Biological Journal of the Linnean Society, 2020, **129**, 259–272. With 3 figures.

Population genomics of *Bombus terrestris* reveals high but unstructured genetic diversity in a potential glacial refugium

SARA E. SILVA[1,*], SOFIA G. SEABRA[1], LUÍSA G. CARVALHEIRO[1,2], VERA L. NUNES[1], EDUARDO MARABUTO[1], RAQUEL MENDES[1], ANA S. B. RODRIGUES[1], FRANCISCO PINA-MARTINS[1], SELÇUK YURTSEVER[3], TELMA G. LAURENTINO[4], ELISABETE FIGUEIREDO[5], MARIA T. REBELO[6] and OCTÁVIO S. PAULO[1]

[1]*Centre for Ecology, Evolution and Environmental Changes (cE3c), Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal*
[2]*Department of Ecology, Universidade Federal de Goiás, Goiânia, Brazil*
[3]*Biology Department, Science Faculty, Trakya University, Edirne, Turkey*
[4]*Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland*
[5]*Linking Landscape, Environment, Agriculture and Food (LEAF), Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, Lisboa, Portugal*
[6]*Centre for Environmental and Marine Research (CESAM), Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal*

Received 17 August 2019; revised 13 November 2019; accepted for publication 14 November 2019

Ongoing climate change is expected to cause an increase in temperature and a reduction of precipitation levels in the Mediterranean region, which might cause changes in many species distributions. These effects negatively influence species gene pools, decreasing genetic variability and adaptive potential. Here, we use mitochondrial DNA and RADseq to analyse population genetic structure and genetic diversity of the bumblebee species *Bombus terrestris* (subspecies *Bombus terrestris lusitanicus*), in the Iberian Peninsula. Although this subspecies shows a panmictic pattern of population structure across Iberia and beyond, we found differentiation between subspecies *B. t. lusitanicus* and *B. t. africanus*, probably caused by the existence of barriers to gene flow between Iberia and North Africa. Furthermore, the results revealed that the Iberian Peninsula harbours a large fraction of *B. terrestris* intraspecific genetic variation, with the highest number of mitochondrial haplotypes found when compared with any other region in Europe studied so far, suggesting a potential role for the Iberian Peninsula as a glacial refugium. Our findings strengthen the idea that Iberia is a very important source of diversity for the global genetic pool of this species, because rare alleles might play a role in population resilience against human- or climate-mediated changes.

ADDITIONAL KEYWORDS: buff-tailed bumblebee – Iberian Peninsula – mitochondrial DNA – phylogeography – RADseq.

INTRODUCTION

As a consequence of several human-driven environmental changes (e.g. habitat loss, agricultural intensification, use of pesticides, the introduction of new parasites and climate change), insect pollinators have purportedly declined dramatically in recent decades (reviewed by Potts *et al.*, 2010). Moreover, the massive use of managed and commercial bees for crop pollination and honey production (e.g. *Apis mellifera*, *Bombus terrestris*) has facilitated the introduction and spread of diseases and parasites (Goulson, 2010) and interfered with the genetic composition of natural populations through hybridization (Jaffé *et al.*, 2016; Seabra *et al.*, 2019).

Bumblebees (*Bombus* spp.) have been particularly affected worldwide by the above-mentioned problems. Several species have endured dramatic changes in their distribution or become locally extinct in developed regions, such as central and western Europe and North

*Corresponding author. E-mail: saraema.silva@gmail.com

America (Goulson *et al.*, 2008; Bommarco *et al.*, 2012). Although several studies have reported high levels of gene flow among European populations (Woodard *et al.*, 2015; Lecocq *et al.*, 2017), some species have experienced a decline in genetic diversity at a local scale (Woodard *et al.*, 2015), which might increase the risk of inbreeding and hinder the ability of populations to cope with environmental change (Goulson *et al.*, 2008; Maebe *et al.*, 2015).

Populations that have persisted in glacial refugia are expected to show higher levels of intraspecific genetic variation compared with populations outside these regions, owing to long-term population persistence and isolation (Hewitt, 1999). The Iberian Peninsula was one of the largest Mediterranean refugia during the Quaternary (2.6 Mya to the present) glaciations and is also at the southernmost latitude limit (rear edge) for many continental species ranges (Hewitt, 2000; Arias *et al.*, 2006). Therefore, many species in Iberia present geographically structured genetic lineages (Miraldo *et al.*, 2011; Rodrigues *et al.*, 2014; Chávez-Galarza *et al.*, 2015). Additionally, the proximity of Iberia to Africa, especially at the Strait of Gibraltar, and with episodic bridges between the two continents, enabled occasional dispersal of the more vagile organisms, particularly during periods of glacial southern contraction or of postglacial northern expansion (Pinto *et al.*, 2013). However, according to Rasmont *et al.* (2015), the Iberian Peninsula is expected to experience major reductions of bumblebee-suitable climatic conditions within the forthcoming decades, alongside other southern European regions. Intensive land-use regimens and degradation of semi-natural areas might already be affecting species richness patterns for many taxa across the peninsula (Martins *et al.*, 2014; Newbold *et al.*, 2015). Moreover, the use of commercial bumblebees for crop pollination, which is a common practice in several parts of the peninsula, might have a negative impact on native natural populations through pathogen spillover and introgression of maladaptive alleles (Murray *et al.*, 2013; Seabra *et al.*, 2019).

The buff-tailed bumblebee *Bombus terrestris* (Linnaeus, 1758) (Hymenoptera: Apidae) presents a wide distribution in the West Palaearctic region (Rasmont *et al.*, 2008) and has been introduced deliberately as a crop pollinator into several areas worldwide (Ings *et al.*, 2005; Goulson, 2010). Nine subspecies were described based on morphology, particularly in coat colour variation, with additional differences in behaviour, phenology, physiological traits and resistance to parasites (Fig. 1; Rasmont *et al.*, 2008). Studies of mitochondrial and microsatellite variation have shown a clear differentiation of northern African and islander *B. terrestris* populations from European

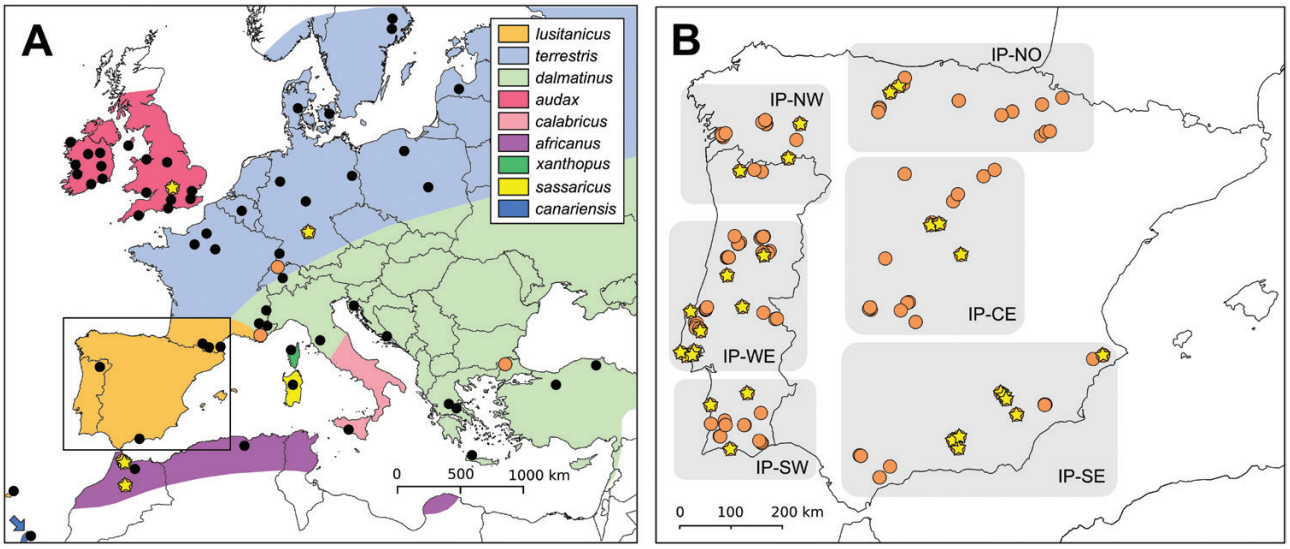

Figure 1. Sampling locations of samples used in this study. A, geographical distribution of each *Bombus terrestris* subspecies, according to Lecocq *et al.* (2016), here represented by different colours; geographical location of collected samples of *B. t. terrestris*, *B. t. dalmatinus*, *B. t. audax* and *B. t. africanus*, and of *COI* sequences [from previous studies by Coppée (2010), Williams *et al.* (2012a, b), Moreira *et al.* (2015) and Schmidt *et al.* (2015)] downloaded from GenBank. B, sampling locations of collected samples of *B. t. lusitanicus* across the Iberian Peninsula. Abbreviations: IP-CE, Iberian Peninsula, centre; IP-NO, Iberian Peninsula, north; IP-NW, Iberian Peninsula, north-west; IP-SE, Iberian Peninsula, south-east; IP-SW, Iberian Peninsula, south-west; IP-WE, Iberian Peninsula, west. B, magnified version of the black rectangle in A. Sampling locations are represented as follows: orange dots, samples used for *COI* analyses; yellow stars, samples used for both *COI* and RAD analyses; black dots, *COI* sequences downloaded from GenBank.

