MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2973–2992 doi: 10.1111/mec.13223

Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia

JULIO CHÁVEZ-GALARZA,*† DORA HENRIQUES,*† J. SPENCER JOHNSTON,‡ MIGUEL CARNEIRO,§ JOSÉ RUFINO,¶ JOHN C. PATTON** and M. ALICE PINTO*

**Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, †Centre of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar 4710-057 Braga, Portugal, ‡Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA, §CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, University of Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal, ¶Polytechnic Institute of Bragança, 5301-857 Bragança, Portugal, **Department of Forestry and Natural Resources, Purdue University, 715 W State St., West Lafayette, IN 4797-2061, USA*

Abstract

Dissecting diversity patterns of organisms endemic to Iberia has been truly challenging for a variety of taxa, and the Iberian honey bee is no exception. Surveys of genetic variation in the Iberian honey bee are among the most extensive for any honey bee subspecies. From these, differential and complex patterns of diversity have emerged, which have yet to be fully resolved. Here, we used a genome-wide data set of 309 neutrally tested single nucleotide polymorphisms (SNPs), scattered across the 16 honey bee chromosomes, which were genotyped in 711 haploid males. These SNPs were analysed along with an intergenic locus of the mtDNA, to reveal historical patterns of population structure across the entire range of the Iberian honey bee. Overall, patterns of population structure inferred from nuclear loci by multiple clustering approaches and geographic cline analysis were consistent with two major clusters forming a well-defined cline that bisects Iberia along a northeastern–southwestern axis, a pattern that remarkably parallels that of the mtDNA. While a mechanism of primary intergradation or isolation by distance could explain the observed clinal variation, our results are more consistent with an alternative model of secondary contact between divergent populations previously isolated in glacial refugia, as proposed for a growing list of other Iberian taxa. Despite current intense honey bee management, human-mediated processes have seemingly played a minor role in shaping Iberian honey bee genetic structure. This study highlights the complexity of the Iberian honey bee patterns and reinforces the importance of Iberia as a reservoir of *Apis mellifera* diversity.

Keywords: *Apis mellifera iberiensis*, geographic cline analysis, honey bee, Iberia, secondary contact, SNPs, sPCA, structure

Received 3 August 2014; revision received 16 April 2015; accepted 21 April 2015

Introduction

Clinal patterns in gene frequencies can be generated by random genetic drift under an isolation-by-distance scenario. Alternatively, clinal variation may be shaped by selection acting within a continuous population (primary intergradation) or, more frequently, may originate from contact between populations that diverged in allopatry (secondary contact). Distinguishing primary intergradation from secondary contact can, however, be a difficult undertaking because both processes may

Correspondence: M. Alice Pinto, Fax: +351273325405; E-mail: apinto@ipb.pt

generate similar patterns of genetic variation (Endler 1977; Barton & Hewitt 1985, 1989). Population genomics provides a suitable framework in which to more effectively unravel such levels of complexity. In population genomics, outlier tests are applied to genome-wide sampling of multiple populations to dissect out adaptive variation, leaving a background of neutral and near-neutral variation (Luikart *et al.* 2003). Cline analysis can then help reveal whether dissected patterns of variation originated from secondary contact or primary intergradation. If multiple coincident clines are identified (Endler 1977; Barton & Hewitt 1985) and these clines reflect changes in neutral loci, there is strong support for recent secondary contact (Durrett *et al.* 2000). Unless many independent loci respond similarly to a single environmental gradient or mosaic, clinal patterns of neutral variation and multiple coincident clines are not expected when primary intergradation is the leading process shaping variation (Durrett *et al.* 2000).

The Iberian Peninsula provides one of the most interesting settings in Europe for studying contact zones. High geological, physiographical and climatic complexity and diversity, together with isolation from Europe and proximity to Africa (especially at the Strait of Gibraltar), made this southernmost European region an important refuge during the Quaternary glaciations (reviewed by Hewitt 2000) and a bridge, for the more vagile organisms, between the two continents (Carranza *et al.* 2004; Cosson *et al.* 2005; Guillaumet *et al.* 2006; Whitfield *et al.* 2006; Wallberg *et al.* 2014). These features made Iberia not only a place of divergence during periods of isolation but also a contact zone during periods of expansion as reported for a wide array of plant and animal taxa (extensively reviewed by Weiss & Ferrand 2007), including the focal organism of this study: the Iberian honey bee, *Apis mellifera iberiensis*.

Disentangling diversity patterns in populations that have possibly experienced recurrent cycles of contraction, expansion, admixture, and adaptation, typical of long-term glacial refugia, is a challenging endeavour. Contemporary human-mediated processes, which in the case of the honey bee may involve movement of colonies within (transhumance) and between lineages (introduction of commercial queens), selective breeding, and accidental introductions of exotic pests and diseases, may further complicate this effort by erasing or obscuring the genetic signatures imprinted by evolutionary and demographic processes. Fortunately, however, Iberia is the best-studied refugial area in Europe, and common patterns are emerging from comparative phylogeography (Gómez & Lunt 2007) that are of great assistance in elucidating patterns exhibited by the Iberian honey bee. Additionally, the honey bee genome has been sequenced and a SNP panel is available for conducting genome-wide sampling of multiple populations (Whitfield *et al.* 2006; Chávez-Galarza *et al.* 2013).

The honey bee native range spans Africa, Europe and the Middle East where it evolved into 30 subspecies (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al.* 2011). This vast amount of variation has been grouped into four (western European, M; eastern European, C; African, A; Middle Eastern, O) largely parapatric evolutionary lineages (Ruttner 1988; Garnery *et al.* 1992; Whitfield *et al.* 2006; Wallberg *et al.* 2014), with contact zones identified in Italy (Franck *et al.* 2000), Turkey (Kandemir *et al.* 2006), Libya (Shaibi *et al.* 2009), and Iberia (Smith *et al.* 1991; Franck *et al.* 1998; Garnery *et al.* 1998a; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008). Among these, the Iberian contact zone formed by A and M lineages has received the greatest attention with numerous studies using a wide array of markers ranging from morphology (Cornuet & Fresnaye 1989; Arias *et al.* 2006; Miguel *et al.* 2011), allozymes (Smith & Glenn 1995; Arias *et al.* 2006), mitochondrial DNA (Smith *et al.* 1991; Garnery *et al.* 1995, 1998a; Franck *et al.* 1998; De la Rúa *et al.* 2001, 2004, 2005; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008; Pinto *et al.* 2012, 2013), microsatellites (Franck *et al.* 1998; Garnery *et al.* 1998b; De la Rúa *et al.* 2002, 2003; Miguel *et al.* 2007, 2011; Cánovas *et al.* 2011) to SNPs (Chávez-Galarza *et al.* 2013).

