A peer-reviewed version of this preprint was published in PeerJ on 16 March 2017.

Madeira PM, Chefaoui RM, Cunha RL, Moreira F, Dias S, Calado G, Castilho R. 2017. High unexpected genetic diversity of a narrow endemic terrestrial mollusc. PeerJ 5:e3069
https://doi.org/10.7717/peerj.3069

High unexpected genetic diversity of a narrow endemic terrestrial mollusc

Pedro M Madeira [1], **Rosa M. Chefaoui** [1], **Regina L Cunha** [1], **Francisco Moreira** [2, 3, 4], **Susana Dias** [2, 4], **Gonçalo Calado** [2], **Rita Castilho** Corresp. [1]

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

[2] Departamento de Ciências da Vida, Escola de Psicologia e Ciências da Vida, Universidade Lusófona, Campo Grande, 376, 1749-024 Lisboa, Portugal

[3] REN Biodiversity Chair, CIBIO/InBIO Associate Laboratory, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal

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

Corresponding Author: Rita Castilho
Email address: rcastil@ualg.pt

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 analyse phylogeographic patterns in a group of endemic land snails of the genus *Candidula* from a valley of central Portugal (Vale da Couda) showing an exceptionally narrow distributional range. Phylogenetic analyses recovered Vale da Couda specimens in two main clades that do not share a common ancestry. Considering the restricted geographic distribution, an unusual high number of haplotypes was found. These haplotypes were 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.

High-unexpected genetic diversity of a narrow endemic terrestrial mollusc

Madeira, P.M.[1], Chefaoui, R.M.[1], Cunha, R. L. [1], Moreira, F. [2,3,4], Dias, S. [2,4], Calado, G. [2] and Castilho, R.[1,*]

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

[2]Departamento de Ciências da Vida, Escola de Psicologia e Ciências da Vida, Universidade Lusófona, Campo Grande, 376, 1749-024 Lisboa – Portugal

[3]REN Biodiversity Chair, CIBIO/InBIO Associate Laboratory, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal

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

E-mail address: rcastil@ualg.pt (R. Castilho).

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 analyse phylogeographic patterns in a group of endemic land snails of the genus *Candidula* from a valley of central Portugal (Vale da Couda) showing an exceptionally narrow distributional range. Phylogenetic analyses recovered Vale da Couda specimens in two main clades that do not share a common ancestry. Considering the restricted geographic distribution, an unusual high number of haplotypes was found. These haplotypes were 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.

Introduction

Phylogeography combines evidence from both population genetics and phylogeny, to understand the evolutionary processes that shape geographic population structure (Avise 2000). These evolutionary processes include divergence between 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 species distribution patterns is also essential to understand the species dynamics in space and time. Species distribution modeling (SDM) allows relating statistically the identified presence records of a 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

65 geographical range of a species (Elith & Leathwick 2009). Recently, there has been a growing
66 trend towards the integration of SDM hindcasts with phylogeography as a useful approach to
67 obtain consistent eco-evolutionary hypotheses. This combination allows getting insights how the
68 distribution of climatic refuges and postglacial colonization pathways may have influenced
69 genetic diversity of current populations (see e.g. Hewitt 2004).
70
71 Land snails are good models for evolutionary studies, since phylogeographic patterns are often
72 preserved due to their limited dispersal capabilities and specific habitat requirements (Pfenninger
73 et al. 2007). Also, snails display an unusual high intraspecific genetic variation, ca. 10-30% in
74 mtDNA sequences (Bond et al. 2001; Hayashi & Chiba 2000; Pinceel et al. 2005; Shimizu &
75 Ueshima 2000), which renders the taxa appropriate to understand processes shaping the
76 partitioning of genetic variation in space. Additionally, many land snail examples in the literature
77 show the existence of cryptic species in sympatry (Köhler & Burghardt 2015).
78
79 The land snails of the genus *Candidula* present in Europe, from eastern Canary Islands to the
80 Balkans and northwards to Scotland and southernmost Sweden are represented by 24 putative
81 species. Portugal has 8 endemic species (*C. coudensis*, *C. setubalensis, C. scabiosula, C.*
82 *arrabidensis, C. belemensis, C. carrapateirensis, C. codia* and *C. strucki*) from a total of 12 (*C.*
83 *gigaxii, C. intersecta, C. ponsulensis, C. olisippensis*) (Holyoak & Holyoak 2014) (see Figure 1
84 for details on geographic distribution). Most species are hard to distinguish using only
85 conchological characters and it takes a combination of morphological characters, such as the size
86 of the penial flagellum or shell shape, to classify the specimens (Holyoak & Holyoak 2014).
87 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) (see Figure 1A).

Candidula coudensis (*Candidula coudensis* Holyoak & Holyoak, 2010) is a recently described endemismwith a highly restricted geographic distribution of ca. 13.5 km2 in Vale da Couda, Leiria, Portugal (inset Figure 1) (Moreira et al. 2015). 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: (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) the population from Vale da Couda may show reduced levels of haplotype and nucleotide diversities, consistent with isolated population on a limited geographical area, and

(3) the population from Vale da Couda 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) reconstructing the phylogenetic relationship of *C. coudensis* regarding other Portuguese *Candidula* species from different locations; (2) revealing the genetic diversity and geographic structure of contemporary *C. coudensis*, and (3) reconstructing the demographic history of the species. 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 *C. coudensis* that can assist future fieldwork.

Material and Methods

Taxon sampling. Sampling in Vale da Couda resulted in 73 individuals collected from 4 different sites (Figure 1). In order to place the Vale da Couda samples in a broader phylogenetic context (see Phylogenetic estimation section below) a few *Candidula spp.* individuals were collected in different locations (Table 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%. 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 95ºC, 35 cycles of 30 seconds at 94ºC, 30 seconds at 53ºC followed by an extension for 1 minute at 72ºC 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 97ºC, 35 cycles of 1 minute at 95ºC, 1 minute at 55ºC and 2 minutes at 72º, followed by a final extension of 5 minutes at 72ºC. 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).

Phylogenetic estimation. In this study we used phylogenetic inference with the sole purpose to ascertain the monophyly of the individuals from Vale da Couda. It is not our intention to produce

a phylogeny for the genus *Candidula*. We followed Holyoak and Holyoak (2014) taxonomy to identify 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).

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 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 (nuDNA) 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 *C. coudensis* 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 $T\text{exp} = \tau / (2\ \mu\ n)$, 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 1 year (Pfenninger et al. 2003).

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. 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.

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 (n=89) 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.

Results

Phylogenetic estimation. The ML analysis (- ln L = -1622.11) based on the COI data set yielded the topology depicted in Figure 2. Specimens from Vale da Couda grouped into two main clades that did not cluster together. One clade included three lineages supported by high BP values (A, B, C). Lineage I grouped with specimens assigned to *C. olisippensis*. The other clade included the fourth lineage (D) and a group of eight specimens from Vale da Couda that showed an unresolved phylogenetic position. These specimens grouped with the clade of *C. setubalensis* from Arrábida (Fig. 2).

Population genetics. MtDNA sequence data of 73 putative *C. coudensis* individuals 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 2A).