mainland ones, with no differentiation amongst the latter. In fact, mainland populations are largely homogeneous, with nearly panmixia patterns (Estoup *et al.*, 1996; Widmer *et al.*, 1998; Lecocq *et al.*, 2013b, 2016; Woodard *et al.*, 2015). Genetic cohesiveness found across broad geographical scales has been attributed to the absence of effective barriers to gene flow and the high dispersal capability of bumblebees (Estoup *et al.*, 1996; Woodard *et al.*, 2015; Lecocq *et al.*, 2016). The flight radius of *B. terrestris* males, for example, varies between 2.6 and 9.9 km (Kraus *et al.*, 2009). Queen dispersal in *B. terrestris* is unknown, but estimates for *Bombus pascuorum* (Scopoli, 1763) and *Bombus lapidarius* (Linnaeus, 1758) queens show that they are able to disperse by ≥ 3 and 5 km, respectively, during their lifetime (Lepais *et al.*, 2010). In the case of *B. terrestris*, two hypotheses have been suggested to explain the genetic homogeneity in the European mainland: (1) a recent population expansion from a single glacial refuge, although low sampling efforts hinder definite conclusions (Lecocq *et al.*, 2016); and (2) the erosion of past population structure owing to genetic homogenization linked to translocations of commercial *B. terrestris*, although this is unlikely because the same pattern of genetic homogenization was found in the early 1990s, when bumblebee commercialization was barely a practice (Estoup *et al.*, 1996). Unfortunately, studies on *B. terrestris* largely undersampled the Iberian Peninsula (Estoup *et al.*, 1996; Widmer *et al.*, 1998; Moreira *et al.*, 2015; Lecocq *et al.*, 2016), despite the importance of this area as a potential glacial refuge for the species and its present location at the south-western edge of the mainland distribution of *B. terrestris*.

To address this gap, we explore the role of the Iberian Peninsula in the differentiation dynamics of *B. terrestris*, by assessing the following factors: (1) whether Iberian *Bombus terrestris lusitanicus* (Krüger, 1956) is panmictic with the rest of the *B. terrestris* distribution; and (2) whether Iberian populations of *B. terrestris* harbour standing genetic variation in order to adapt to the currently changing environment. We address these questions by determining: (1) the level of genetic differentiation between *B. t. lusitanicus* and other recognized subspecies of *B. terrestris*; (2) the population genetic structure within the Iberian Peninsula; and (3) the levels of genetic diversity within *B. t. lusitanicus*. We contrast the use of mitochondrial cytochrome *c* oxidase I (*COI*) marker, which has been used commonly to study inter- and intraspecific relationships in bumblebees (Lecocq *et al.*, 2013a, 2016; Dellicour *et al.*, 2015; Moreira *et al.*, 2015), with a new genome-wide dataset of restriction site-associated DNA sequencing (RADseq). This method readily provides thousands of single nucleotide polymorphisms (SNPs) and has proved to be effective in biogeography, adaptation, association and conservation studies, even when individual and population sampling is limited (Lozier, 2014; Woodard *et al.*, 2015; Lozier & Zayed, 2017). With RADseq, we obtained the first comprehensive dataset of nuclear loci from *B. t. lusitanicus* and evaluated its population structure at a finer-scale resolution. This is the first step to investigate the spatial patterns of population structure and genetic diversity on a bumblebee species within the Iberian Peninsula and to identify the main priorities for future research on the Iberian bumblebees' conservation, evolution and environmental adaptation.

MATERIAL AND METHODS

Sampling and DNA extraction

A total of 198 individuals of *B. t. lusitanicus* were collected from six regions within the Iberian Peninsula, covering most of the Iberian distribution and habitat heterogeneity of this subspecies. To minimize the probability of sampling individuals from the same colony, individuals were caught from locations separated by ≥ 1 km, within each region, whenever possible (Fig. 1B; Supporting Information, Table S1). Additional samples from other subspecies of *B. terrestris* were collected: one *B. t. terrestris* from Switzerland, from Germany and from France, one *B. t. dalmatinus* from Turkey, one *B. t. audax* from Great Britain and four *B. t. africanus* from Morocco (Fig. 1A; Supporting Information, Table S1). We focused on diploid individuals (females, mainly workers) for the genetic analyses, in order to capture the genetic variation in the populations. The only exception was one male from France, from where we did not collect females.

For the genetic analyses, *B. t. lusitanicus* samples were grouped according to the six defined regions, whereas samples from other subspecies were grouped according to the country where they were collected (Supporting Information, Table S1). Thirteen females from closely related species were collected to serve as an outgroup: one individual of *Bombus hortorum* (Linnaeus, 1761), three of *B. lapidarius*, three of *Bombus lucorum* (Linnaeus, 1761), four of *B. pascuorum* and two of *Bombus ruderatus* (Fabricius, 1775). Samples were preserved for DNA extraction either in absolute ethanol and stored at −20 °C or dried and stored at −80 °C. Total genomic DNA was extracted from fore- and mid-legs and the head; for smaller individuals, a portion of the thorax was also used. DNA was isolated with the DNeasy Blood & Tissue extraction kit (Qiagen), following the manufacturer's standard

protocol. To maximize DNA yield, some samples were eluted in a lower volume of buffer AE (minimum of 60 µL), and the eluted volume was transferred again to the silica column of the kit for a second elution, and incubation times with buffer AE were extended up to 30 min.

COI AMPLIFICATION, CONSTRUCTION OF RAD LIBRARIES AND SEQUENCING

For all samples, a fragment from the mitochondrial cytochrome *c* oxidase I (*COI*) gene was amplified by polymerase chain reaction (PCR) with the primer set LepF/LepR (Hajibabaei *et al.*, 2006). The PCR amplifications were carried out in 20 µL volumes containing ~10–45 ng of template DNA, 1× reaction buffer, 1.8 mM of $MgCl_2$, 1.0 mM dNTPs, 0.04 units of *GoTaq Flexi* DNA polymerase and 0.4 µM of each primer. The thermocycling profile consisted of one cycle of 1 min at 94 °C, five cycles of 30 s at 94 °C, 1 min at 45 °C and 1 min at 72 °C, followed by 30 cycles of 1 min at 94 °C, 1 min 30 s at 50 °C and 1 min at 72 °C, with a final step of 5 min at 72 °C. All PCR products were purified with the SureClean (Bioline) purification kit and sequenced in the forward direction on an ABI3730XL by Macrogen Europe. DNA sequences were quality controlled with *Sequencher* v.4.0.5 (Gene Codes Corporation).

A subset of 55 individuals, including 37 individuals of *B t. lusitanicus* from the six defined Iberian regions, five individuals from other *B. terrestris* subspecies and 13 individuals from outgroup species (see Fig. 1A, B; Supporting Information, Table S1) was used for RAD sequencing analyses. RADseq libraries for Illumina paired-end sequencing were prepared following the protocol described by Etter *et al.* (2011), available at https://www.wiki.ed.ac.uk/display/RADSequencing/Home, with some minor modifications as reported by Seabra *et al.* (2019). We used the restriction enzyme *Pst*I-HF (New England Biolabs). Sequencing took place on a Illumina HiSeq 2000/2005 at Edinburgh Genomics, Ashworth Laboratories. The 55 individuals were run together with another 53 samples from another study (Seabra *et al.*, 2019), over two lanes.

COI DATA ANALYSIS

Mitochondrial *COI* sequences obtained in this study were deposited in GenBank (accession numbers MN652675–MN652877; Supporting Information, Table S1a). We also included 17 sequences from our previous study (Seabra *et al.*, 2019), and we followed the designation of haplotypes in that study (Supporting Information, Table S1a). In order to extend our *B. terrestris COI* dataset (207 sequences) to the geographical range of the species, we downloaded from GenBank five additional sets of *COI* sequences from previous studies, namely Coppée (2010), Williams *et al.* (2012a, b), Moreira *et al.* (2015) and Schmidt *et al.* (2015) (Fig. 1; for GenBank accession numbers, see Supporting Information, Table S1b). In this way, 233 *COI* sequences were added to our dataset, giving a total of 441 sequences. Sequences related to commercially reared or introduced populations were not considered.