Differential and complex diversity patterns emerged from the numerous biparental and maternal surveys of Iberian honey bees and the underlying processes shaping genetic variation remain controversial. Arguments based on selection, demography, and contemporary human-mediated processes have been favoured by different authors. Morphology (Ruttner *et al.* 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al.* 1994; Smith & Glenn 1995; Arias *et al.* 2006) exhibited a smooth latitudinal cline extending from North Africa to France supporting a hypothesis of primary intergradation (Ruttner *et al.* 1978). In contrast, the abrupt transition from highly divergent M mitotypes in the northeastern half of Iberia to A mitotypes in the southwestern half (Garnery *et al.* 1995; Franck *et al.* 1998; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008) was more compatible with secondary contact (Smith *et al.* 1991). To complicate matters further, microsatellites did not capture the signal of a contact zone in Iberia (Franck *et al.* 1998; Garnery *et al.* 1998b; Cánovas *et al.* 2011; Miguel *et al.* 2011), but detected instead a sharp disruption between Iberian and northern African populations (Franck *et al.* 1998). This latter finding prompted a third hypothesis that explained occurrence of A mitotypes in Iberia by human-assisted introductions of African colonies during Muslim occupation, with selection acting to maintain

the morphological, allozymic and maternal clines (Franck *et al.* 1998). The hypothesis of selection as the driving force shaping the Iberian cline was recently addressed in a genome-wide SNP scan conducted in a fine-scale sample that covered the entire Iberian honey bee range (Chávez-Galarza *et al.* 2013). This study detected signatures of selection in the Iberian honey bee genome, suggesting that this evolutionary force has had an important role in structuring Iberian honey bee diversity.

Here, we built from those findings to provide, at both geographic and genomic levels, the most comprehensive characterization of the Iberian honey bee diversity patterns performed until now. We employed multiple clustering approaches and cline analysis to examine the genome-wide SNP data set using a population-genomics framework. After analysing the patterns of variation generated by the complete SNP data set, we removed any SNPs putatively associated with selection identified by Chávez-Galarza *et al.* (2013) and then used concurrently a mtDNA locus and 309 remaining neutrally tested SNPs to address the following questions: (i) How effective are SNPs in capturing clinal variation? – a signal that microsatellites have failed to detect, (ii) How concordant are the patterns generated by the complete and the neutral SNP data sets? (iii) Do neutral SNPs capture the clinal signal? (iv) Does the mtDNA marker confirm the presence of a cline formed by two highly divergent lineages, as documented by earlier studies? and How concordant are the patterns of neutral and maternal variation? If variation originated from secondary contact, then we expect neutral SNPs to detect clinal patterns and multiple coincident clines. In contrast, coincident neutral clines are not expected if variation originated via primary intergradation. Further evidence for secondary contact will come from comparisons of mtDNA and nuclear DNA. If a maternal cline formed by two highly divergent lineages is observed and this cline is paralleled by nuclear DNA variation, then there is strong support for secondary contact. However, given that Iberian honey bees are managed organisms, it is possible that human-mediated processes have obscured historical patterns. Therefore, we also asked the question: (v) To what extent do contemporary human-mediated forces influence the Iberian honey bee structure?

Methods

Sampling

Sampling in Iberia was conducted in 2010 across three north–south transects (Fig. 1). One transect extended along the Atlantic coast [Atlantic transect (AT)], one through the centre [central transect (CT)], and another along the Mediterranean coast [Mediterranean transect (MT)]. A total of 711 honey bee haploid males were collected in the three transects from 23 sites (AT=8, CT=9; MT=6) representing 237 apiaries and 711 colonies. Samples were stored in absolute ethanol at -20 °C until molecular analysis. Global positioning system (GPS) coordinates were recorded in the field for each apiary. Further details on sampling procedure can be found in Chávez-Galarza *et al.* (2013).

DNA extraction and marker genotyping

Using a phenol–chloroform–isoamyl alcohol (25:24:1) protocol (Sambrook *et al.* 1989), total DNA was extracted from the thorax of the 711 individuals, each representing a single colony. Analysis of mtDNA was based on the very popular $tRNA^{leu}$-cox2 intergenic locus. This region was amplified using the primers and PCR conditions detailed elsewhere (Garnery *et al.* 1993). After PCR amplification, the products were sequenced in both directions. Analysis of nuclear DNA was based on SNPs, which were genotyped using Illumina's BeadArray Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay (Illumina, San Diego, CA, USA) following manufacturer's protocols. Genotype calling was performed using ILLUMINA'S GENOMESTUDIO® data analysis software.

Of the 1536 SNPs included in the GoldenGate array, 383 passed the quality filtering and were polymorphic in *A. m. iberiensis*, using a cut-off criterion of > 0.98 for the most common allele. The 383 SNPs (referred hereafter as the complete SNP data set) were scanned for signatures of selection using four multiple-population F_{ST}-based methods and the spatial method matSAM (Chávez-Galarza *et al.* 2013). The 74 outlier loci detected by at least one of the five methods were removed from the complete data set, leaving 309 neutrally tested SNP loci (referred hereafter as the neutral SNP data set). Chromosomal locations in honey bee linkage groups and a summary of physical distances of the SNPs are shown in Table S1 (Supporting information). Details of SNP genotyping and detection of outlier loci can be found in Chávez-Galarza *et al.* (2013).

Mitochondrial DNA analysis

The sequences produced for the $tRNA^{leu}$-cox2 intergenic locus were aligned using MEGA version 5.03 (Tamura *et al.* 2011). For the identification of mitotypes, the sequences were compared with those available in GenBank. Each mitotype was then assigned to western European (M), eastern European (C), and African (A) lineages or to an African sublineage (A_I, A_{II}, A_{III}), according to the complex architecture of the $tRNA^{leu}$-cox2

Fig. 1 Map of the Iberian Peninsula showing the centroids of the sampling sites at each transect (AT, Atlantic; CT, central; MT, Mediterranean), sample size per site (n), site codes (CT1 to MT6), and the west–east transect (dashed line) traced from Lisbon to Girona for the geographic cline analysis (see Fig. 6). The approximate location of putative inferred refugia for the Iberian honey bee and other Iberian fauna that supports them (adapted from Gómez & Lunt 2007) is also shown.