These haplotypes were organized into five main divergent haplogroups, with 22 to 63 mutation steps apart (Figure 3A). 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%). In spite of 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 3 and 4).

NuDNA 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 2A). These haplotypes constitute two haplogroups separated by 12 mutation steps (Figure 3B). 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.

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 and in site 4 only haplogroups B and D are represented. The two ITS-1 haplogroups are just present in two sampled sites, 1 and 2. We found no association between nuDNA and mtDNA haplogroups. All COI haplogroups, except haplogroup C, are represented within ITS-1 haplogroups with individuals from haplogroup R of ITS-1 belonging to COI haplogroups A, B, D and group of 8

non monophyletic individuals, while ITS-1 S haplogroup individuals belong to the most common COI haplogroup (D) (Figure 4).

The mismatch analysis for Vale da Couda haplogroup D shows signs of historical demographic expansion. The frequency of pairwise differences between haplotypes of lineage A showed a distribution consisting of a unimodal curve, which combined with non-significant values of SDD ($p = 0.13$) and raggedness ($p = 0.08$) do not reject the null hypothesis of expansion. Adding to this, a low and significant value of R2 ($p = 0.0035$) also supports the hypothesis of population growth for Vale da Couda haplogroup A. 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. We detected a clear signal of demographic expansion for lineage D (Fig. 5). For a population expansion of this clade to be compatible with LGM, the mutation rate would have to be between ca. 15%/MY, 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).

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 3). 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 3). Maxent models produced a strong discrimination between presence and background data regardless of

294 the threshold used (Table 4). Overall validation scores of models calibrated under current
295 conditions were: mean AUC = 0.981 ± 0.002, mean sensitivity = 0.979 ± 0.017, and mean
296 specificity = 0.982 ± 0.012 (Table 4). The resulting ensemble for the current distribution showed
297 two main areas with high probability of presence of Vale da Couda lineages: (1) one around the
298 presently known distribution, and (2) different patches at the north of the Iberian Peninsula
299 (Figure 6A). LGM projection indicates that past distribution of suitable habitats could have been
300 wider, with also appropriate conditions in the Andalusian region and in a smaller area in the
301 Central System (Figure 6B).
302
303 **Discussion**
304 The mitochondrial marker used in this study supports an unequivocal Vale da Couda complex
305 phylogeny, with four highly divergent and genetically diverse lineages. Our results show a
306 departure from the expectation that species with restricted distributions have low genetic
307 variability, adding to a list of exceptions that keeps growing (e.g. Coates et al. 2006; Ellis et al.
308 2006; Gevaert et al. 2013; Young et al. 1996). The maintenance of diversity in rare species can
309 be explained by the existence of a large effective population size (Ellstrand & Elam 1993), which
310 is likely the case with *Candidula* from Vale da Couda that are usually locally abundant. The
311 distribution models obtained for Vale da Couda lineages should be interpreted as regions with
312 environmental conditions analogous to those where the species presently occurs (Pearson et al.
313 2007) and represent a probability of occurrence of the species.

314 **Population genetics.** Studies have shown populations with high genetic structure existing during
315 the LGM in the Iberian Peninsula (Gómez & Lunt 2007). The particular geographical
316 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 2B). 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.

Species distribution modeling. 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. 6B). This predicted distribution implies a co-occurrence between Vale da Couda lineages and other species of the genus (e.g. *C. setubalensis* and *C. belemensis*) 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 3) 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.

363 Despite we have addressed some common hindcasting uncertainties by using different GCMs
364 and a clamping mask hindcast approach, we could not solve the lack of accurate lithological data
365 for emerged coastal land in the LGM. Thus, further appropriate habitats not depicted in our
366 models could have existed in regions near the coast.
367
368 **Biogeographic scenario.** The putative past and contemporary models of distribution of Vale da
369 Couda lineages are compatible with a biogeographic LGM scenario of species co-existence in
370 more southern locations followed by a northern dispersal tracking the species optimal thermal,
371 humidity and soil physical conditions. This co-existence is plausible given the fact that *C.*
372 *setubalensis* and *C. arrabidensis* occur in sympatry, as well as *C. olisippensis* and *C. coudensis*
373 (Holyoak & Holyoak 2014) and share habitat requirements with Vale da Couda lineages (Figure
374 2). Moreover, Roucoux (2001) shows low but fluctuating tree pollen through the LGM, along
375 with abundant grass and some herb pollen, indicating likely widespread suitability of the grassy
376 habitats for *Candidula* species throughout the LGM. Similar events have already been detected
377 for other land snail species (Harl et al. 2014; Sauer & Hausdorf 2010; Shimizu & Ueshima
378 2000). After the LGM, environmental conditions during deglaciation were such that promoted
379 northward dispersal of land snails and the establishment of populations in locations of suitable
380 isothermality and precipitation like Vale da Couda. *C. setubalensis* and *C. arrabidensis*
381 maintained a southern distribution, in the Setubal Peninsula. Specifically, we hypothesize that
382 Pleistocene conditions may have isolated populations into pockets of suitable habitats in more
383 southern locations, which promoted population differentiation and intra-specific diversification
384 without apparent geological barriers.
385

386 **Taxonomic implications.** Divergence between COI sequences has been used as a tool to
387 identify new species based on the premise that there is a "DNA barcoding gap", meaning that
388 the genetic differences between intra- and inter-specific individuals do not overlap, with the
389 latter being bigger than the former (Bucklin et al. 2011). Several studies have shown that high
390 levels of mtDNA divergence are extremely common within land snail species, reaching values
391 as high as 30% (Chiba 1999; Davison et al. 2009; Haase et al. 2003; Thomaz et al. 1996), and
392 values between species going as low as 1% (Davison et al. 2009). These findings preclude the
393 application of the DNA barcoding approach to land snails because there is no known "barcoding
394 gap" and as such species cannot be discriminated solely based on a threshold value of sequence
395 divergence. Nevertheless, a conservative approach to our results imply the existence in Vale da
396 Couda of at least two species, represented by the two most divergent lineages (Figure 2). These
397 lineages are likely to display the highest genetic variability of the two species that according to
398 Holyoak (2014) coexist in that region, *C. coudensis* and *C. olisippensis*. As discussed before,
399 patterns of high mitochondrial divergence can be attributed to the maintenance of ancient
400 polymorphisms from past isolations (Dillon Jr & Robinson 2009) combined with the effects of
401 present large overall population size and gene flow between fragmented sub-populations.
402
403 **Conclusions.** The genetic survey presented here revealed the existence of four lineages in Vale
404 da Couda with independent evolutionary histories. These results do not corroborate previous
405 morphological studies that considered the existence of a single species, *Candidula coudensis*.
406 LGM hindcasts revealed the existence of putative glacial refugia south of the current
407 distribution of the lineages of Vale da Couda. We hypothesize a biogeographic scenario, which
408 is consistent with the inferred LGM distribution and the high genetic diversity observed in a