The *COI* dataset of *B. terrestris* was aligned in MAFFT v.7.271 (Katoh & Standley, 2013) using default settings. The final alignment was checked for accuracy, and sequences were trimmed to the same length (597 bp) to eliminate missing data, using BioEdit v.7.2.5 (Hall, 1999). Median-joining haplotype networks were constructed in PopART v.1.7 (Bandelt *et al.*, 1999; Leigh & Bryant, 2015), in order to visualize the relationship among *B. terrestris* haplotypes. We used ARLEQUIN v.3.5.2.2 (Excoffier & Lischer, 2010) and only the individuals from the Iberian Peninsula to perform a standard analysis of molecular variance (AMOVA) with 10 000 permutation steps and to calculate haplotype (h) and nucleotide (π) diversities. We also calculated haplotype and nucleotide diversities for the individuals of the remaining *B. terrestris* distribution range in order to make a comparison with the Iberian Peninsula. File format conversion for PopART and ARLEQUIN were performed using PGDSpider v.2.1.0.3 (Lischer & Excoffier, 2012).

RADSEQ DATA ANALYSIS

The RADseq data obtained in this study are available at Sequence Read Archive (PRJNA578045). All console commands used for RADseq data filtering, SNP discovery and subsequent analyses are available in the Supporting Information (Appendix S1). Quality control of the RADseq raw read data was performed using FastQC v.0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The script *process_radtags* implemented in STACKS v.1.29 (Catchen *et al.*, 2013) was used (with default settings) to remove low-quality (with a phred quality score < 33) and unidentifiable reads and to demultiplex the data. BOWTIE 2 v.2.1.0 (Langmead & Salzberg, 2012) was used to trim the last three bases from the 3′ end of each read before alignment, because this region is richer in low-quality base calls, and to align the reads for each sample to the assembled reference genome of *B. terrestris* (NCBI assembly GCA_000214255.1) with the 'sensitive' option. SAMtools v.0.1.19 (Li *et al.*, 2009) was used to remove low-quality alignments (mapping quality < 20) and any unmapped reads.

After data filtering, RAD loci were identified and SNPs called using the STACKS pipeline v.1.45 (Catchen *et al.*, 2013). In order to minimize the impact of differences among taxa in the number of SNPs obtained and the amount of missing data, three datasets were created from the initial 55 individuals for different analyses (Supporting Information, Table S1): (1) BT_OUT includes all the studied subspecies (all *B. terrestris*) and the outgroup species; (2) BT_SSP includes all *B. terrestris* subspecies and does not include the outgroup species; and (3) BT_BTL includes the Iberian *B. t. lusitanicus* only.

Preliminary tests were carried out to identify optimal STACKS parameters (Supporting Information, Appendix S2; Figs S1, S2, S3). In *pstacks*, the minimum stack depth was set to six. In *cstacks*, the maximum number of mismatches allowed when building catalogue loci was two. In *populations*, individuals were grouped into putative locations based on geographical region (see Supporting Information, Table S1), and one random SNP per RAD locus was used to avoid confounding signals of linkage disequilibrium. The SNPs were retained only if they were present in 50% of individuals in at least $N - 1$ (where N is the number of geographical regions) in all datasets. To test for differences among outputs with different randomly selected SNPs, the module *populations* was run several times, but no significant differences were obtained in the final results (data not shown).

Finally, we performed an additional filtering step to remove loci with minor allele frequencies < 0.05 and > 25% of missing data across all samples using VCFtools v.0.1.15 (Danecek *et al.*, 2011). VCFtools was also used to obtain the mean coverage per site per individual. The resulting SNP datasets were then used in subsequent population genomic and phylogeographical analyses: dataset BT_OUT to check for differentiation between *B. terrestris* and outgroup species; dataset BT_SSP to determine differentiation between *B. t. lusitanicus* and the other subspecies; and dataset BT_BTL to determine population genetic structure and genetic diversity of *B. t. lusitanicus* within the Iberian Peninsula.

Principal components analyses (PCAs) of the three RADseq datasets were performed using the package SNPRelate v.1.12.0 (Szulkin *et al.*, 2016) as implemented in the R script *snp_pca_static.R* (https://github.com/CoBiG2/RAD_Tools) as of commit 'bb2fc45'. To test for differentiation between *B. t. lusitanicus* and the other subspecies and within *B. t. lusitanicus*, clustering analyses of population structure were performed using the datasets BT_SSP and BT_BTL and *MavericK* v.1.0.4 (Verity & Nichols, 2016). Given that the model used by *MavericK* assumes that markers are neutral (Verity & Nichols, 2016), we first performed an outlier analysis of both datasets to identify and remove non-neutral markers, using both BayeScan v.2.1 (Foll & Gaggiotti, 2008) and SelEstim v.1.1.7 (Vitalis *et al.*, 2014). BayeScan was run using a matrix of SNP genotypes, with prior odds for the neutral model turned to five and assuming a detection threshold of 0.05. The remaining parameters were set to default values. Plots and convergence were checked using the R script *plot_R.r* available within the BayeScan package, and the package CODA v.0.19-1 (Plummer *et al.*, 2006). SelEstim was run after randomizing the reference allele for each locus (using the R script SelEstim.R available within SelEstim package) and using 50 pilot runs of 1000 length, followed by a main run of one million length, with a burn-in of 100 000, a thinning interval of 20 and a detection threshold of 0.01. The R script SelEstim.R was also used to obtain the list of outliers and check for convergence. The total number of outliers identified by BayeScan and SelEstim were removed from the datasets using the Python script *outlier_removal.py* (https://github.com/CoBiG2/RAD_Tools) as of commit 'ba731f2'. Datasets with only neutral markers were analysed using *MavericK* v.1.0.4 (Verity & Nichols, 2016), wrapped under *Structure_threader* v.1.2.4 (Pina-Martins *et al.*, 2017) for values of K (number of demes) between one and five for dataset BT_SSP, and values between one and four for the dataset BT_BTL. We first performed a single 'pilot' run of 5000 iterations, with a burn-in of 500 using an admixture model, a free α parameter of one with a standard deviation of the normal proposed distribution of 0.10 and 'thermodynamic integration' (TI) turned off. The posterior median and posterior standard deviation of α were obtained from the 'pilot' run and used in a 'tuned' run as parameters for the admixture model as follows: α was set to ten times the posterior median and alphaPropSD to five times the posterior SD. This 'tuned' run was composed of five runs of 10 000 iterations, with a burn-in of 2000, with TI turned on and another set of 20 runs of 10 000 samples, with a burn-in of 2000. The most suitable value of K was calculated for both datasets using the TI method. The R script *MavericK1.0_functions.R* available within the *MavericK* package was used to produce diagnostic plots in order to check for convergence and autocorrelation, and the Qmatrix plots.

Finally, a locus-by-locus AMOVA was performed in ARLEQUIN with RADseq dataset BT_BTL, which includes all the SNPs, in order to examine the variance within and among geographical regions, and significance was calculated using 10 000 permutation steps. Genome-wide measures of genetic diversity, including per-SNP nucleotide diversity (π_{SNP}), and the mean per-individual observed and expected heterozygosities (H_O and H_E) were calculated using VCFtools and the same dataset.

File format conversions for BayeScan, *MavericK* and ARLEQUIN were performed using PGDSpider, whereas file format conversion for SelEstim was performed with the bash script *GESTE2SelEstim.sh* (https://github.com/Telpidus/omics_tools) as of commit 'f74f66b'.

RESULTS

The final alignment of the mitochondrial *COI* dataset included no indels and consisted of a total of 26 variable sites, of which nine were parsimony informative. A total of 16 haplotypes were found (Fig. 2; Supporting Information, Table S1a and S1b).

After Illumina sequencing of RAD libraries for 55 individuals, we obtained an average of 7.7 million paired end reads of 125 bp, per individual. Of those, an average of 7.4 million were retained after filtering with *process_radtags*, representing ~8.18% of the genome sequenced. Quality scores of the retained reads ranged from 36 to 37, with a GC content of ~40%. An average of 52.69% of the quality-filtered reads aligned to the *B. terrestris* genome. Of the 55 sequenced individuals, nine were excluded from further analyses (six *B. t. lusitanicus* from four of the Iberian defined regions and three representatives of the outgroup species) owing to lower mean coverage per individual (≤ 12×). Missing data of the remaining samples averaged 5.4%. Information concerning the output of RADseq filtering steps for each sample (number of initial and mapped reads, coverage, datasets to which samples belong and missing data) are provided in the Supporting Information (Table S1). Final individual counts and statistics per dataset after filtering were as follows: BT_OUT dataset comprised 46 individuals and 5357 SNPs (~22.7 markers/Mb), with a mean coverage of 57.7× per site, per individual; BT_SSP dataset comprised 36 individuals and 10 765 SNPs (~45.6 markers/Mb), with a mean coverage of 52.3× per site, per individual; and BT_BTL dataset comprised 31 individuals and 11 369 SNPs (~48.2 markers/Mb), with a mean coverage of 56.6× per site, per individual.