intergenic region described elsewhere (Garnery *et al.* 1993; Franck *et al.* 1998, 2001; Rortais *et al.* 2011; Pinto *et al.* 2012). In short, the intergenic region is composed of two elements: the P (size varies between ~53 and 68 bp) and the Q (size varies between ~194 and 196 bp). A combination of point mutations and indels in the P element distinguishes honey bee subspecies from different lineages. The number of Q elements can vary between one and four, although the number of repeats is not lineage specific. Given the highly variable size of the sequences resulting predominantly from the variable number of Q elements, the sequences were trimmed at the end of the first Q with additional Qs coded as present/absent. Therefore, the ~627-bp sequence fragment analysed herein encompassed the 5′ end of the $tRNA^{leu}$ gene, the P element, the first Q element, the coding relative to the other Q elements, and the 3′ end of the cox2 gene. Relationships among the sequences were inferred using the median-joining network algorithm (Bandelt *et al.* 1999), as implemented in the program NETWORK version 4.6.1.1 (Fluxus Engineering, Clare, UK; http://www.fluxus-engineering.com). The trimmed ~627-bp sequences examined in this study did not diverge from those downloaded from GenBank. Therefore, GenBank sequences were not included in the network analysis.

Estimation of structure by nonspatial approaches

Structure was inferred from the complete and the neutral SNP data sets using two approaches: the Bayesian model-based STRUCTURE and the model-free discriminant analysis of principal components (DAPC). These approaches were used to estimate the proportion of an individual's genome (*Q*) that originated from a given genetic group or cluster.

The Bayesian clustering approach was implemented in STRUCTURE 3.4 (Pritchard *et al.* 2000) for haploid data using the admixture ancestry and correlated allele frequency models run with the unsupervised option. The program was set up for 750 000 Markov chain

Monte Carlo (MCMC) iterations after an initial burn-in of 250 000, which was sufficient to reach convergence. Over 20 independent runs for each number of clusters (K), from 1 to 7, were performed to confirm consistency across runs. The Greedy algorithm, implemented in the software CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007), was used to compute the pairwise 'symmetric similarity coefficient' between pairs of runs and to align the 20 runs for each K. The means of the permuted results were plotted using the software DISTRUCT 1.1 (Rosenberg 2004). The optimal K value was determined using Evanno's ΔK (Evanno *et al.* 2005) and Campana's ΔF_{st} methods (Campana *et al.* 2011) in STRUCTURE HARVESTER web v0.6.93 (Earl & Von Holdt 2012) and CORRSIEVE 1.6-8 package (Campana *et al.* 2011), respectively.

The DAPC clustering approach was implemented in ADEGENET 1.3-9 package for R (Jombart 2008). Simulation studies have shown that DAPC performs as well or better than STRUCTURE, particularly under more complex structuring scenarios (Jombart *et al.* 2010; Klaassen *et al.* 2012). DAPC provides a description of the genetic clustering using coefficients of the alleles (loadings) in linear combinations and seeks to maximize between-groups variance and minimize within-group variance in these loadings (Jombart *et al.* 2010). Successive K-means clustering runs (from 1 to 40) were also incorporated in the analysis to estimate the optimal number of distinct clusters (K) based on the Bayesian information criterion (BIC). The optimal K value is associated with the lowest BIC value (Jombart *et al.* 2010).

Estimation of structure by spatial approaches

Spatial structure was inferred from the complete and the neutral SNP data sets using two approaches that explicitly incorporate information on geographic coordinates for genotyped individuals: the model-free multivariate spatial principal component analysis (sPCA) and the Bayesian model-based TESS. The sPCA is a modification of PCA which takes into account not only the genetic variance of individuals or populations but also their spatial autocorrelation (measured by Moran's I). This approach disentangles global structures (clines, patches or intermediates) from local structures (strong genetic differences among neighbours), and from random noise (random distribution of allelic frequencies among individuals or populations on a connection network). While global structures display positive spatial autocorrelation (high positive eigenvalue), local structures display negative spatial autocorrelation (high negative eigenvalue) (Jombart *et al.* 2008). The sPCA was performed in ADEGENET using the K-nearest neighbours to model the spatial connectivity among individuals. To test for global and local structures, a Monte Carlo test was implemented using 10 000 permutations.

The Bayesian model-based clustering approach implemented by the software TESS (Chen *et al.* 2007) incorporates spatial population models assuming geographic continuity of allele frequencies by including the interaction parameter Ψ, which defines the intensity of two neighbouring individuals belonging to the same genetic cluster. The incorporation of trend surface and a Gaussian autoregressive residual term allows for capturing global and local patterns. The software TESS 2.3.1 was run for haploid data using the convolution admixture model (BYM), correlated allele frequency and a trend degree surface of 1. A Euclidean distance matrix was used to weight the spatial connectivity among individuals. Five runs were carried out at each K, from 2 to 7, with 5 000 000 MCMC total sweeps including a burn-in of 1 250 000 sweeps. For each run, the deviance information criterion (DIC) was calculated and the values of all runs were averaged and plotted against K. The first lower DIC value represents the optimal K for the data. As in STRUCTURE, the software programs CLUMPP and DISTRUCT were used to obtain the average matrix of membership proportions (Q) for each K and for graphical representation.

Geographic cline analysis

Geographic clines were estimated for mtDNA frequency, individual SNPs frequency, and mean Q per sampling site inferred from both the complete and the neutral SNP data sets by STRUCTURE at $K = 2$. Cline analysis of individual SNPs was performed on a subset of loci with an absolute allele frequency difference > 0.2. The rationale behind this choice was that the SNP panel used in this study was ascertained from the reference genome of *A. mellifera* (sequenced from the North American DH4 strain, which was primarily *A. m. ligustica*, a subspecies belonging to the C lineage) and genome sequence traces of Africanized honey bees (largely the African *A. m. scutellata* admixed with the genomes of *A. m. ligustica* and the M-lineage *A. m. mellifera*), and is thus underrepresented for diagnostic SNPs and SNPs with large allele frequency differences between the two maternal lineages identified in Iberian honey bees. Using a subset of loci with larger allele frequency difference between the groups, we expected to increase the ancestry information of individual loci for cline analysis. Of the 383 SNPs, 33 (17 neutral and 16 under selection, as identified by Chávez-Galarza *et al.* 2013) conformed to the frequency criterion.