409 terrestrial snail with an extremely narrow known distributional range. Although unexpected in
410 organisms with such a restricted distribution, the high genetic diversity found in Vale da Couda
411 lineages adds to the growing list of exceptions where low-dispersal species show high levels of
412 genetic variability. These findings have implications for the understanding of the genetic
413 characteristics of rare and endemic species. From a conservation perspective, Vale da Couda
414 lineages do not seem to be endangered, with high genetic diversity within and between lineages
415 maintained by putative large effective population sizes.
416
417 **Acknowledgements**. We wish to thank Manuela Coelho (FCUL) for relevant comments on
418 earlier versions of this manuscript and David and Geraldine Holyoack for critical remarks.
419
420
421 **References**
422
423 Akaike H. 1974. A new look at the statistical model identifications. *IEEE Transactions on*
424 *automatic control* 19:716–723.
425 Austin MP, Belbin L, Meyers JA, Doherty MD, and Luoto M. 2006. Evaluation of statistical
426 models used for predicting plant species distributions: role of artificial data and theory.
427 *Ecological Modelling* 199:197-216. 10.1016/j.ecolmodel.2006.05.023
428 Avise JC. 2000. *Phylogeography: the history and formation of species.* Cambridge,
429 Massachussets: Harvard University Press.
430 Bandelt H-J, Forster P, and Röhl A. 1999. Median-joining networks for inferring intraspecific
431 phylogenies. *Mol Biol Evol* 16:37–48.
432 Bird CE, Karl SA, Smouse PE, and Toonen RJ. 2011. Detecting and measuring genetic
433 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.

Bond JE, Hedin MC, Ramirez MG, and Opell BD. 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. *Molecular ecology* 10:899-910. 10.1046/j.1365-294X.2001.01233.x

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

456 Chiba S. 1999. Accelerated evolution of land snails *Mandarina* in the oceanic Bonin Islands:
457 evidence from mitochondrial DNA sequences. *Evolution*:460-471. 10.2307/2640782
458 Coates DJ, Tischler G, and McComb JA. 2006. Genetic variation and the mating system in the
459 rare *Acacia sciophanes* compared with its common sister species *Acacia anfractuosa*
460 (*Mimosaceae*). *Conservation Genetics* 7:931-944. 10.1007/s10592-006-9136-7
461 Davison A, Blackie RL, and Scothern GP. 2009. DNA barcoding of stylommatophoran land
462 snails: a test of existing sequences. *Molecular ecology resources* 9:1092-1101.
463 10.1111/j.1755-0998.2009.02559.x
464 Dillon Jr RT, and Robinson JD. 2009. The snails the dinosaurs saw: are the pleurocerid
465 populations of the Older Appalachians a relict of the Paleozoic Era? *Journal of the North*
466 *American Benthological Society* 28:1-11. 10.1899/08-034.1
467 Doyle J, and Doyle JL. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method.
468 *Phytochem Bull* 19:11-15.
469 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
470 throughput. *Nucleic Acids Research* 32:1792-1797. 10.1093/nar/gkh340
471 Elith J, and Leathwick JR. 2009. Species distribution models: ecological explanation and
472 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*
473 40:677-697. 10.1146/annurev.ecolsys.110308.120159
474 Ellis J, Pashley C, and McCauley D. 2006. High genetic diversity in a rare and endangered
475 sunflower as compared to a common congener. *Molecular ecology* 15:2345-2355.
476 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

499 Haase M, Misof B, Wirth T, Baminger H, and Baur B. 2003. Mitochondrial differentiation in a
500 polymorphic land snail: evidence for Pleistocene survival within the boundaries of
501 permafrost. *Journal of Evolutionary Biology* 16:415-428. 10.1046/j.1420-
502 9101.2003.00542.x
503 Harl J, Páll-Gergely B, Kirchner S, Sattmann H, Duda M, Kruckenhauser L, and Haring E. 2014.
504 Phylogeography of the land snail genus *Orcula (Orculidae, Stylommatophora)* with
505 emphasis on the Eastern Alpine taxa: speciation, hybridization and morphological
506 variation. *BMC evolutionary biology* 14:1-26. 10.1186/s12862-014-0223-y
507 Harpending H. 1994. Signature of ancient population growth in a low-resolution mitochondrial
508 DNA mismatch distribution. *Human Biology* 66:591–600.
509 Hartmann J, and Moosdorf N. 2012. The new global lithological map database GLiM: A
510 representation of rock properties at the Earth surface. *Geochemistry, Geophysics,*
511 *Geosystems* 13. 10.1029/2012GC004370
512 Hayashi M, and Chiba S. 2000. Intraspecific diversity of mitochondrial DNA in the land snail
513 *Euhadra peliomphala* (Bradybaenidae). *Biological Journal of the Linnean Society*
514 70:391-401. 10.1111/j.1095-8312.2000.tb01230.x
515 Hermida J, Ondina M, and Rodriguez T. 2000. The relative importance of edaphic factors on the
516 distribution of some terrestrial gastropod species: autecological and synecological
517 approaches. *Acta Zoologica Academiae Scientiarum Hungaricae* 46:265-274.
518 Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the*
519 *Linnean Society* 68:87-112. 10.1111/j.1095-8312.1999.tb01160.x

520 Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary.
521 *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*
522 359:183–195. 10.1098/rstb.2003.1388
523 Hijmans RJ, Cameron SE, Parra JL, Jones PG, and Jarvis A. 2005. Very high resolution
524 interpolated climate surfaces for global land areas. *International journal of climatology*
525 25:1965-1978. 10.1002/joc.1276
526 Hillis DM, and Dixon MT. 1991. Ribosomal DNA: molecular evolution and phylogenetic
527 inference. *Quarterly Review of Biology* 66:411-453.
528 Hirzel AH, Hausser J, Chessel D, and Perrin N. 2002. Ecological-niche factor analysis: how to
529 compute habitat-suitability maps without absence data? *Ecology* 83:2027-2036.
530 10.2307/3071784
531 Holyoak DT, and Holyoak GA. 2014. A review of the genus *Candidula* in Portugal with notes on
532 other populations in Western Europe (*Gastropoda, Pulmonata, Hygromiidae*) *Journal of*
533 *Conchology* 41:629-672.
534 Jost L. 2008. GST and its relatives do not measure differentiation. *Molecular ecology* 17:4015-
535 4026. 10.1111/j.1365-294X.2008.03887.x
536 Kadmon R, and Heller J. 1998. Modelling faunal responses to climatic gradients with GIS: land
537 snails as a case study. *Journal of Biogeography* 25:527-539. 10.1046/j.1365-
538 2699.1998.2530527.x
539 Katoh K, and Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
540 Improvements in Performance and Usability. *Mol Biol Evol* 30:772-780.
541 10.1093/molbev/mst010