DIFFERENTIATION OF *BOMBUS TERRESTRIS LUSITANICUS*

Seven of the 16 *B. t. lusitanicus COI* haplotypes were present in the Iberian Peninsula. The most common haplotype (H1) was common to the entire Peninsula and to the remaining European mainland regions analysed, whereas H2 was shared with central Europe (Switzerland) and Ireland (Fig. 2). The other five haplotypes found in Iberia were exclusive to this peninsula (Fig. 2). No shared haplotypes were detected between the Iberian Peninsula and North Africa. Haplotypes H9 and H16 were exclusive to North Africa, and H14 was shared between that region and Italy. Some haplotypes found in Great Britain, Sardinia, Corsica and the Canary Islands were also exclusively from those regions (Fig. 2A). Most of the haplotypes differed from H1 in only one or two mutational steps, with the exception of the haplotypes found only in islands or in North Africa, and the haplotypes from Sardinia and Morocco being the most differentiated (Fig. 2C).

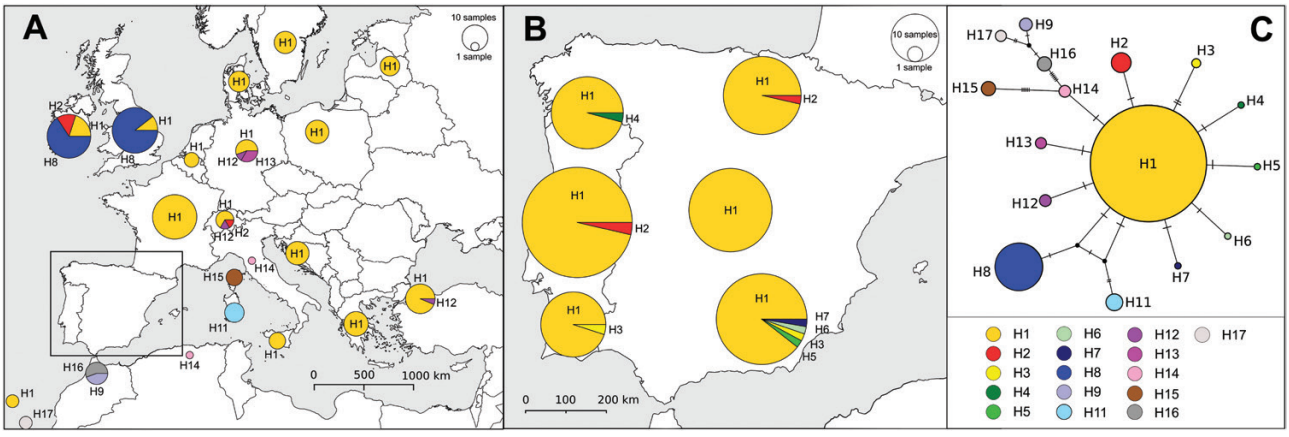

Figure 2. Geographical distribution and frequency of mitochondrial DNA *COI* haplotypes. A, map of haplotype frequencies for *Bombus terrestris* across Europe and North Africa. B, map of haplotype frequencies for each Iberian geographical region. B is a magnified version of the black rectangle in A. C, median-joining network representing the relationship among haplotypes, coloured by haplotype. Sequences from Coppée (2010), Williams *et al.* (2012a, b), Moreira *et al.* (2015) and Schmidt *et al.* (2015) were included. The size of the pie charts in A and B and of the circles in C is in proportion to the haplotype frequencies. Each haplotype is represented by the respective colour and designation (H1–H9 and H11–H17).

The PCA of RADseq dataset BT_OUT revealed a clear separation between *B. terrestris* and the outgroup species, with the exception of *B. lucorum*, which was the closest to *B. terrestris* (Fig. 3A). The PCA using dataset BT_SSP, with *B. terrestris* samples only, revealed a very narrow separation between samples from the Iberian Peninsula and those from Great Britain and Germany across EV2 (eigenvector 2; EV2 explained 3.08% of the variance). In contrast, individuals from Morocco showed a greater separation across EV1 (eigenvector 1) from the remaining samples (EV1 explained 10.14% of the variance; Fig. 3B). Outlier tests of dataset BT_SSP revealed that a total of 44 SNPs were non-neutral: 43 SNPs (0.40% of the total SNPs) when using SelEstim and two SNPs (0.02% of the total SNPs) when using BayeScan (one SNP was identified by both softwares). Clustering analyses of the dataset BT_SSP with non-neutral loci removed, using *MavericK*, determined the existence of two groups ($K = 2$) as the most likely scenario (Fig. 3C). These two groups corresponded to: (1) the individuals from the Iberian Peninsula, Great Britain and Germany; and (2) individuals from Morocco, which was consistent with what was observed in the PCA (Fig. 3B).

Genetic structure of Iberian populations

The most common *COI* haplotype, H1, was present in 95% of the *B. terrestris* sampled in the Iberian Peninsula. H2 was present in three samples from the north and west Iberian Peninsula regions, whereas H3 was present in only two samples from southern locations, one from the south-east Iberian Peninsula and another from the south-west. All the remaining Iberian *COI* haplotypes (H4–H7) were represented by a single sample (Fig. 2B). The Iberian region with the highest haplotype diversity was the south-east Iberian

