservation biology*. Sunderland, Massachusetts: Sinauer Associates.
- R Development Core Team. 2013. R: A language and environment for statistical computing.
- Ramos-Onsins SE, and Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092–2100.
- Rogers AR, and Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552–569.
- Roucoux KH, Shackleton NJ, de Abreu L, Schönfeld J, and Tzedakis PC. 2001. Combined marine proxy and pollen analyses reveal rapid Iberian vegetation response to North Atlantic millennial-scale climate oscillations. *Quaternary Research* 56:128-132. 10.1006/qres.2001.2218

- Ryman N, and Leimar O. 2009. GST is still a useful measure of genetic differentiation - a comment on Jost's D. *Molecular ecology* 18:2084-2087.
- Sambrook JF, and Russell DW. 2001. *Molecular cloning: a laboratory manual*: Cold spring harbor laboratory press, New York.
- Sauer J, and Hausdorf B. 2010. Reconstructing the evolutionary history of the radiation of the land snail genus *Xerocrassa* on Crete based on mitochondrial sequences and AFLP markers. *BMC evolutionary biology* 10:1-13. 10.1186/1471-2148-10-299
- Schmatz D, Luterbacher J, Zimmermann N, and Pearman P. 2015. Gridded climate data from 5 GCMs of the Last Glacial Maximum downscaled to 30 arc s for Europe. *Climate of the Past Discussions* 11:2585-2613.
- Schneider S, and Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152:1079-1089.
- Shimizu Y, and Ueshima R. 2000. Historical biogeography and interspecific mtDNA introgression in *Euhadra peliomphala* (the Japanese land snail). *Heredity* 85:84-96. 10.1046/j.1365-2540.2000.00730.x
- Tamura K, and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512-526.
- Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725-2729. 10.1093/molbev/mst197
- Tattersfield P, Warui C, Seddon M, and Kiringe J. 2001. Land-snail faunas of afro-montane forests of Mount Kenya, Kenya: ecology, diversity and distribution patterns. *Journal of Biogeography* 28:843-861. 10.1046/j.1365-2699.2001.00606.x
- Thomaz D, Guiller A, and Clarke B. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society of London B: Biological Sciences* 263:363-368. 10.1098/rspb.1996.0056
- Tsoar A, Allouche O, Steinitz O, Rotem D, and Kadmon R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions* 13:397-405. 10.1111/j.1472-4642.2007.00346.x
- Weir BS, and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370. 10.2307/2408641
- White TJ, Bruns T, Lee S, and Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* 18:315-322.
- Whitlock MC. 2011. G_{ST} and D do not replace F_{ST} . *Molecular ecology* 20:1083-1091. 10.1111/j.1365-294X.2010.04996.x
- Young A, Boyle T, and Brown T. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends in ecology & evolution* 11:413-418. 10.1016/0169-5347(96)10045-8

3. Final remarks

This chapter concludes the thesis, therefore I will present a summary of the main findings interspersed with the initial questions and hypothesis that prompted the present work.

How much diversity is there in the *Candidula* populations of Vale da Couda?

Contrary to theoretical expectations, the individuals of *Candidula* from Vale da Couda exhibit high genetic diversity, with four non-monophyletic lineages present in the region. Our expectation relied on the assumption that such a restricted geographic distribution would comprehend one single genetic lineage, corresponding to the previously identified endemism *Candidula coudensis*.

Is *Candidula coudensis* a monophyletic clade within the *Candidula* genus in Portugal?

There are four highly divergent mtDNA haplogroups suggesting an intricate biogeographic scenario possibly driven by the simultaneous action of the following factors: (1) allopatric divergence during the LGM followed by secondary contact in Vale da Couda haplotypes (Thomaz *et al.* 1996) ; (2) rapid mtDNA evolution promoted by high mutation rates (Thomaz *et al.* 1996); (3) natural selection acting to preserve variation (Thomaz *et al.* 1996) and (4) population structure of land snails favors the maintenance of ancient haplotypes (Thomaz *et al.* 1996, Dillon Jr & Robinson 2009).