542 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
543 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, and Drummond A. 2012.
544 Geneious Basic: an integrated and extendable desktop software platform for the
545 organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
546 10.1093/bioinformatics/bts199
547 Keenan K, McGinnity P, Cross TF, Crozier WW, and Prodöhl PA. 2013. diveRsity: An R
548 package for the estimation and exploration of population genetics parameters and their
549 associated errors. *Methods in Ecology and Evolution* 4:782–788. 10.1111/2041-
550 210X.12067
551 Köhler F, and Burghardt I. 2015. Cryptic diversity in a widespread land snail: revision of the
552 genus *Xanthomelon* Martens, 1860 from the Australian Monsoon Tropics (Pulmonata,
553 Camaenidae). *Zoologica Scripta* 45:127-144. 10.1111/zsc.12144
554 Laikre L, Nilsson T, Primmer CR, Ryman N, and Allendorf FW. 2009. Importance of genetics in
555 the interpretation of favourable conservation status. *Conservation Biology* 23:1378-1381.
556 10.1111/j.1523-1739.2009.01360.x
557 Librado P, and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA
558 polymorphism data. *Bioinformatics* 25:1451-1452. 10.1093/bioinformatics/btp187
559 Lobo JM, Jiménez-Valverde A, and Hortal J. 2010. The uncertain nature of absences and their
560 importance in species distribution modelling. *Ecography* 33:103-114. 10.1111/j.1600-
561 0587.2009.06039.x
562 Meirmans PG, and Hedrick PW. 2011. Assessing population structure: F(ST) and related
563 measures. *Molecular ecology resources* 11:5–18. Doi 10.1111/J.1755-
564 0998.2010.02927.X

565 Moreira F, Calado G, and Dias S. 2015. Conservation status of a recently described endemic land
566 snail, *Candidula coudensis*, from the Iberian Peninsula. *PLoS ONE* 10:e0138464.
567 10.1371/journal.pone.0138464
568 Nei M. 1987. Genetic distance and molecular phylogeny. In: Ryman N, and Utter FW, eds.
569 *Population Genetics & Fishery Management*. Seattle: Washington Sea Grant Program,
570 University of Washington, 193–223.
571 Nei M, and Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases.
572 *Genetics* 97:145–163.
573 Palacios D, de Andrés N, López-Moreno JI, and García-Ruiz JM. 2015. Late Pleistocene
574 deglaciation in the upper Gállego Valley, central Pyrenees. *Quaternary Research* 83:397-
575 414. 10.1016/j.yqres.2015.01.010
576 Pearson RG, Raxworthy CJ, Nakamura M, and Townsend Peterson A. 2007. Predicting species
577 distributions from small numbers of occurrence records: a test case using cryptic geckos
578 in Madagascar. *Journal of Biogeography* 34:102-117. 10.1111/j.1365-2699.2006.01594.x
579 Pfenninger M, Nowak C, and Magnin F. 2007. Intraspecific range dynamics and niche evolution
580 in *Candidula* land snail species. *Biological Journal of the Linnean Society* 90:303-317.
581 10.1111/j.1095-8312.2007.00724.x
582 Pfenninger M, Posada D, and Magnin F. 2003. Evidence for survival of Pleistocene climatic
583 changes in northern refugia by the land snail *Trochoidea geyeri* (Soós 1926)
584 (*Helicellinae, Stylommatophora*). *BMC evolutionary biology* 3:1-13. 10.1186/1471-2148-
585 3-8

586 Phillips S, Anderson R, and Schapire R. 2006. Maximum entropy modeling of species
587 geographic distributions. *Ecological Modelling* 190:231–259.
588 10.1016/j.ecolmodel.2005.03.026
589 Pinceel J, Jordaens K, Pfenninger M, and Backeljau T. 2005. Rangewide phylogeography of a
590 terrestrial slug in Europe: evidence for Alpine refugia and rapid colonization after the
591 Pleistocene glaciations. *Molecular ecology* 14:1133–1150. 10.1111/j.1365-
592 294X.2005.02479.x
593 Primack RB. 2006. *Essentials of conservation biology*. Sunderland, Massachusetts: Sinauer
594 Associates.
595 R Development Core Team. 2013. R: A language and environment for statistical computing.
596 Ramos-Onsins SE, and Rozas J. 2002. Statistical properties of new neutrality tests against
597 population growth. *Mol Biol Evol* 19:2092–2100.
598 Rogers AR, and Harpending H. 1992. Population growth makes waves in the distribution of
599 pairwise genetic differences. *Mol Biol Evol* 9:552–569.
600 Roucoux KH, Shackleton NJ, de Abreu L, Schönfeld J, and Tzedakis PC. 2001. Combined
601 marine proxy and pollen analyses reveal rapid Iberian vegetation response to North
602 Atlantic millennial-scale climate oscillations. *Quaternary Research* 56:128-132.
603 10.1006/qres.2001.2218
604 Ryman N, and Leimar O. 2009. GST is still a useful measure of genetic differentation - a
605 comment on Jost's D. *Molecular ecology* 18:2084-2087.
606 Sambrook JF, and Russell DW. 2001. *Molecular cloning: a laboratory manual*: Cold spring
607 harbor laboratory press, New York.

608 Sauer J, and Hausdorf B. 2010. Reconstructing the evolutionary history of the radiation of the
609 land snail genus *Xerocrassa* on Crete based on mitochondrial sequences and AFLP
610 markers. *BMC evolutionary biology* 10:1-13. 10.1186/1471-2148-10-299
611 Schmatz D, Luterbacher J, Zimmermann N, and Pearman P. 2015. Gridded climate data from 5
612 GCMs of the Last Glacial Maximum downscaled to 30 arc s for Europe. *Climate of the*
613 *Past Discussions* 11:2585-2613.
614 Schneider S, and Excoffier L. 1999. Estimation of past demographic parameters from the
615 distribution of pairwise differences when the mutation rates vary among sites: application
616 to human mitochondrial DNA. *Genetics* 152:1079-1089.
617 Shimizu Y, and Ueshima R. 2000. Historical biogeography and interspecific mtDNA
618 introgression in *Euhadra peliomphala* (the Japanese land snail). *Heredity* 85:84-96.
619 10.1046/j.1365-2540.2000.00730.x
620 Tamura K, and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control
621 region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512-526.
622 Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: molecular
623 evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725-2729.
624 10.1093/molbev/mst197
625 Tattersfield P, Warui C, Seddon M, and Kiringe J. 2001. Land-snail faunas of afromontane
626 forests of Mount Kenya, Kenya: ecology, diversity and distribution patterns. *Journal of*
627 *Biogeography* 28:843-861. 10.1046/j.1365-2699.2001.00606.x
628 Thomaz D, Guiller A, and Clarke B. 1996. Extreme divergence of mitochondrial DNA within
629 species of pulmonate land snails. *Proceedings of the Royal Society of London B:*
630 *Biological Sciences* 263:363-368. 10.1098/rspb.1996.0056

631 Tsoar A, Allouche O, Steinitz O, Rotem D, and Kadmon R. 2007. A comparative evaluation of
632 presence-only methods for modelling species distribution. *Diversity and Distributions*
633 13:397-405. 10.1111/j.1472-4642.2007.00346.x
634 Weir BS, and Cockerham CC. 1984. Estimating F-statistics for the analysis of population
635 structure. *Evolution* 38:1358–1370. 10.2307/2408641
636 White TJ, Bruns T, Lee S, and Taylor J. 1990. Amplification and direct sequencing of fungal
637 ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and*
638 *applications* 18:315-322.
639 Whitlock MC. 2011. G'st and D do not replace FST. *Molecular ecology* 20:1083-1091.
640 Young A, Boyle T, and Brown T. 1996. The population genetic consequences of habitat
641 fragmentation for plants. *Trends in ecology & evolution* 11:413-418. 10.1016/0169-
642 5347(96)10045-8
643

Figure 1(on next page)

Distribution of *Candidula* sampling sites.