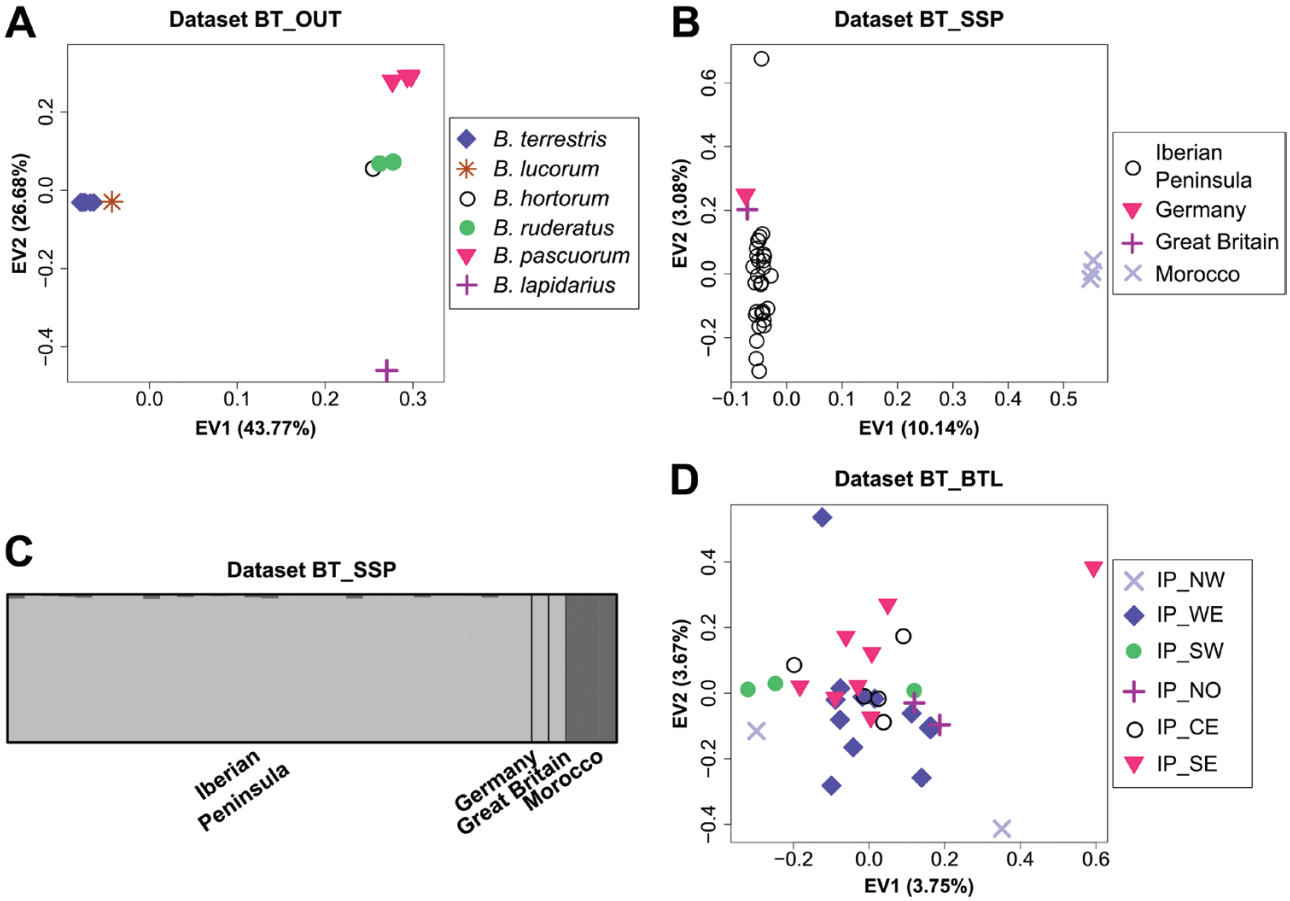

Figure 3. Principal components analyses (PCA) and *MavericK* analyses of RADseq data. A, PCA of dataset BT_OUT, comparing *Bombus terrestris* with other species. B, PCA of dataset BT_SSP comparing *Bombus terrestris lusitanicus* samples with samples of other *B. terrestris* subspecies. C, *MavericK* clustering plot of dataset BT_SSP, for $K = 2$. D, PCA of dataset BT_BTL comparing *B. t. lusitanicus* among Iberian geographical regions. In *MavericK* results, the estimated membership of each individual to each cluster is shown by vertical bars, with the clusters represented by different shades of grey.

Peninsula, with five haplotypes (Fig. 2B; Supporting Information, Table S2). The haplotype network did not show a structured phylogeographical pattern, with the most common haplotype being shared among geographically distant regions, and the less frequent and unique haplotypes being closely related to H1, in a 'star-like' configuration (Fig. 2C).

The PCA of RADseq dataset BT_BTL showed no clear separation between any Iberian geographical regions (EV1 and EV2 explained only 3.75 and 3.67%, respectively, of the variation; Fig. 3D). The most segregated individuals were BTL_075 (north-west Iberian Peninsula) and BTL_136 (south-east Iberian Peninsula) along EV1, and BTL_306 (west Iberian Peninsula) along EV2. Missing data values for these individuals did not explain their separation from the remainder (3.7, 5.7 and 10.6% respectively), and their *COI* haplotype was the most common one, H1. Outlier tests of dataset BT_BTL revealed a total of 31 non neutral SNPs: 29 (0.26% of the total SNPs) when using SelEstim, and two SNPs (0.02% of the total SNPs) when using BayeScan. Clustering analyses of the dataset BT_BTL with non-neutral loci removed, using *MavericK*, determined the existence of one group ($K = 1$) as the most likely scenario (data not shown), which was concordant with the PCA results.

Genetic diversity

The AMOVA results revealed an absence of genetic structure for the *B. terrestris COI* dataset, suggesting that the overall source of variation was within geographical regions instead of among them (Table 1). Measures for haplotype (h) and nucleotide (π) diversities calculated using the *B. terrestris COI* dataset for the whole Iberian Peninsula were 0.08470 and 0.00018, respectively (Table 2). More than 43% (seven of 16) of the haplotypes found were present in the Iberian Peninsula, and 31% (five of 16) were exclusive to this area. No other European region showed such a high number of haplotypes. It is important to note, however, that sample sizes differed substantially (Fig. 2; Table 2).

The AMOVA results using the RADseq dataset BT_BTL mirrored those of the *COI* dataset, with an absence of genetic structure, indicating that the overall source of variation was within and not among geographical regions (Table 1). Measures of per-SNP nucleotide diversity and mean per-individual observed and expected heterozygosities for the total Iberian Peninsula were 0.2780, 0.2326 and 0.2773, respectively (Table 2).

DISCUSSION

We conducted a population genetic study, with mitochondrial and nuclear genome-wide markers, to measure differentiation of *B. t. lusitanicus* from other *B. terrestris* subspecies and to investigate the population structure of this subspecies and its genetic diversity within the Iberian Peninsula. We found no evident differentiation pattern in mitochondrial DNA between *B. t. lusitanicus* and the other European mainland subspecies, *B. t. terrestris*, *B. t. dalmatinus* and *B. t. calabricus*. In contrast, we found a clear differentiation of North African *B. t. africanus* from the remaining subspecies, including the geographically close *B. t. lusitanicus*, which was in accordance with previous studies (Coppée, 2010; Lecocq *et al.*, 2016) and was confirmed here with samples from southern Iberia, where any evidence of admixture with *B. t. africanus* would be more likely to occur. Considering the genetic diversity of *B. terrestris* at the species level, it was homogeneous across mainland populations, whereas subspecies from the islands and North Africa appeared to be more differentiated, although with evidence of some admixture, particularly between the British and continental populations (as previously reported by Moreira *et al.*, 2015). The presence of the same haplotype in both Algeria and central Italy could also indicate some admixture in this region. Our study corroborated those of Estoup *et al.* (1996), Widmer *et al.* (1998), Moreira *et al.* (2015) and Lecocq *et al.* (2016) based on mitochondrial and microsatellite markers. Our genome-wide analyses with RAD sequencing showed very slight distinction of *B. t. lusitanicus* from *B. t. terrrestris* (Germany) and *B. t. audax* (Great Britain) samples in the PCA, which is not supported by *Maverick* results, but corroborated the clear differentiation from *B. t. africanus*. However, larger sample sizes of *B. t. terrestris* and *B. t. audax*

Table 1. Results of the analyses of molecular variance considering mitochondrial DNA *COI* data and RADseq data

DNA marker	Source of variation	Sum of squares	Variation components	Percentage of variation
COI	Within geographical regions	19.62	0.06	100.03
	Among geographical regions	0.27	−0.00002	−0.03
RADseq	Within geographical regions	79454.05	1532.05	97.81
	Among geographical regions	9121.49	34.29	2.19

Table 2. The sample size and diversity indices across *Bombus terrestris lusitanicus* samples from the Iberian Peninsula considering mitochondrial DNA *COI* data and RADseq data, along with data for other populations of *B. terrestris* or for other species (from Lozier, 2014 and Jackson *et al.*, 2018)

Species	Geographical region	Reference	Sample size	DNA marker	h (number of haplotypes)	π	π_{SNP}	H_O	H_E
B. t. lusitanicus	Iberian Peninsula	This study	208	*COI*	0.08470 (seven haplotypes)	0.00018	–	–	–
Bombus terrestris	Europe (continental)	This study	340	*COI*	0.09740 (ten haplotypes)	0.00034	–	–	–
Bombus terrestris	Europe (continental + islands) and North Africa	This study	441	*COI*	0.39900 (16 haplotypes)	0.00209	–	–	–
B. t. lusitanicus	Iberian Peninsula	This study	31 ♀	RADseq, *Pst*I(10 938 SNPs)	–	–	0.278	0.233	0.277
Bombus impatiens	Eastern USA	Lozier *et al.* (2014)	24 ♀	RADseq, *Sgr*AI(2387–9148 SNPs)	–	–	0.136–0.289	n.d.	n.d.
Bombus pensylvanicus	Eastern USA	Lozier *et al.* (2014)	24 ♀	RADseq, *Sgr*AI(3240–9376 SNPs)	–	–	0.135–0.276	n.d.	n.d.
Bombus bifarius	Mountain regions of western USA	Jackson *et al.* (2018)	383 ♀	RADseq, *Pst*I(598–37 474 SNPs)	–	–	0.122–0.140	n.d.	n.d.
Bombus vosnesenskii	Mountain regions of western USA	Jackson *et al.* (2018)	587 ♀	RADseq, *Pst*I(356–18 700 SNPs)	–	–	0.105–0.116	n.d.	n.d.

Note that for RADseq markers, the restriction enzyme used and the number of SNPs obtained are also indicated.
Abbreviations: h, haplotype diversity; H_E, mean per-individual expected heterozygosity; H_O, mean per-individual observed heterozygosity; n.d., no data; π, nucleotide diversity; π_{SNP}, per-single nucleotide polymorphism nucleotide diversity; SNP, single nucleotide polymorphism; '–', 'non-applicable'.

are needed to evaluate this small differentiation at the genomic level better.

The lack of differentiation of *B. t. lusitanicus* from other European mainland subspecies can be explained by: (1) a common origin, with subsequent local differentiation, which is supported by the star-like pattern in the mitochondrial *COI* network, with rarer haplotypes deriving from a single ancestral haplotype (H1); (2) high dispersal ability of these insects across large distances (Kraus *et al.*, 2009; Lepais *et al.*, 2010) and extensive mountain ranges, such as the Pyrenees; or (3) erosion of genetic differentiation caused by hybridization with commercial hives of allochthonous origin, which are used in several areas in Europe for crop pollination [commercial stocks used in the Iberian Peninsula include mostly subspecies *B. t. terrestris* and *B. t. dalmatinus* (Velthuis & van Doorn, 2006; Lecocq *et al.*, 2016)]. Putative hybrids with commercial hives in the western Iberian Peninsula have already been detected (Seabra *et al.*, 2019), but a widespread genetic erosion is not expected because the transfer of colonies of this species across Europe for crop pollination is a relatively recent phenomenon (Estoup, 1996).