Sampling sites were arranged along a transect beginning at the westernmost location (Lisbon, Portugal) and ending at the easternmost location (Girona, Spain)

(Fig. 1). Each sampling site was assigned a distance along this transect, which corresponded to the shortest straight-line distance between it and Lisbon, calculated using the 'harversine' approach (www.movable-type.co.uk/scripts/latlong.html). The cline shape was modelled using the package HZAR v2.5 (Derryberry *et al.* 2014), which fits allele frequency data to equilibrium geographic cline models (Szymura & Barton 1986, 1991; Barton & Gale 1993; Gay *et al.* 2008) using the Metropolis–Hastings Markov chain Monte Carlo algorithm. The following cline shape parameters were estimated: centre (c, distance from sampling location), width (w, 1/maximum slope), delta (δ, distance between the centre and the tail) and tau (τ, slope of the tail). The allele frequencies at the top and bottom of the cline (*Pmin* and *Pmax*) were either fixed or free to vary. Three sets of five cline models were fitted: model set 1 had no scaling (*Pmin* = 0, *Pmax* = 1), model set 2 had fixed scaling (*Pmin* = observed minimum, *Pmax* = observed maximum), and model set 3 allowed *Pmin* and *Pmax* to vary. Within each model set, scaling and tails were fixed or free to vary. These models were compared to a null model of no clinal transition using the Akaike Information Criterion corrected (AICc). The best-fitting model had the lowest AICc value. To evaluate coincidence among cline centre positions and concordance among cline widths, the composite likelihood method (Phillips *et al.* 2004) was used. Likelihood profiles were constructed for both c and w to compare alternative hypotheses across loci: H1, all loci are characterized by statistically indistinguishable c and w values and are likely to share a common c/w; and H2, each locus has its own independent c and w values. Composite log-likelihood profiles were constructed by summing log-likelihood (ML) profiles for all individual SNP loci ML(H1). This composite log-likelihood profile was compared to the sum of all maximum-likelihood estimates for individual SNP loci ML(H2) using a likelihood ratio test (LRT). If the clines of individual SNP loci coincide and have the same c/w values, ML(H1) is not significantly different from ML (H2) (ML = ML(H2) – ML(H1) $\approx$ 0). Conversely, if the clines do not coincide, ML(H1) is significantly smaller than ML(H2) (ML > 0). The significance of any difference of ML(H1) and ML(H2) was determined using a chi-square test with $n-1$ degrees of freedom ($\alpha = 0.05$). This approach was similarly employed to evaluate coincidence and concordance of mtDNA and both SNP data sets.

Linkage disequilibrium and genetic diversity

Linkage disequilibrium (LD) between all pairs of neutral SNPs was estimated using the statistic r^2 (Hill & Robertson 1968), as implemented by the software DNASP 5.10.1 (Librado & Rozas 2009). Significant LD was identified at the 5% level using Fisher's exact test. Unbiased haploid genetic diversity (*uh*) for the neutral SNPs was calculated using the program GENALEX 6.5 (Peakall & Smouse 2012). Values of LD and *uh* were calculated for each sampling site and then projected along the Lisbon–Girona transect (Fig. 1).

Statistical tests

Differences in individuals' Qs between clustering approaches were assessed using the Mann–Whitney–Wilcoxon test (Wilcoxon 1945; Mann & Whitney 1947). Genetic structure inferred by the different clustering approaches was compared using Pearson's correlation coefficient (r). Whenever applicable, statistical significance levels were adjusted for multiple comparisons using the Bonferroni procedure to correct for type I error (Weir 1996). These analyses were implemented in R (R Development Core Team 2013).

Results

Structure estimated by nonspatial approaches

Genetic structure inferred from the 711 Iberian honey bee individuals and the 309 neutral SNP data set using the Bayesian model-based clustering algorithm, implemented in STRUCTURE, and the model-free DAPC clustering algorithm, implemented in ADEGENET, is shown in Fig. 2a (for $K = 2$) and Fig. S1 (Supporting information) ($K = 3$ to 5). The optimal number of clusters (K) varied between two, when estimated by ΔF_{st} and BIC, and four, when estimated by ΔK (Fig. S2, Supporting information). Incongruent optimal K values are often obtained by different methods (Campana *et al.* 2011), especially in the presence of low levels of population differentiation (Waples & Gaggiotti 2006), which is the case of the Iberian honey bee (global $\Phi_{PT} = 0.020$; pairwise Φ_{PT} values ranged from 0.000 to 0.046, but see Table S2 and Fig. S3, Supporting information for pairwise comparisons across the study area). Given that two of three measures agreed on an optimal $K = 2$ and the presence of two maternal lineages in Iberia (this and previous studies), it is likely that the number of clusters that best represents the maximum population structure is two.

At the optimal $K = 2$ (Fig. 2a, Fig. S4, Supporting information), a concordant geographic pattern was produced by DAPC and STRUCTURE ($r = 0.79$, P-value = 0.0000 for individual Q values), although a deeper subdivision was inferred by the former than by the latter clustering approach, as measured by Q (P-value = 0.0000 for comparisons of individual Q values, Mann–Whitney–Wilcoxon test). Membership proportions