Figure 1. Distribution of *Candidula* sampling sites in mainland Portugal and number of samples. Inset: detail of Vale da Couda collection sites.

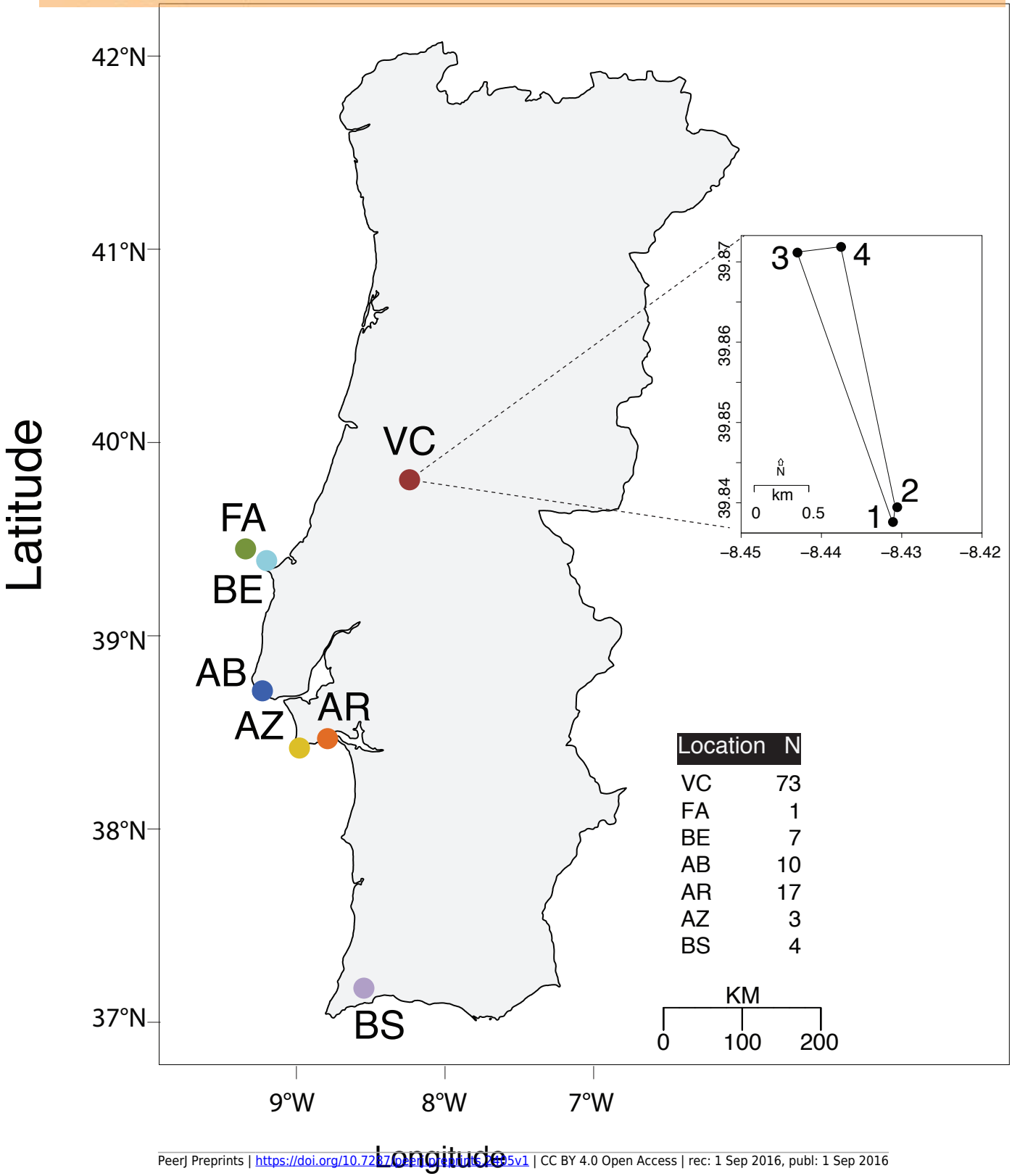
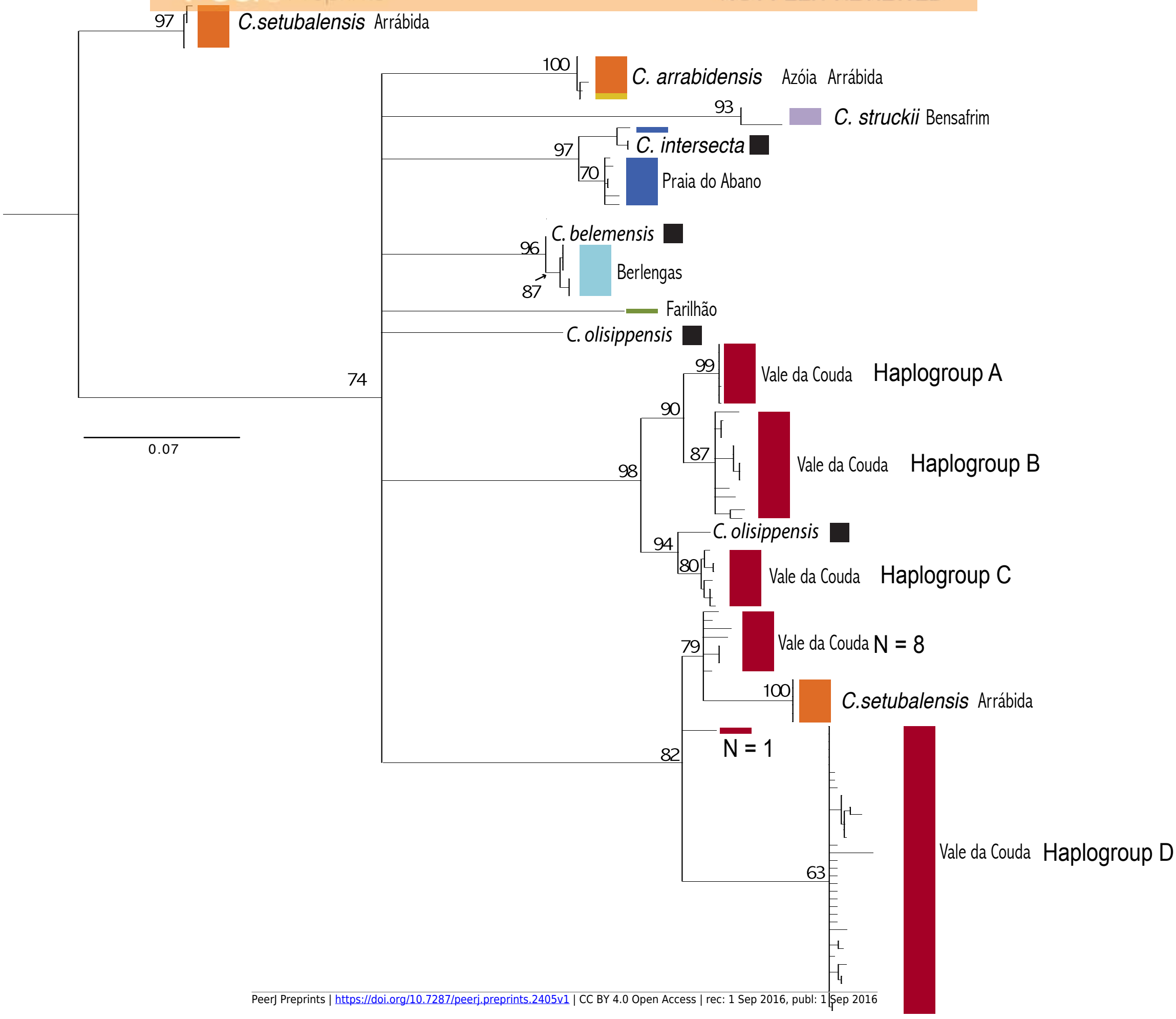