In contrast, the differentiation found between *B. t. lusitanicus* and *B. t. africanus* suggests that gene flow between the Iberian Peninsula and North Africa is much lower than gene flow to elsewhere in mainland Europe. The number of accumulated differences in mitochondrial DNA also suggests that these two subspecies probably started to diverge earlier than the others. The Mediterranean sea thus seems to be an effective barrier to gene flow, although the two continents are geographically very close at the Strait of Gibraltar (< 15 km at the closest point), and despite the good dispersal capability of bumblebees. According to the information retrieved from http://www.atlashymenoptera.net/, *B. terrestris* is currently present right up to the coast on both sides of the Strait. Both Estoup *et al.* (1996) and Moreira *et al.* (2015) reported evidence of *B. terrestris* migrating over the sea, across the English Channel and between the Isle of Man and Ireland, albeit at recognizably very low rates. Also, other bumblebee species, such as *Bombus jonellus* (Kirby, 1802), were found to be able to disperse over sea barriers ≤ 30 km (Darvill *et al.*, 2010). Given that bumblebees are known to disperse over such large distances, two hypotheses might explain this result: (1) migration is conditioned by the prevailing wind conditions, characterized by strong winds from easterly or westerly directions, which are known to have an important role in dispersion patterns of several species (e.g. the moth *Cornifrons ulceratalis*; Dantart *et al.*, 2009); or (2) migration occurs between both continents, but local differences in environmental conditions and/or sexual selection might be acting against migrants and preventing effective gene flow. The Strait of Gibraltar seems to hinder the dispersal of other flying species between North Africa and the Iberian Peninsula (e.g. the butterfly *Pararge aegeria*; Weingartner *et al.*, 2006), whereas it has acted as a route of dispersal for others, mainly during periods of lower sea level (e.g. the Iberian honey bee *Apis mellifera iberiensis*; Chávez-Galarza *et al.*, 2015).

We did not find population genetic structure within *B. t. lusitanicus* across the Iberian Peninsula, on the contrary to what was reported for the Iberian honey bee *A. m. iberiensis* (Chávez-Galarza *et al.*, 2015). In this latter species, two highly divergent genetic lineages are observed, which form a north-eastern–south-eastern cline, better explained by secondary contacts between divergent populations from distinct and isolated glacial refugia (Chávez-Galarza *et al.*, 2015). The panmictic pattern of *B. t. lusitanicus* within Iberia is probably attributable to the capacity of *B. terrestris* for long-distance flights, coupled with the absence of effective geographical barriers to its dispersal. Long-distance flights of queens and males contribute towards regular gene flow and have been suggested to be sufficient to maintain genetic cohesion of common bumblebee species over large areas (Lepais *et al.*, 2010). Also, the fact that *B. terrestris* is a short-tongued generalist bumblebee (Chapman *et al.*, 2003), having a large foraging range (Walther-Hellwig & Frankl, 2000), probably increases its capacity to find suitable habitats in a variety of conditions. In addition, this species tolerates a broad range of climates (Penado *et al.*, 2016), from Mediterranean beaches at high temperatures to high mountains of crio–oro-Mediterranean regimens. Thus, individuals are more capable of dispersing and occupying a vast majority of habitats when compared with other species, contributing to the observed large-scale connectivity.

The above-mentioned hypothesis of a recent population expansion from a single periglacial refuge in the Iberian Peninsula could also explain not only the observed pattern of homogeneity but also the star-like pattern in the mitochondrial *COI* network. This refuge might have existed in the Iberian Peninsula or elsewhere in Europe (as also discussed by Estoup *et al.*, 1996), because the most common haplotype, H1, is widespread across the continent. Nonetheless, some rarer haplotypes could have evolved in a smaller refuge in the Betic ranges of southern Spain ('refugia within refugia' paradigm of Gómez & Lunt, 2007), which might explain the higher genetic diversity found in mitochondrial DNA for the south-eastern Iberian Peninsula region (Fig. 2B; Table S2). This region is characterized by semi-arid lowlands drastically contrasting with steep changes in vegetation and climate along an elevational cline. It is considered to be

a hotspot for Mediterranean biodiversity, harbouring many endemic species or lineages, and therefore, the high genetic diversity found in this region (five haplotypes from a total of ten in continental Europe) is not unexpected (Hewitt, 2011; Nunes *et al.*, 2014).

The hypothesis of admixture between this region and North Africa is unlikely, because there are no haplotypes shared between both regions, as already mentioned. The use of commercial hives of *B. terrestris* for crop pollination in the south-east Iberian Peninsula, and in the south-west and west Iberian Peninsula, is a common practice (Cejas *et al.*, 2018). In these regions, commercial bumblebees have been found foraging outside greenhouses and on natural habitats, and introgression between commercial and native bumblebees was detected (Cejas *et al.*, 2018, 2019; Seabra *et al.*, 2019;Trillo et al., 2019). Thus some of the variation found in the south-east Iberian Peninsula could have been introduced artificially. We found one *COI* haplotype (H3) in the south-east and south-west Iberian Peninsula regions that is also relatively common in commercial stocks and in individuals collected from greenhouse areas investigated by Seabra *et al.* (2019). This haplotype was found in two specimens: one collected ~300 km from the area where Cejas *et al.* (2018, 2019) detected potential hybrids between commercial and native populations, based on morphological and mitochondrial 16S data; and one collected ~30 km from the area where Seabra *et al.* (2019) also detected potential introgression between native and escaped individuals, based on RADseq data. None of the other unique haplotypes from the south-east Iberian Peninsula was found within the commercial samples analysed by Seabra *et al.* (2019). Moreira *et al.* (2015) found that commercially reared populations were differentiated from the majority of the wild populations from Ireland, having the highest number of private microsatellite alleles. Thus the introduction of non-native specimens can lead to changes in the genetic structure of the native ones and, ultimately, increase the risk of displacement and the consequent loss of rare beneficial alleles, especially in populations with low genetic diversity. Consequently, we cannot underestimate the potential impact of the use of allochthonous commercial bumblebees for local crop pollination, independent of their origin.

The extended sampling in the Iberian Peninsula revealed this region as one of the richest in genetic diversity for *B. terrestris*, with a higher number of mitochondrial *COI* haplotypes than any other region in Europe studied so far (although some mainland regions remain undersampled). Iberia seems to be an important source of diversity for the global genetic pool of this species, because rare alleles might play a role in population resilience against human- or climate-mediated changes (Barret & Schluter, 2008), especially at the extremes of the species range. Given that this is the first study to evaluate population genetic diversity in *B. terrestris* with RADseq, we were not able to compare our results with other regions in Europe regarding the diversity at the genomic level. When compared with North American *Bombus* species (Lozier, 2014; Jackson *et al.*, 2018), despite the different RADseq markers used, *B. t. lusitanicus* from the Iberian Peninsula showed similar or higher diversity values, even when comparing with *Bombus impatiens* Cresson, 1863, a common species in eastern North America.

Further ecological studies comparing habitats, phenology and phenotypic characteristics of *B. terrestris* from the south of the peninsula with those from North Africa could help to describe barriers to dispersion and to gene flow within this species in more detail. Also, the absence of population genetic structure will facilitate analyses of the adaptive potential of *B. t. lusitanicus* to environmental changes within the Iberian Peninsula, by finding adaptive genetic diversity and by modelling the response of the species to future changes in land use and/or climate. Ecological and genetic studies focusing on arid regions or on other Mediterranean peninsulas are also needed, in order to understand how environmental change is affecting natural populations of *B. terrestris*. It would also be helpful to evaluate the impacts of global warming on the efficiency of crop pollination of commercial hives at the extremes of this species range and in economically important regions.

ACKNOWLEDGEMENTS

This work was funded by national funds through Fundação para a Ciência e a Tecnologia [FCT; project UID/BIA/00329/2013 (2015–2018); UID/AGR/04129/2013; grant PD/BD/113548/2015, under the PhD programme 'Biology and Ecology of Global Changes', Universidade Aveiro and Universidade Lisbon, Portugal, attributed to S.E.S.; grant SFRH/BPD/108413/2015, attributed to S.G.S.] and through Programa de Desenvolvimento Rural (ProDer; project ref. ProDeR 4.1. 46221-3). L.G.C. was supported by Fundação para Ciência e Tecnologia (FCT) and European Union via the programa operacional regional de Lisboa 2014/2020 (project EUCLIPO-028360). Thanks are due for the financial support also to the Centre for Environmental and Marine Research (UID/AMB/50017/2019), to Fundação para a Ciência e a Tecnologia (FCT)/ Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through national funds, and the co-funding by the Fundo Europeu de Desenvolvimento Regional (FEDER), within the PT2020 Partnership Agreement and Compete 2020.