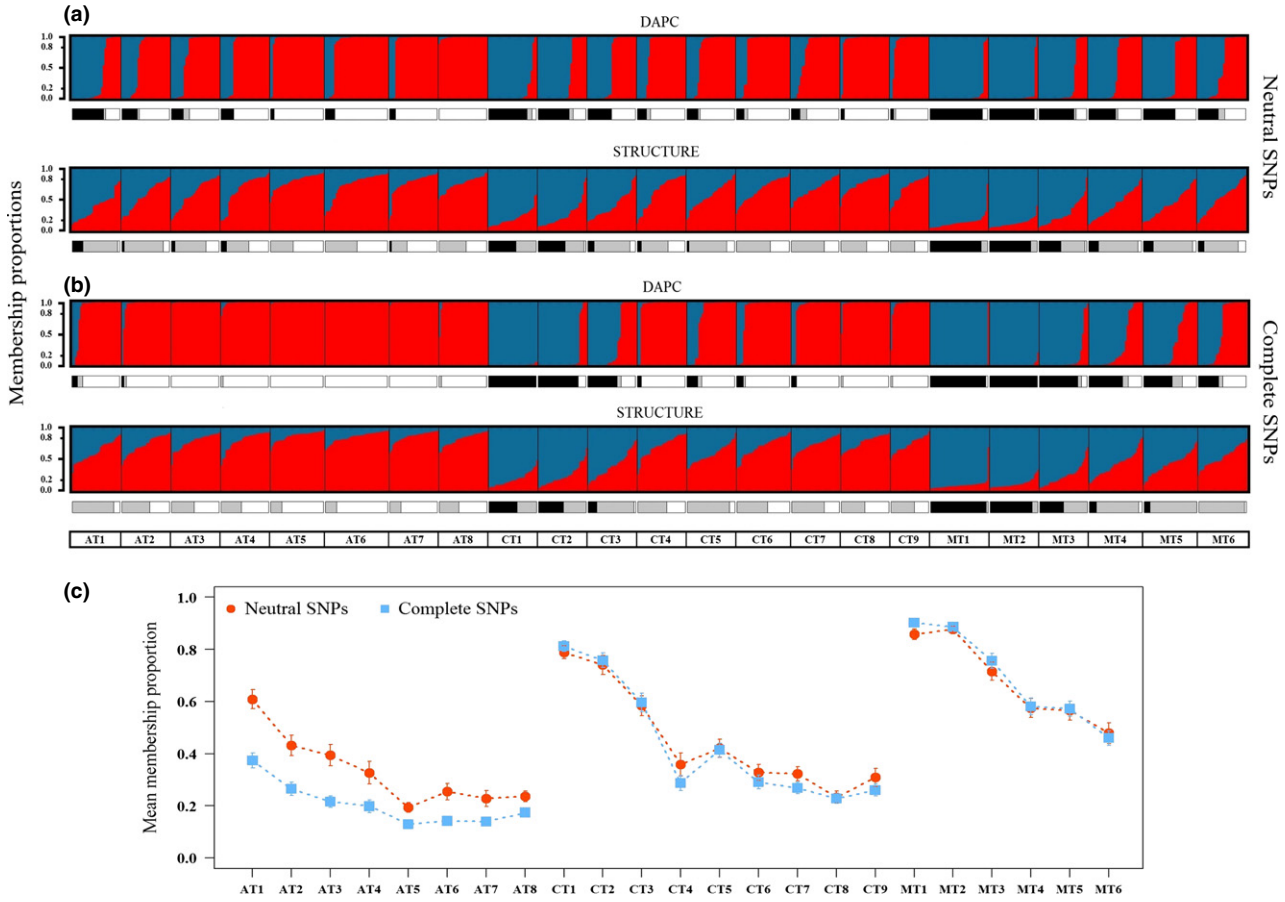

Fig. 2 Population structure of *Apis mellifera iberiensis* estimated by DAPC and STRUCTURE at $K = 2$ clusters. The 23 sampling sites are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT, Atlantic; CT, central; MT, Mediterranean). Plots represent each of the 711 individuals by a vertical bar partitioned into two coloured segments (blue and red) corresponding to membership proportions (Q) in each of the two clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from high Q (left) to low Q (right) in the blue cluster. The frequency of individuals per sampling site exhibiting $Q \geq 0.80$ (black), $0.20 > Q > 0.80$ (grey), and $Q \leq 0.20$ (white) in the blue cluster is indicated below each plot. Structure estimated from (a) the neutral SNP data set (309 loci) and (b) the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). (c) Mean membership proportion (±SE) in the blue cluster inferred from the neutral and complete SNP data sets with STRUCTURE for each sampling site.

estimated by STRUCTURE showed that most individuals (ranging from 53.3% in CT2 to 88.9% in MT1; see the bar below each clustering plot in Fig. 2a) from sampling sites near the Pyrenees were assigned with high posterior probability ($Q \geq 0.80$) to the blue cluster. The percentages increased considerably when Q was inferred by DAPC, ranging from 50.0% in CT3 to 93.3% in MT2. Individuals with $Q \geq 0.80$ in the red cluster were common in the southern sampling sites of the Atlantic transect (ranging from 60.6% in AT5 to 63.3% in AT7) and rare in the Mediterranean transect (ranging from 0% in MT1-2 to 13.3% in MT6). However, again, the percentages increased considerably when Q was inferred by DAPC. Individuals exhibiting admixed proportions (Q to any cluster ≤ 0.80) prevailed in the northern part of the Atlantic transect and in the southern part of the central and Mediterranean transects when inferred by STRUCTURE, although they were rare when inferred by DAPC (Fig. 2a).

When genetic structure was inferred from the complete SNP data set (Fig. 2b), a deeper phylogeographical signal was captured by both clustering approaches, as measured by Q (P-value = 0.0000 for comparisons of individual Q values inferred from the complete and neutral data sets with STRUCTURE and DAPC; Mann–Whitney–Wilcoxon test). Nonetheless, inclusion of the 74 putatively selected SNPs in the data set did not qualitatively change the overall geographic patterns of hybridization across Iberia, while providing additional support for an optimal $K = 2$, this time simultaneously estimated by the three methods ΔK, ΔF_{ST} and BIC (Fig. S2, Supporting information).

The mean *Q* estimated with STRUCTURE from both the neutral and complete SNP data sets is shown at the sampling site level in Fig. 2c (see Fig. S5, Supporting information for the corresponding DAPC plot). While further confirming the nearly concordant patterns across Iberia, this representation revealed a more abrupt transition from the blue to the red cluster in the central and Mediterranean transects than in the Atlantic transect, which suggests a contact zone located towards the northeastern part of Iberia.

Comparing maternal pattern with neutral structure

A median-joining network of a ~627-bp fragment of the tRNA^{leu}-cox2 intergenic mitochondrial region confirms the presence of the two highly divergent African (A) and western European (M) lineages in Iberia (Fig. 3a). The two lineages show a highly structured geographic pattern of distribution in Iberia. Mitotypes belonging to the M lineage were predominant in the northeastern half, whereas mitotypes belonging to the A lineage were fixed or almost fixed in the southwestern half of Iberia. While some sampling sites displayed a mixture of M and A mitotypes, the geographic distribution of the maternal lineages reveals a sharp northeastern–southwestern trend.