Figure 2(on next page)

Phylogenetic relationships between individuals from Vale da Couda (in red) and other locations.

Figure 2. Phylogenetic relationships between *Candidula* individuals from Vale da Couda (in red) and other locations. Black squares represent individuals identified morphologically and anatomically. Colours of locations as in Figure 1. Outgroups removed from figure for illustrative purposes.



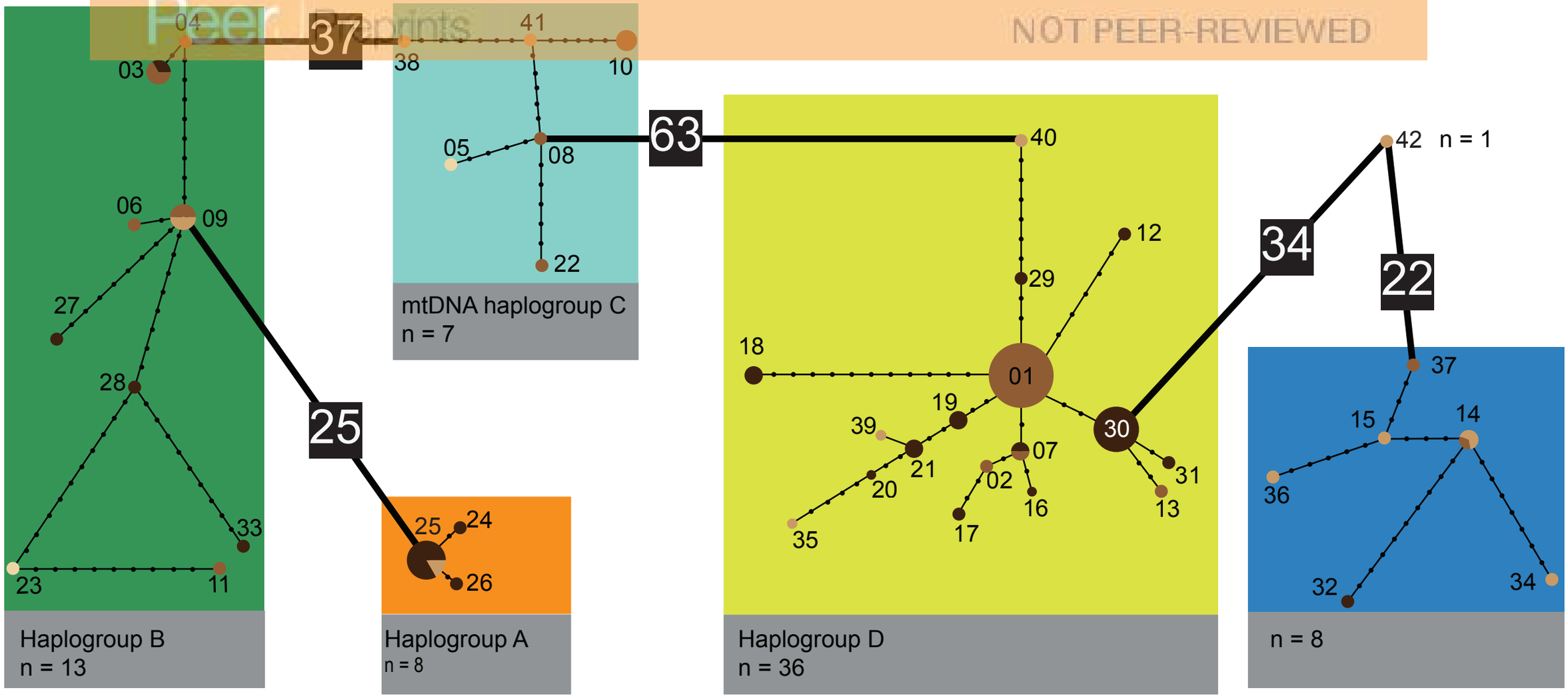
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2405v1 | CC BY 4.0 Open Access | rec: 1 Sep 2016, publ: 1 Sep 2016

Figure 3(on next page)

Statistical parsimony haplotype networks for Vale da Couda individuals.

Figure 3. (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.