We thank Edinburgh Genomics, particularly Karim Gharbi, for their assistance with RAD sequencing and Moises Mallo (Instituto Gulbenkian de Ciência) for the use of Bioruptor. We also thank the anonymous referees for helpful comments on an earlier version of the manuscript.

AUTHOR CONTRIBUTIONS

S.E.S., S.G.S., L.G.C. and O.S.P. designed the study. S.E.S., V.L.N., R.M., A.S.B.R., E.M., S.Y., T.G.L., E.F., M.T.R. and O.S.P. were responsible for sampling. S.E.S., S.G.S. and V.L.N. were responsible for DNA extraction and mitochondrial DNA amplification. V.L.N. constructed RAD libraries. S.E.S. and S.G.S. performed the bioinformatic analyses, with important contributions from V.L.N., F.P.-M. and O.S.P. S.E.S. wrote the manuscript, with contributions from all the other authors.

REFERENCES

Arias MC, **Rinderer TE**, **Sheppard WS. 2006.** Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research* **45:** 188–196.

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

Barrett RD, **Schluter D. 2008.** Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23:** 38–44.

Bommarco R, **Lundin O**, **Smith HG**, **Rundlöf M. 2012.** Drastic historic shifts in bumble-bee community composition in Sweden. *Proceedings of the Royal Society B: Biological Sciences* **279:** 309–315.

Catchen J, **Hohenlohe PA**, **Bassham S**, **Amores A**, **Cresko WA. 2013.** Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22:** 3124–3140.

Cejas D, **López-López A**, **Muñoz I**, **Ornosa C**, **De la Rua P. 2019.** Unveiling introgression in bumblebee (*Bombus terrestris*) populations through mitogenome-based markers. *Animal Genetics*. Advanced online publication. DOI: 10.1111/age.12874.

Cejas D, **Ornosa C**, **Muñoz I**, **De la Rua P. 2018.** Searching for molecular markers to differentiate *Bombus terrestris* (Linnaeus) subspecies in the Iberian Peninsula. *Sociobiology* **65:** 558–565.

Chapman RE, **Wang J**, **Bourke AF. 2003.** Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. *Molecular Ecology* **12:** 2801–2808.

Chávez-Galarza J, **Henriques D**, **Johnston JS**, **Carneiro M**, **Rufino J**, **Patton JC**, **Pinto MA. 2015.** Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. *Molecular Ecology* **24:** 2973–2992.

Coppée A. 2010. Bombus terrestris *(L. 1758): A complex species or a species complex? - Intraspecific pheromonal and genetic variations of* Bombus terrestris *(L.). Impacts on the speciation.* Unpublished PhD thesis. Laboratoire de Zoologie, Université de Mons.

Danecek P, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G**, **Durbin R**; 1000 Genomes Project Analysis Group. **2011.** The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

Dantart J, **Stefanescu C**, **Àvila A**, **Alarcón M. 2009.** Long-distance wind-borne dispersal of the moth *Cornifrons ulceratalis* (Lepidoptera: Crambidae: Evergestinae) into the northern Mediterranean. *European Journal of Entomology* **106:** 225–229.

Darvill B, **O'Connor S**, **Lye GC**, **Waters J**, **Lepais O**, **Goulson D. 2010.** Cryptic differences in dispersal lead to differential sensitivity to habitat fragmentation in two bumblebee species. *Molecular Ecology* **19:** 53–63.

Dellicour S, **Michez D**, **Mardulyn P. 2015.** Comparative phylogeography of five bumblebees: impact of range fragmentation, range size and diet specialization. *Biological Journal of the Linnean Society* **116:** 926–939.

Estoup A, **Solignac M**, **Cornuet JM**, **Goudet J**, **Scholl A. 1996.** Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. *Molecular Ecology* **5:** 19–31.

Etter PD, **Bassham S**, **Hohenlohe PA**, **Johnson EA**, **Cresko WA. 2011.** SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology* **772:** 157–178.

Excoffier L, **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.

Foll M, **Gaggiotti O. 2008.** A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180:** 977–993.

Gómez A, **Lunt DH. 2007.** Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, eds. *Phylogeography of southern European refugia*. Amsterdam: Springer, 155–188.

Goulson D. 2010. Impacts of non-native bumblebees in Western Europe and North America. *Applied Entomology and Zoology* **45:** 7–12.

Goulson D, **Lye GC**, **Darvill B. 2008.** Decline and conservation of bumble bees. *Annual Review of Entomology* **53:** 191–208.

Hajibabaei M, **Janzen DH**, **Burns JM**, **Hallwachs W**, **Hebert PD. 2006.** DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 968–971.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907–913.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87–112.

Hewitt GM. 2011. Mediterranean peninsulas: the evolution of hotspots. In: Zachos FE, Habel JC, eds. *Biodiversity hotspots*. Amsterdam: Springer, 123–147.

Ings TC, **Schikora J**, **Chittka L. 2005.** Bumblebees, humble pollinators or assiduous invaders? A population comparison of foraging performance in *Bombus terrestris*. *Oecologia* **144:** 508–516.

Jackson JM, **Pimsler ML**, **Oyen KJ**, **Koch-Uhuad JB**, **Herndon JD**, **Strange JP**, **Dillon ME**, **Lozier JD. 2018.** Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. *Molecular Ecology* **27:** 2926–2942.

Jaffé R, **Pope N**, **Acosta AL**, **Alves DA**, **Arias MC**, **De la Rúa P**, **Francisco FO**, **Giannini TC**, **González-Chaves A**, **Imperatriz-Fonseca VL**, **Tavares MG**, **Jha S**, **Carvalheiro LG. 2016.** Beekeeping practices and geographic distance, not land use, drive gene flow across tropical bees. *Molecular Ecology* **25:** 5345–5358.

Katoh K, **Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

Kraus FB, **Wolf S**, **Moritz RF. 2009.** Male flight distance and population substructure in the bumblebee *Bombus terrestris*. *The Journal of Animal Ecology* **78:** 247–252.

Langmead B, **Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9:** 357–359.

Lecocq T, **Coppée A**, **Michez D**, **Brasero N**, **Rasplus J-Y**, **Valterová I**, **Rasmont P. 2016.** The alien's identity: consequences of taxonomic status for the international bumblebee trade regulations. *Biological Conservation* **195:** 169–176.

Lecocq T, **Dellicour S**, **Michez D**, **Lhomme P**, **Vanderplanck M**, **Valterová I**, **Rasplus JY**, **Rasmont P. 2013a.** Scent of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (*Bombus lapidarius*). *BMC Evolutionary Biology* **13:** 263.

Lecocq T, **Gérard M**, **Michez D**, **Dellicour S. 2017.** Conservation genetics of European bees: new insights from the continental scale. *Conservation Genetics* **18:** 585–596.

Lecocq T, **Vereecken NJ**, **Michez D**, **Dellicour S**, **Lhomme P**, **Valterová I**, **Rasplus JY**, **Rasmont P. 2013b.** Patterns of genetic and reproductive traits differentiation in Mainland vs. Corsican populations of bumblebees. *PLoS ONE* **8:** e65642.

Leigh JW, **Bryant D. 2015.** PopART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6:** 1110–1116.

Lepais O, **Darvill B**, **O'Connor S**, **Osborne JL**, **Sanderson RA**, **Cussans J**, **Goffe L**, **Goulson D. 2010.** Estimation of bumblebee queen dispersal distances using sibship reconstruction method. *Molecular Ecology* **19:** 819–831.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**; 1000 Genome Project Data Processing Subgroup. **2009.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lischer HE, **Excoffier L. 2012.** PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28:** 298–299.

Lozier JD. 2014. Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology* **23:** 788–801.

Lozier JD, **Zayed A. 2017.** Bee conservation in the age of genomics. *Conservation Genetics* **18:** 713–729.

Maebe K, **Meeus I**, **Ganne M**, **De Meulemeester T**, **Biesmeijer K**, **Smagghe G. 2015.** Microsatellite analysis of museum specimens reveals historical differences in genetic diversity between declining and more stable *Bombus* species. *PLoS ONE* **10:** e0127870.

Martins IS, **Proença V**, **Pereira HM. 2014.** The unusual suspect: land use is a key predictor of biodiversity patterns in the Iberian Peninsula. *Acta Oecologica* **61:** 41–50.