Membership proportions inferred by STRUCTURE (Fig. 3b, c) and DAPC (Fig. S6, Supporting information) from neutral SNPs were contrasted with mtDNA data, at both sampling site and individual levels. At the

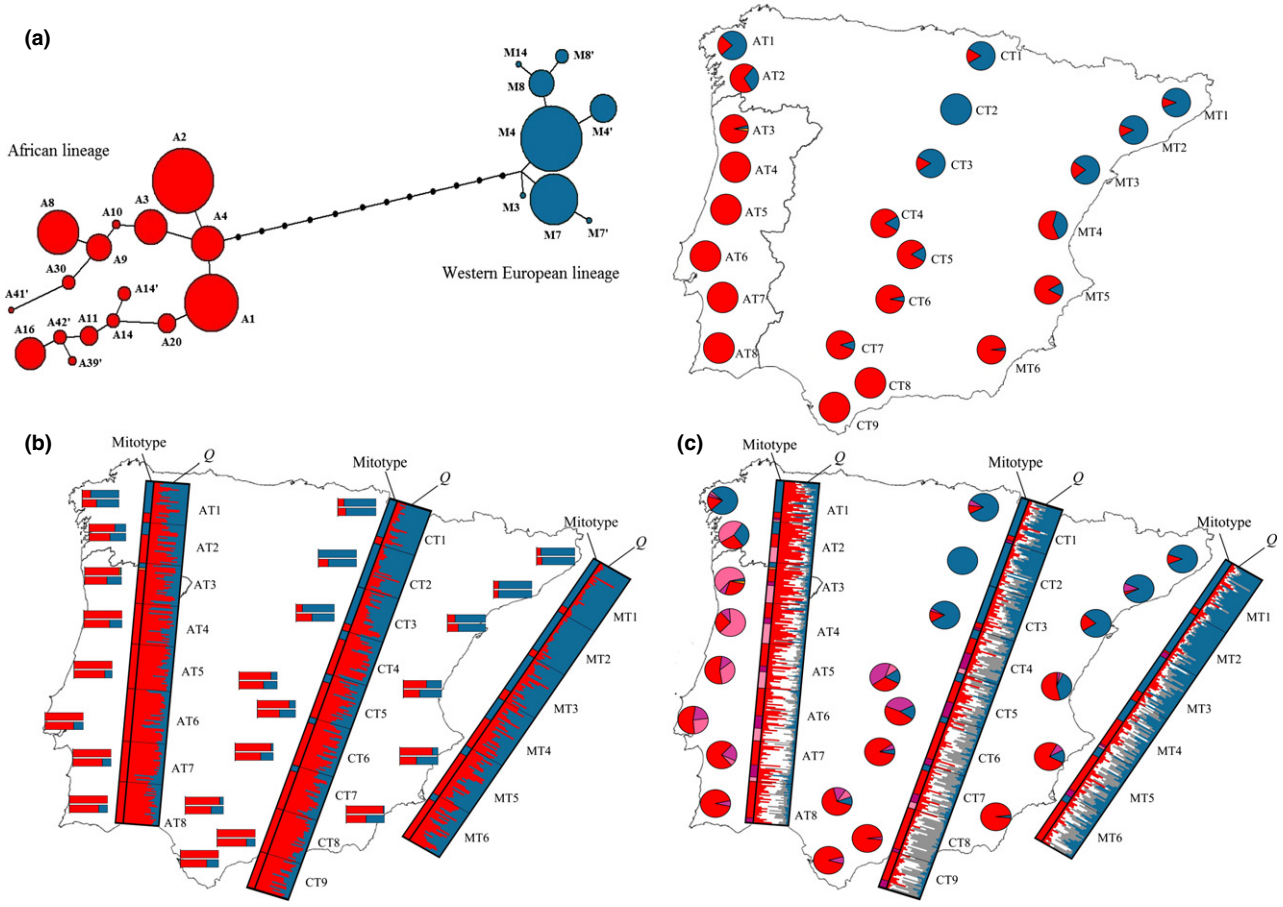

Fig. 3 Maternal pattern and estimated structure inferred from the neutral SNP data set (309 loci) with STRUCTURE. Patterns of variation are displayed at both individual and sampling site level for each transect (AT, Atlantic; CT, central; MT, Mediterranean). (a) Median-joining network relating the mtDNA sequences of a ~627-bp fragment of the tRNA^{leu}-cox2 intergenic region. The African (A) and western European (M) mitotypes form two divergent lineages. The sizes of the circles size are proportional to the mitotype frequencies. The pie charts displayed in the Iberian map at the right side show the frequencies of the A and M mitotypes at each sampling site. (b) Patterns shown at $K = 2$ clusters. Vertical plots display the mitotype (A in red; M in blue; C in orange, one single individual in AT3) and the membership proportions (*Q*) for each of the 711 individuals. Horizontal bar plots show mitotype frequencies (top) and the mean *Q* in blue and red clusters (bottom) at each sampling site. (c) Patterns shown at $K = 4$ clusters. Vertical plots display the mitotype and *Q* for each of the 711 individuals. Maternal data are represented by M lineage (blue), C lineage (orange) and A sub-lineages (A_I in red, A_{II} in magenta, and A_{III} in pink). Pie charts show mitotype frequencies at each sampling site.

sampling site level, the partitioning of neutral SNP variation into two clusters corresponded remarkably to M and A maternal lineages ($r = 0.81$ for both DAPC and STRUCTURE *vs* mtDNA; P-value < 0.0000). At the individual level, the correlations were weaker ($r = 0.46$ for DAPC, $r = 0.60$ for STRUCTURE), yet significant (P-value < 0.0000), suggesting differential gene flow among genomic compartments.

Genome partitioning of individuals at greater values of K produced increasingly complex patterns (Fig. 3c and Fig. S1, Supporting information). At $K = 4$, a pronounced east–west structuring of neutral variation was revealed. The Atlantic populations were clearly distinct from the other populations, and a north–south trend becomes more apparent in this transect. The nuclear pattern is consistent with maternal variation partitioned into African sublineages (Fig. 3c). Sublineage A_{III} mitotypes were common in northern Atlantic populations and were gradually replaced by sublineage A_I mitotypes towards the south. In contrast to Atlantic populations, sublineage A_{III} mitotypes were virtually absent in populations of central and Mediterranean transects, which were dominated by sublineage A_I mitotypes.

Structure estimated by spatial approaches

A number of studies have questioned the use of STRUCTURE for studying populations exhibiting continuous spatial distribution of genetic diversity (Serre & Pääbo 2004; Rosenberg *et al*. 2005). To address this issue, patterns of variation in Iberia were further investigated using spatially explicit approaches implemented by sPCA (Fig. 4) and TESS (Fig. 5).