Is there a range shift in the putative distribution of *Candidula* during the LGM in comparison to present-day distributions?

Both models of geographic distribution of Vale da Couda lineages, past and contemporary, support a larger co-existence in more southern locations with individuals trailing after their optimal thermal, humidity and soil physical conditions in northern locations.

What is the demographic history of Vale da Couda individuals?

The mismatch analysis showed signs of population expansion, compatible with a post-LGM period when considering a high mutation rate of ca. 15%MY. This is consistent with a scenario of population expansion towards the present-day known geographical distribution.

The present-day existence of different genetic lineages in Vale da Couda is compatible with a scenario in which Pleistocene conditions have promoted genetic divergence without obvious geological barriers, by having compartmentalized populations into appropriate habitats. From a conservation perspective, Vale da Couda lineages do not seem to be endangered, with high genetic diversity within and between lineages maintained by putative large effective population sizes.

Although the present work does not deal with taxonomy *per se*, the detection of four genetically distant haplogroups, deserves consideration. Our treatment of Vale da Couda clades as separate units is not an endorsement of their recognition as distinct species but simply the recognition that the subject merits further integrative studies to assess their taxonomic status. A more intensive analysis, including denser geographical sampling, more high-resolution molecular markers, together with paleo-niche model approaches we produced, are needed to fully evaluate the taxonomic questions raised by this work, regarding the Vale da Couda population.

The hypotheses formulated for *C. coudensis* were based on *a priori* theoretical premises likely to shape the phylogeographic structure of terrestrial species with low dispersal abilities and restricted geographical distributions. Our results (1) do not support the existence of a single monophyletic clade in Vale da Couda; (2) indicate high levels of haplotype and nucleotide diversities, not consistent with

isolated population on a limited geographical area and (3) confirm that population from Vale da Couda may have undergone demographic expansion after the LGM, similarly to many other terrestrial species.

3.1. References

- Dillon Jr RT, and Robinson JD. 2009. The snails the dinosaurs saw: are the pleurocerid populations of the Older Appalachians a relict of the Paleozoic Era? *Journal of the North American Benthological Society* 28:1-11. 10.1899/08-034.1
- Thomaz D, Guiller A, and Clarke B. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society of London B: Biological Sciences* 263:363-368. 10.1098/rspb.1996.0056

4. Appendix

Table 4.S1 - Sample location and summary statistics for the genus *Candidula*.

Location	Code	Long	Lat	<i>N</i>	<i>nh</i>	<i>h</i> ± <i>s.d.</i>			π ± <i>s.d.</i>		
Arrábida	AR	38.5	-9.0	17	9	0.860	±	0.068	0.108	±	0.009
Azoia, Espichel	AZ	38.4	-9.2	3	3	1.000	±	0.272	0.105	±	0.047
Berlengas	BE	37.2	-8.7	7	2	0.571	±	0.119	0.004	±	0.000
Bensafrim	BS	39.4	-9.4	4	3	0.833	±	0.222	0.116	±	0.054
Farilhão	FA	39.5	-9.5	1	1	0.000			-		-
Praia do Abano	AB	38.7	-9.5	10	9	0.978	±	0.054	0.025	±	0.006
Vale da Couda	VC	39.8	-8.4	73	42	0.964	±	0.011	0.084	±	0.004
Total				115	69	0.972	±	0.009	0.122	±	0.004

N: number of individuals; *Nh*: number of haplotypes; *h*: haplotype diversity; π : nucleotide diversity; *s.d.*: standard deviation.

Table 4.S2 - Bioclimatic variables for current conditions retrieved from WorldClim dataset (Hijmans *et al.* 2005).