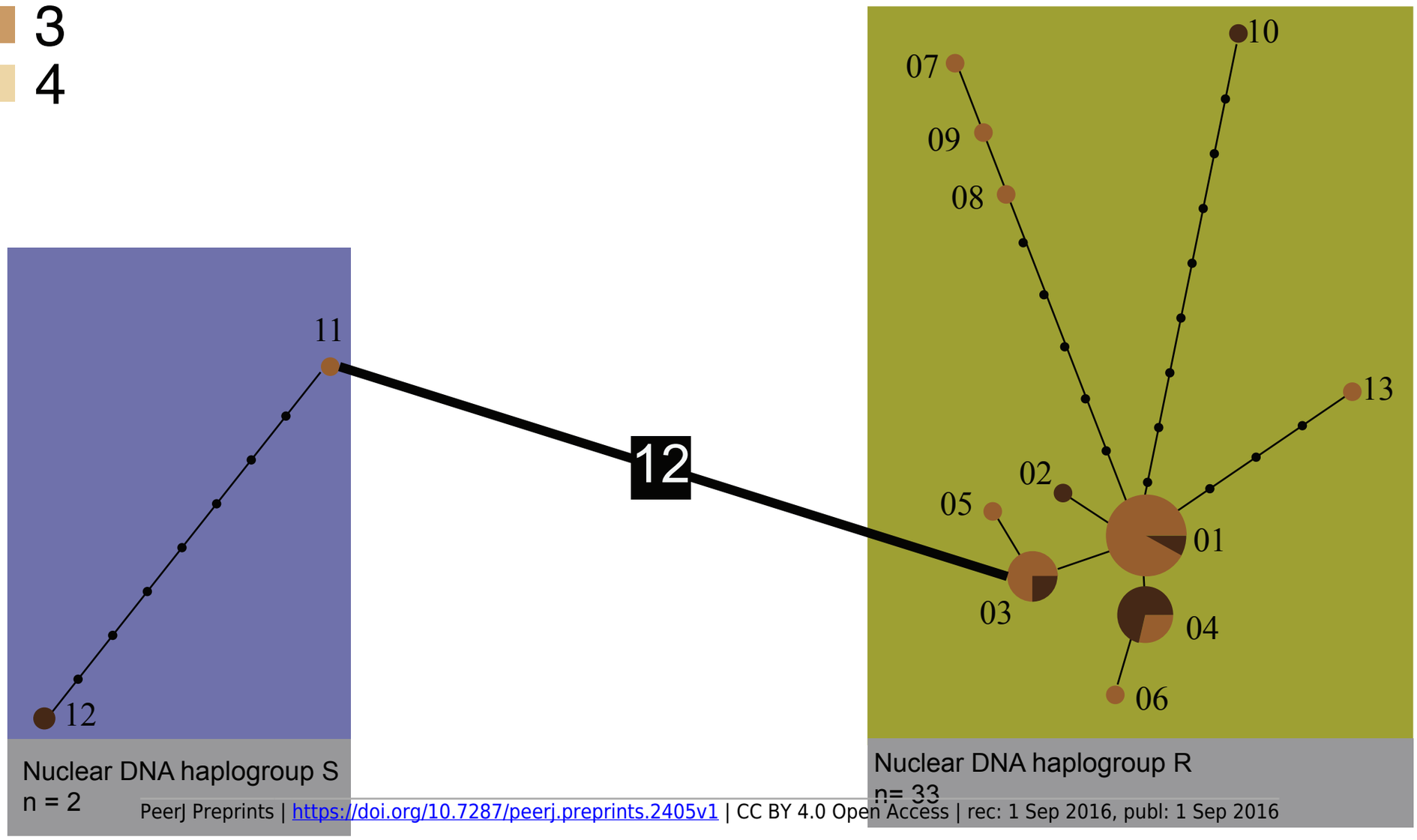
A



Locations

1
2
3
4

B

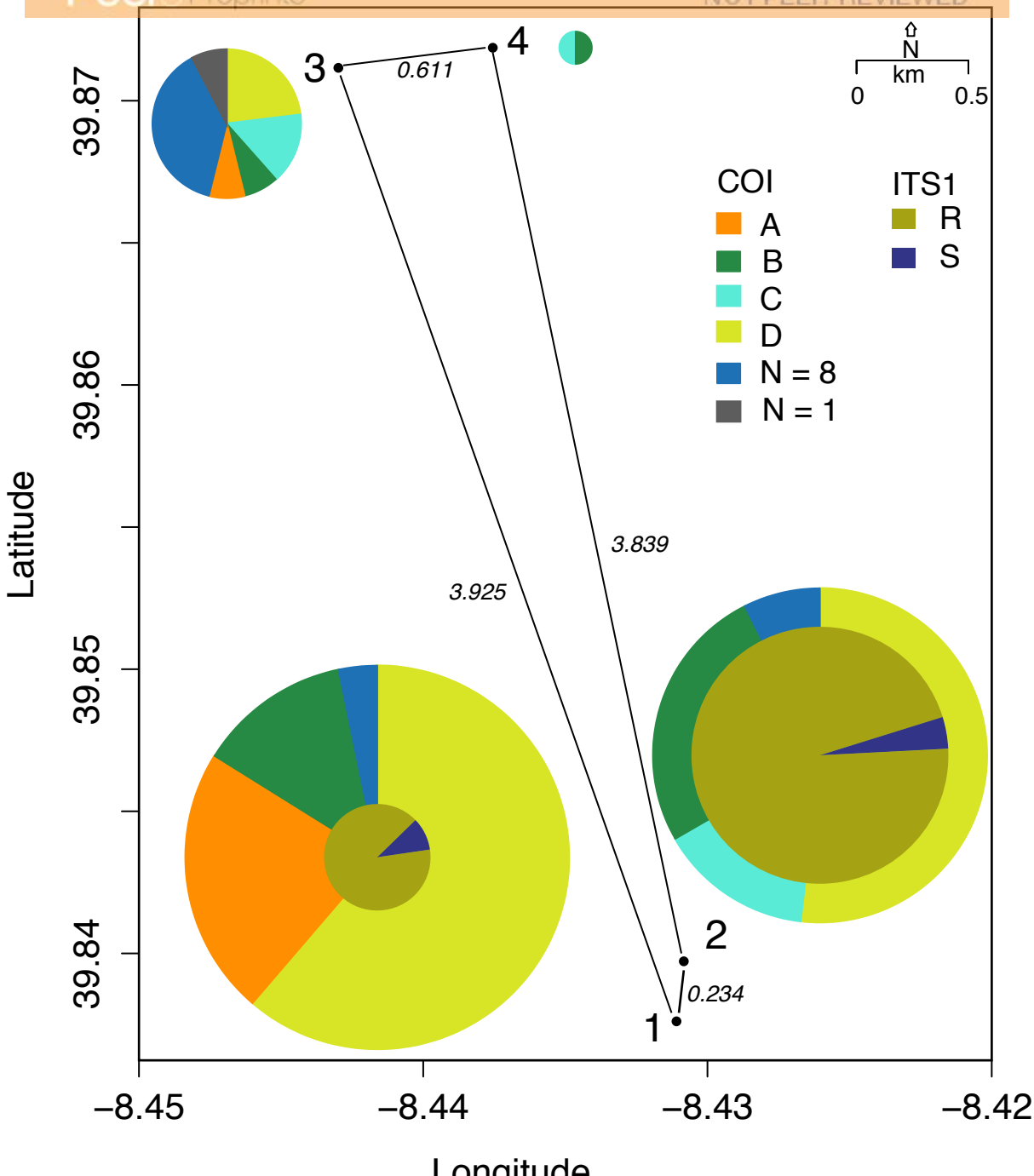


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2405v1 | CC BY 4.0 Open Access | rec: 1 Sep 2016, publ: 1 Sep 2016

Figure 4(on next page)

Distribution of mtDNA lineages in Vale da Couda sites.

Figure 4. 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 3.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2465v1 | CC BY 4.0 Open Access | rec: 1 Sep 2016, publ: 1 Sep 2016

Figure 5(on next page)

Mismatch distribution of lineage D from Vale da Couda

Figure 5. 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 (θ0) and after (θ1) 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 R2.