Analysis of neutral SNPs using sPCA showed that one global axis and one local axis were retained, indicating the existence of both global and local spatial structures in Iberia. The interpolation of the first global score, which was associated with a strong autocorrelation (Moran's $I = 0.639$), detected two clusters forming a cline (Fig. 4a) concordant with nonspatial approaches. The second global score (Moran's $I = 0.560$) clearly differentiated the four northernmost sampling sites of the Atlantic transect and the southern half of central and Mediterranean transects (Fig. 4b). The third global score (Moran's $I = 0.443$) further partitioned the Atlantic populations into two groups (north and south) and the southern half of central transect (Fig. 4c). The northern half of the central transect was differentiated by the

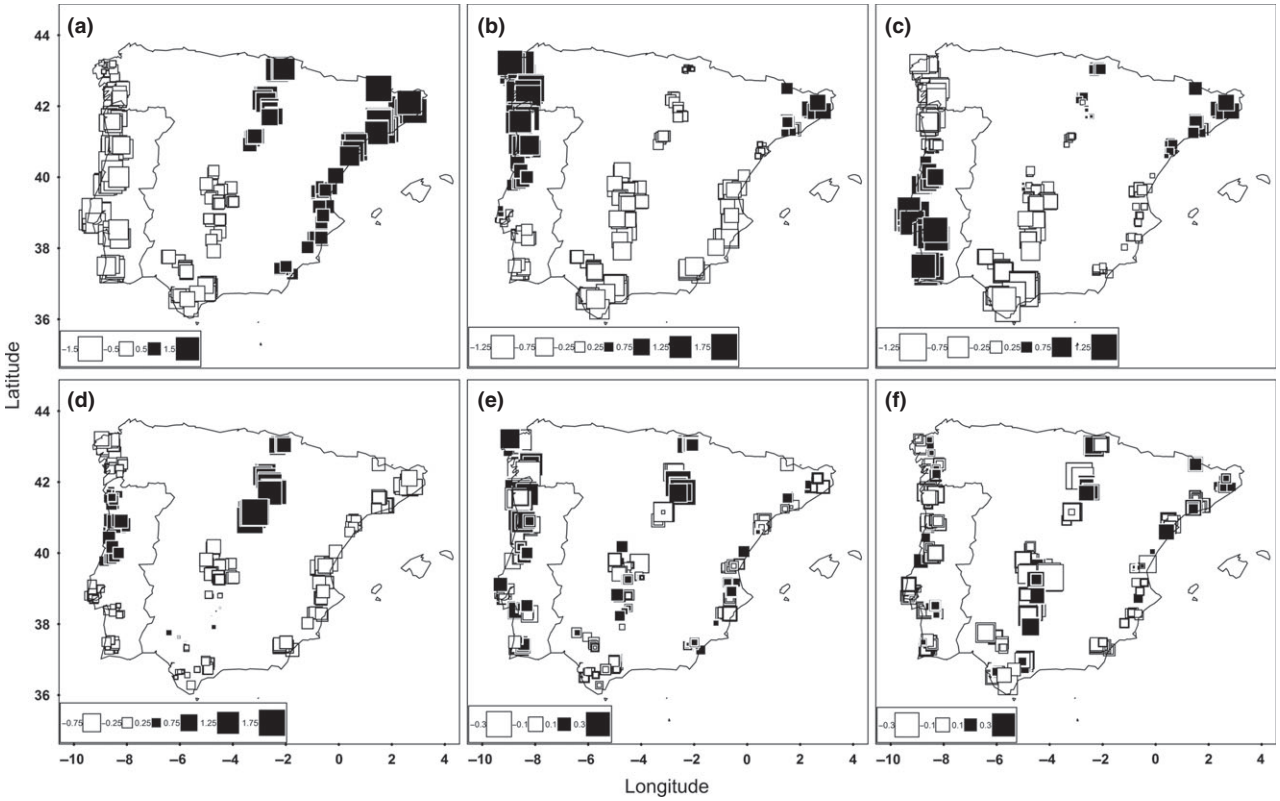

Fig. 4 Analysis of global and local structures, among 711 individuals of *A. m. iberiensis* from 23 sampling sites, by spatial principal component analysis (sPCA) using 309 neutral SNPs. Each square represents the score of an individual, which is positioned by its spatial coordinates. (a–d) The first four global scores of sPCA. (e–f) The first two local scores of sPCA.

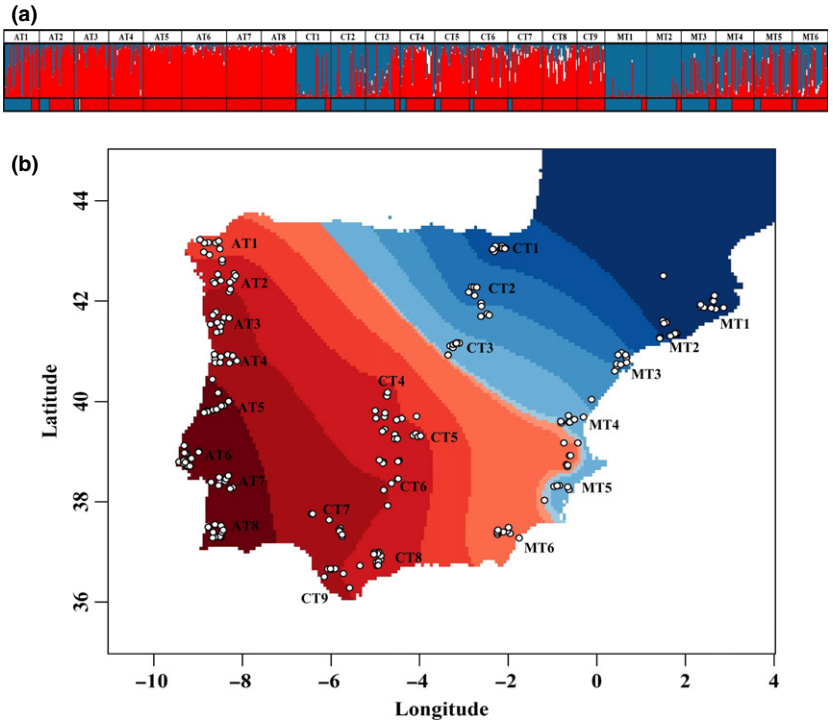

Fig. 5 Spatially explicit analysis implemented by the software TESS for the 711 individuals of *Apis mellifera iberiensis* using the neutral SNP data set (309 loci). (a) Plot of individuals' *Q* at the optimal *K* = 3 clusters. Each of the 711 individuals included in the analysis is represented by a vertical bar partitioned into three coloured segments (blue, red and white) corresponding to *Q* in each of the three clusters. Maternal data (M lineage in blue, C lineage in orange, and A lineage in red) are shown at the bottom. Sampling sites and individuals within sampling sites are arranged as in Fig. 3. (b) Map of the Iberian Peninsula showing the two major clusters ($Q \geq 0.5$) interpolated by TESS. Dots represent the locations of sampled apiaries across the Atlantic (AT1-8), central (CT1-9) and Mediterranean (MT1-6) transects.

fourth global score producing a Moran's $I = 0.392$ (Fig. 4d).