Code	Bioclimatic variables	Initial climatic factor
Bio1	Annual Mean Temperature	T
Bio2	Mean Diurnal Range [Mean of monthly (max temp - min temp)]	T
Bio3	Isothermality [(Bio2/Bio7)*100]	T
Bio4	Temperature Seasonality [standard deviation *100]	T
Bio5	Max Temperature of Warmest Month	T
Bio6	Min Temperature of Coldest Month	T
Bio7	Temperature Annual Range [Bio5-Bio6]	T
Bio8	Mean Temperature of Wettest Quarter	T + R
Bio9	Mean Temperature of Driest Quarter	T + R
Bio10	Mean Temperature of Warmest Quarter	T
Bio11	Mean Temperature of Coldest Quarter	T
Bio12	Annual Precipitation	R
Bio13	Precipitation of Wettest Month	R
Bio14	Precipitation of Driest Month	R
Bio15	Precipitation Seasonality [Coefficient of Variation]	R
Bio16	Precipitation of Wettest Quarter	R
Bio17	Precipitation of Driest Quarter	R
Bio18	Precipitation of Warmest Quarter	T + R
Bio19	Precipitation of Coldest Quarter	T + R

T, Temperature ; R, Rainfall ; T + R, Both

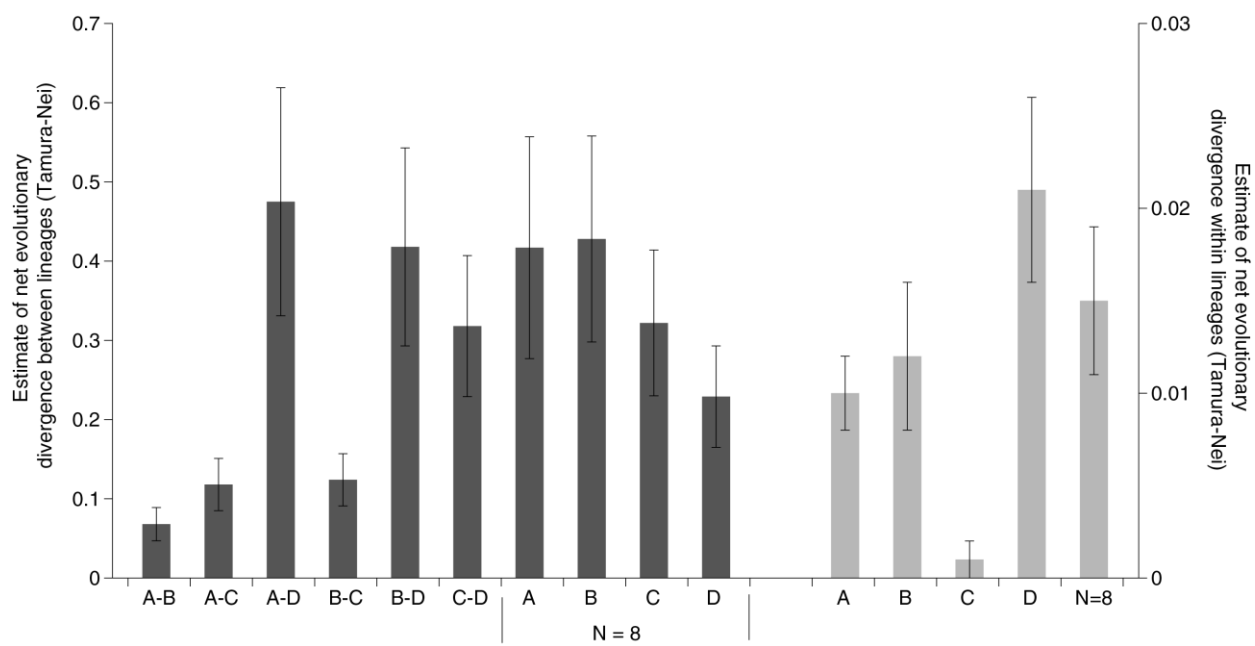


Figure 4.S1 - Estimates of net evolutionary divergence between haplogroups (axis on the left, dark grey bars \pm standard deviation) and within lineages (axis on the right, light grey bars \pm standard deviation), based on Tamura-Nei distances.