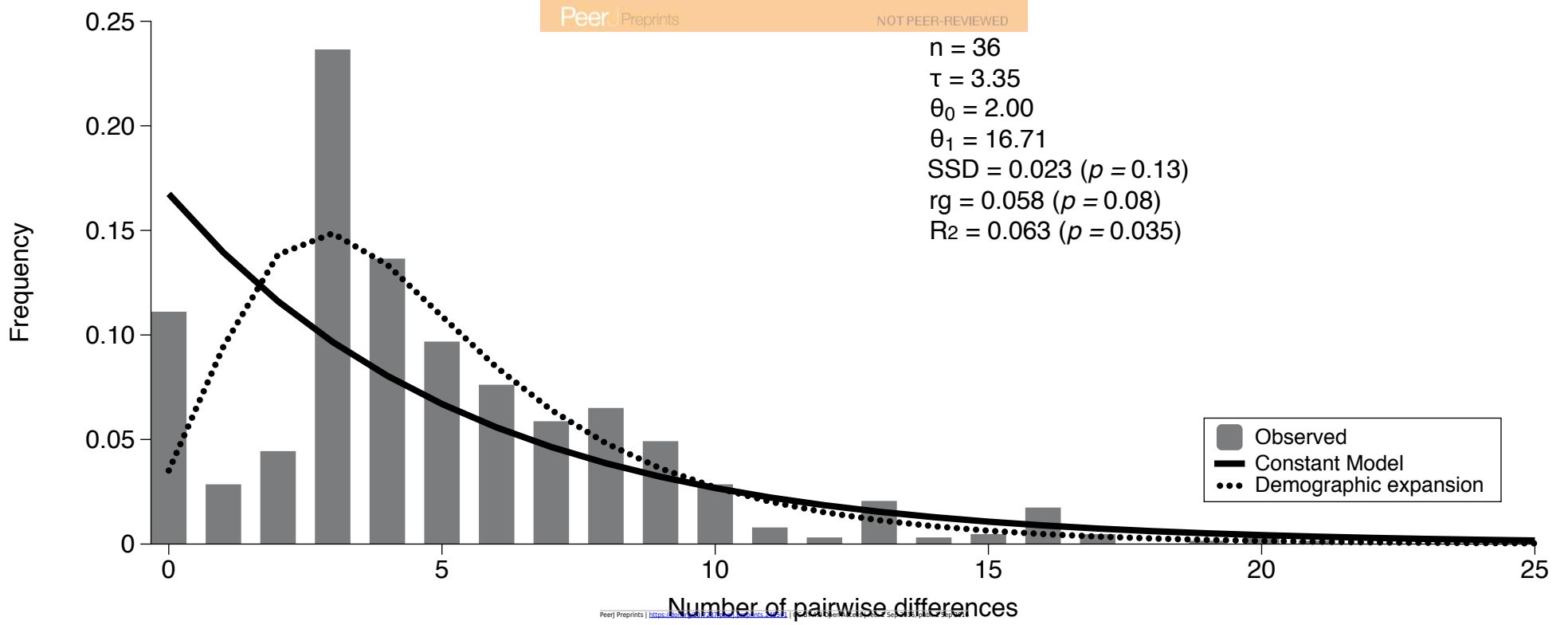
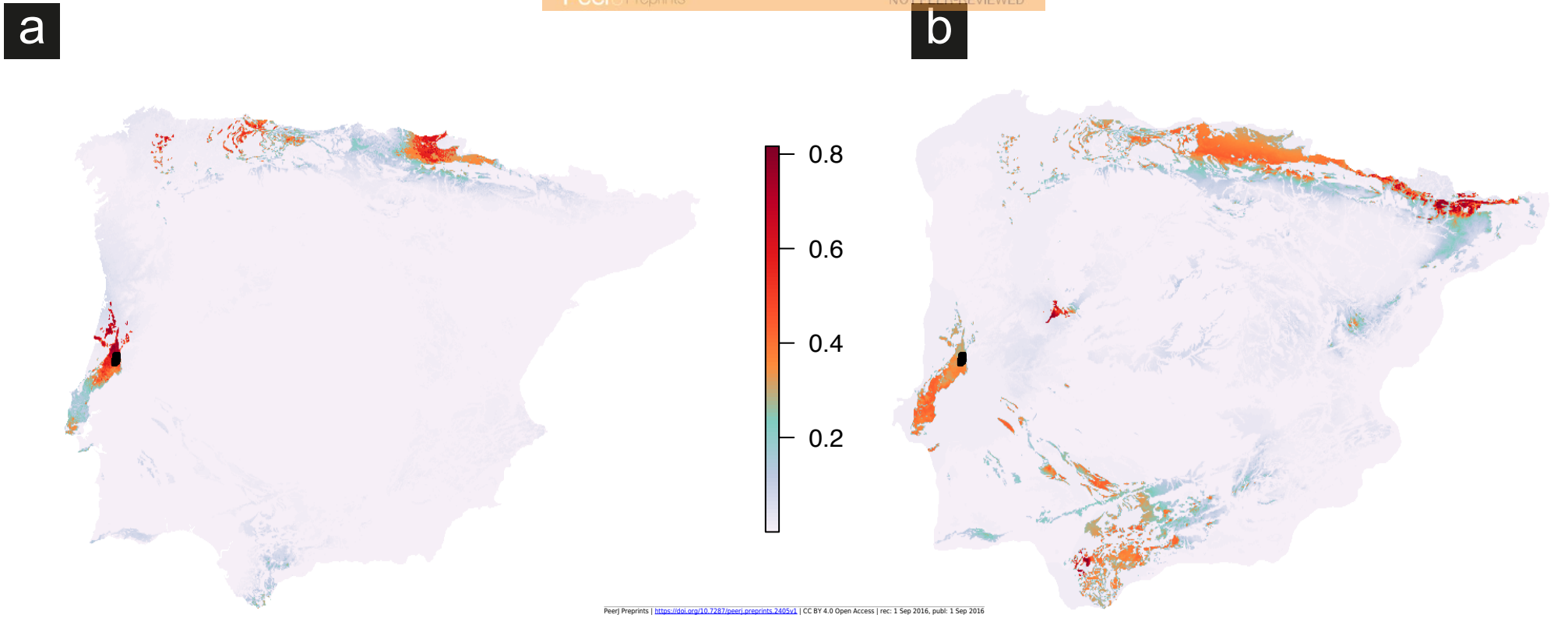


Figure 6(on next page)

LGM and current climate predicted geographic distributions.

Figure 6. 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.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2405v1 | CC BY 4.0 Open Access | rec: 1 Sep 2016, publ: 1 Sep 2016

Table 1(on next page)

Sample location and summary statistics for *Candidula*.

Sample location and summary statistics for *Candidula*.

1

Location	**Code**	**Long**	**Lat**	***N***	***nh***	***h* ± *s.d***	**π ± s.d.**
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

2

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

5

Table 2(on next page)

Vale da Couda lineages and sites statistics.

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)

	COI					ITS-1				
Lineages	***N***	***Nh***	**Locations**	**h ± s.d.**	**? ± s.d.**	***N***	***Nh***	**Locations**	**h ± s.d.**	**? ± s.d.**
A	36	17	1, 2, 3	0.889 ± 0.001	0.009 ± 0.001	33	11	1,2	0.799 0.054	0.006 ± 0.001
B	7	6	2, 3, 4	0.952 ± 0.096	0.010 ± 0.096	2	2	1,2	1.000 0.500	0.016 ± 0.008
C	8	3	1, 3	0.464 ± 0.040	0.001 ± 0.000					
D	13	9	1, 2, 3, 4	0.936 ± 0.051	0.018 ± 0.002					
-	8	6	1, 2, 3	0.893 ± 0.111	0.013 ± 0.003					

(B)

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

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

Table 3(on next page)

Environmental Niche Factor Analysis

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.

1

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

2
3

Table 4(on next page)

Summary of Maxent models

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).

1

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

2