Miraldo A, **Hewitt GM**, **Paulo OS**, **Emerson BC. 2011.** Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology* **11:** 170.

Moreira AS, **Horgan FG**, **Murray TE**, **Kakouli-Duarte T. 2015.** Population genetic structure of *Bombus terrestris* in Europe: isolation and genetic differentiation of Irish and British populations. *Molecular Ecology* **24:** 3257–3268.

Murray TE, **Coffey MF**, **Kehoe E**, **Horgan FG. 2013.** Pathogen prevalence in commercially reared bumble bees and evidence of spillover in conspecific populations. *Biological Conservation* **159:** 269–276.

Newbold T, **Hudson LN**, **Hill SL**, **Contu S**, **Lysenko I**, **Senior RA**, **Börger L**, **Bennett DJ**, **Choimes A**, **Collen B**, **Day J**, **De Palma A**, **Díaz S**, **Echeverria-Londoño S**, **Edgar MJ**, **Feldman A**, **Garon M**, **Harrison ML**, **Alhusseini T**, **Ingram DJ**, **Itescu Y**, **Kattge J**, **Kemp V**, **Kirkpatrick L**, **Kleyer M**, **Correia DL**, **Martin CD**, **Meiri S**, **Novosolov M**, **Pan Y**, **Phillips HR**, **Purves DW**, **Robinson A**, **Simpson J**, **Tuck SL**, **Weiher E**, **White HJ**, **Ewers RM**, **Mace GM**, **Scharlemann JP**, **Purvis A. 2015.** Global effects of land use on local terrestrial biodiversity. *Nature* **520:** 45–50.

Nunes VL, **Mendes R**, **Marabuto E**, **Novais BM**, **Hertach T**, **Quartau JA**, **Seabra SG**, **Paulo OS**, **Simões PC. 2014.** Conflicting patterns of DNA barcoding and taxonomy in the cicada genus *Tettigettalna* from Southern Europe (Hemiptera: Cicadidae). *Molecular Ecology Resources* **14:** 27–38.

Penado A, **Rebelo H**, **Goulson D. 2016.** Spatial distribution modelling reveals climatically suitable areas for bumblebees in undersampled parts of the Iberian Peninsula. *Insect Conservation and Diversity* **9:** 391–401.

Pina-Martins F, **Silva DN**, **Fino J**, **Paulo OS. 2017.** *Structure_threader*: an improved method for automation and parallelization of programs STRUCTURE, FASTSTRUCTURE and MavericK on multicore CPU systems. *Molecular Ecology Resources* **17:** e268–e274.

Pinto MA, **Henriques D**, **Neto M**, **Guedes H**, **Muñoz I**, **Azevedo JC**, **De La Rúa P. 2013.** Maternal diversity

patterns of Ibero-Atlantic populations reveal further complexity of Iberian honeybees. *Apidologie* **44:** 430–439.

Plummer M, **Best N**, **Cowles K**, **Vines K. 2006.** CODA: convergence diagnosis and output analysis for MCMC. *R News* **6:** 7–11.

Potts SG, **Biesmeijer JC**, **Kremen C**, **Neumann P**, **Schweiger O**, **Kunin WE. 2010.** Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25:** 345–353.

Rasmont P, **Coppée A**, **Michez D**, **De Meulemeester T. 2008.** An overview of the *Bombus terrestris* (L. 1758) subspecies (Hymenoptera: Apidae). *Annales de La Société Entomologique de France (N.S.)* **44:** 243–250.

Rasmont P, **Franzén M**, **Lecocq T**, **Harpke A**, **Roberts SPM**, **Biesmeijer K**, **Cederberg B**, **Dvorák L**, **Fitzpatrick Ú**, **Gonseth Y**, **Haubruge E**, **Mahé G**, **Manino A**, **Michez D**, **Neumayer J**, **Ødegaard F**, **Paukkunen J**, **Pawlikowski T**, **Potts SG**, **Reemer M**, **Settele J**, **Straka J**, **Schweiger O. 2015.** Climatic risk and distribution atlas of European bumblebees. *BioRisk* **10**(Special Issue): 1–236.

Rodrigues AS, **Silva SE**, **Marabuto E**, **Silva DN**, **Wilson MR**, **Thompson V**, **Yurtsever S**, **Halkka A**, **Borges PA**, **Quartau JA**, **Paulo OS**, **Seabra SG. 2014.** New mitochondrial and nuclear evidences support recent demographic expansion and an atypical phylogeographic pattern in the spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae). *PLoS ONE* **9:** e98375.

Schmidt S, **Schmid-Egger C**, **Morinière J**, **Haszprunar G**, **Hebert PD. 2015.** DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea *partim*). *Molecular Ecology Resources* **15:** 985–1000.

Seabra SG, **Silva SE**, **Nunes VL**, **Sousa VC**, **Martins J**, **Marabuto E**, **Rodrigues ASB**, **Pina-Martins F**, **Laurentino TG**, **Rebelo MT**, **Figueiredo E**, **Paulo OS. 2019.** Genomic signatures of introgression between commercial and native bumblebees, *Bombus terrestris*, in western Iberian Peninsula—implications for conservation and trade regulation. *Evolutionary Applications* **12:** 679–691.

Szulkin M, **Gagnaire PA**, **Bierne N**, **Charmantier A. 2016.** Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Molecular Ecology* **25:** 542–558.

Trillo A, **Montero-Castaño A**, **González-Varo JP**, **González-Moreno P**, **Ortiz-Sánchez FJ**, **Vilà M. 2019.** Contrasting occurrence patterns of managed and native bumblebees in natural habitats across a greenhouse landscape gradient. *Agriculture, Ecosystems and Environment* **272:** 230–236.

Velthuis HHW, **van Doorn A. 2006.** A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie* **37:** 421–451.

Verity R, **Nichols RA. 2016.** Estimating the number of subpopulations (*K*) in structured populations. *Genetics* **203:** 1827–1839.

Vitalis R, **Gautier M**, **Dawson KJ**, **Beaumont MA. 2014.** Detecting and measuring selection from gene frequency data. *Genetics* **196:** 799–817.

Walther-Hellwig K, **Frankl R. 2000.** Foraging habitats and foraging distances of bumblebees, *Bombus* spp. (Hym., Apidae), in an agricultural landscape. *Journal of Applied Entomology* **124:** 299–306.

Weingartner E, **Wahlberg N**, **Nylin S. 2006.** Speciation in *Pararge* (Satyrinae: Nymphalidae) butterflies—North Africa is the source of ancestral populations of all *Pararge* species. *Systematic Entomology* **31:** 621–632.

Widmer A, **Schmid-Hempel P**, **Estoup A**, **Scholls A. 1998.** Population genetic structure and colonization history of *Bombus terrestris* s.l. (Hymenoptera: Apidae) from the Canary Islands and Madeira. *Heredity* **81:** 563–572.

Williams PH, **An J**, **Brown MJ**, **Carolan JC**, **Goulson D**, **Huang J**, **Ito M. 2012a.** Cryptic bumblebee species: consequences for conservation and the trade in greenhouse pollinators. *PLoS ONE* **7:** e32992.

Williams PH, **Brown MJF**, **Carolan JC**, **An J**, **Goulson D**, **Aytekin AM**, **Best LR**, **Byvaltsev AM**, **Cederberg B**, **Dawson R**, **Huang J**, **Ito M**, **Monfared A**, **Raina RH**, **Schmid-Hempel P**, **Sheffield CS**, **Šima P**, **Xie Z. 2012b.** Unveiling cryptic species of the bumblebee subgenus *Bombus s. str*. worldwide with COI barcodes (Hymenoptera: Apidae). *Systematics and Biodiversity* **10:** 21–56.

Woodard SH, **Lozier JD**, **Goulson D**, **Williams PH**, **Strange JP**, **Jha S. 2015.** Molecular tools and bumble bees: revealing hidden details of ecology and evolution in a model system. *Molecular Ecology* **24:** 2916–2936.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. List of samples used in this study. Table S1a, list of collected samples. Table S1b, list of samples with sequences downloaded from GenBank.
Table S2. The sample size and diversity indices across *B. t. lusitanicus* samples by Iberian region considering mitochondrial DNA *COI* data and RADseq dataset BT_BTL.
Appendix S1. List of command line commands used for RADseq dataset analyses.
Appendix S2. Preliminary tests to identify optimal STACKS parameters.
Figure S1. Results of parameter tests for the STACKS module *pstacks*.
Figure S2. Results of parameter tests for the STACKS modules *ctsacks* and *sstacks*.
Figure S3. Results of parameter tests for the STACKS module *populations*.