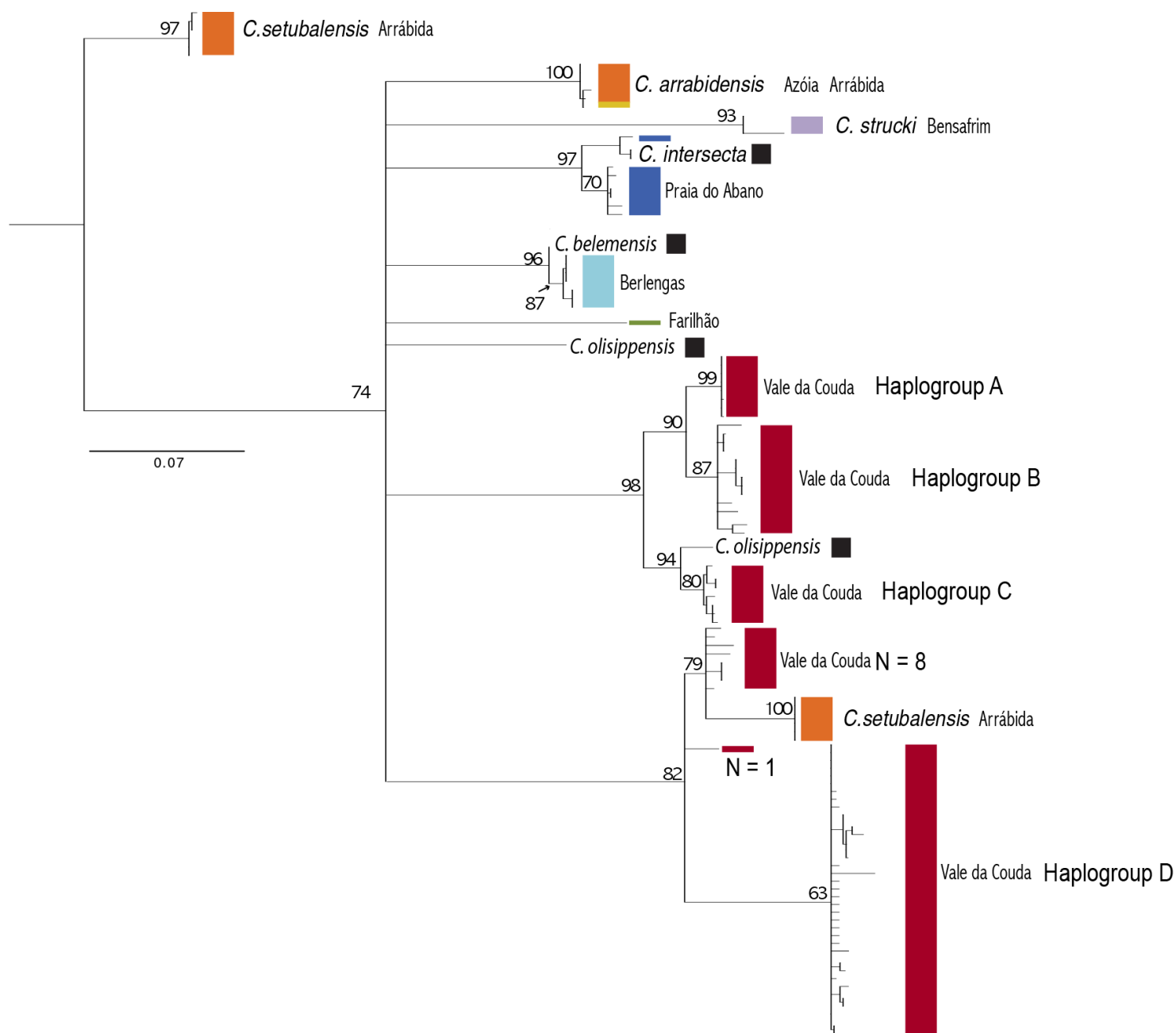


Figure 4.S2 - Phylogenetic relationships between *Candidula* individuals from Vale da Couda (in red) and other locations in Portugal. Black squares represent individuals identified morphologically and anatomically. Colours of locations as in Figure 1.1. Outgroups removed from figure for illustrative purposes.