While the global test corroborated the presence of global spatial structure (max(t) = 0.0017; *P*-value = 0.0001), there was also structure at the local level (max (t) = 0.0019; *P*-value = 0.0001). The first local score (Moran's $I = -0.075$) highlighted the differences among individuals of northern Atlantic and central transects (Fig. 4e), while the second local score (Moran's $I = -0.071$) differentiated the individuals from sampling sites in the middle part of the central transect (Fig. 4f).

The additional spatial approach performed using TESS further confirmed the neutral patterns obtained previously (Fig. 5). Two major clusters that largely overlapped those of nonspatial approaches ($r = 0.76$ for STRUCTURE *vs* TESS and $r = 0.64$ for DAPC *vs* TESS using individual *Q* values, *P*-value < 0.0000) were identified by TESS at each simulated *K* (Fig. S7, Supporting information). At the optimal *K* = 3 (Fig. S8, Supporting information), one additional minor cluster (mean *Q* = 0.023 in the white cluster) further partitioned the nuclear genomes of individuals mainly from the southern portion of the central transect (Fig. 5a). While TESS supported the major northeastern–southwestern cline and the contrasting patterns exhibited by Atlantic and Mediterranean populations, it did not capture the partitioning within the Atlantic transect, which was detected by the other clustering approaches and by mtDNA analysis.

The spatial patterns inferred from the complete SNP data set using both sPCA and TESS were largely concordant with those inferred from the neutral SNP data set, although, as observed with the nonspatial clustering approaches, a deeper phylogeographical signal was captured by the complete SNP data set (Figs S9 and S10, Supporting information).

Geographic cline analysis

The geographic clines were modelled for 33 (17 neutral and 16 selected) individual SNPs, mean *Q* obtained with the complete and neutral SNP data sets, and mtDNA (Fig. 6, Fig. S11, Supporting information). There was considerable variation in the identity of the best-fitting model among individual SNPs, SNP data sets and mtDNA (Table S3, Supporting information).

The model 'Pmin/Pmax observed – no tails' was fitted to 16 of the 33 SNPs and to the neutral SNP data set, whereas 'Pmin/Pmax fixed – right tail' and 'Pmin/Pmax fixed – no tails' were fitted to the mtDNA and the complete SNP data set, respectively. The AICc values obtained for the null model of no clinal variation were higher for mtDNA (cline model = 255.1, null model = 516.3), complete SNP data set (cline model = 25.2, null model = 203.5) and neutral SNP data set (cline model = 33.2, null model = 137.2) than for any individual SNP (Table S3, Supporting information).

Estimates of cline centre positions and widths for the 33 individual SNPs, complete SNP data set, neutral SNP data set and mtDNA were highly variable (Table S3, Supporting information). Coincidence analysis of the 33 SNPs revealed that 18 (Table S3, Supporting information), of which nine were neutral, could be constrained to share a common centre ($LRT_{same\text{-}diff.}$ = 26.88, 17 d.f., *P*-value > 0.05) at the consensus position of 665.2 km, as estimated by the likelihood profiles. The consensus centre of the 18 SNPs was coincident with those estimated for mtDNA (706.7 km) and for the complete (714.7 km) and neutral (725.7 km) SNP data sets ($LRT_{same\text{-}diff.}$ = 0–3.03, 1 d.f., *P*-value > 0.05 for all pairwise comparisons). Concordance analysis of the 33 SNPs revealed that 23 (Table S3, Supporting information), of which 11 were neutral, exhibited a similar width ($LRT_{same\text{-}diff.}$ = 1.60, 22 d.f., *P*-value > 0.05) of 1350 km. The consensus width of the 23 SNPs was concordant ($LRT_{same\text{-}diff.}$ = 0.3 – 3.18, 1 d.f., *P*-value > 0.05 for all pairwise comparisons) with those estimated for the complete (1283.5 km) and neutral SNP data sets (1047.6 km), but not with that estimated for the mtDNA (580.9 km), which was significantly narrower ($LRT_{same\text{-}diff.}$ = 10.10–26.56, 1 d.f., *P*-value < 0.05 for all pairwise comparisons).

Linkage disequilibrium and genetic diversity

Genome-wide analysis of linkage disequilibrium (LD) between all possible pairs of neutral SNPs in each sampling site produced low levels of LD with mean r^2 values varying between 0.014 and 0.045 (Table S4, Supporting information). From a total of 701,362 pairwise comparisons, 14 687 pairs (2.09%), ranging from 1.49% to 3.41%, exhibited significant LD before Bonferroni correction (a single pair in CT4 remained significant after Bonferroni correction). Pairwise comparisons performed by linkage group also produced low LD values (data not shown). Levels of unbiased haploid diversity (*uh*) were low, ranging from 0.281 in the Atlantic transect (AT8) to 0.313 in the Mediterranean transect (M6, Table S4, Supporting information). Interestingly, despite the low levels of LD and *uh* across the study area, a trend of elevated values was observed towards the centre of the cline and overlapping the consensus centre location (Fig. 7).

Discussion

Genetic studies of Iberian honey bees have revealed complex and often incongruent patterns of variation, which have led to competing hypothesis of primary intergradation (Ruttner *et al.* 1978) and secondary contact (Smith *et al.* 1991) as the leading mechanisms shaping patterns of variation. We examined a large number of individuals with a maternal locus and genome-wide SNPs and provided the most comprehensive portrait of clinal change across the entire Iberian honey bee distributional range. Our results support a signature of origin via secondary contact, which was still detectable despite intense beekeeping practices involving selective breeding and large-scale movement of colonies.

Nuclear and maternal patterns and cline origin

The multiple clustering approaches and the geographic cline analysis implemented on genome-wide SNPs collectively revealed a well-defined clinal pattern bisecting Iberia along a northeastern–southwestern axis, contrasting with the lack of microsatellite structure documented earlier (Franck *et al.* 1998; Cánovas *et al.* 2011; Miguel *et al.* 2011).

The most commonly suggested mechanisms underlying clinal patterns in gene frequencies are random genetic drift with isolation-by-distance effects, selection across an environmental gradient (primary intergradation), and secondary contact between previously isolated and genetically divergent populations. The SNP patterns exhibited by the Iberian honey bees could be explained by any of these mechanisms. However, several aspects of our data are more consistent with an alternative model of secondary contact and introgression between divergent populations previously isolated in glacial refugia, as proposed for a growing list of other Iberian taxa (reviewed by Gómez & Lunt 2007; Godinho *et al.* 2008; Gonçalves *et al.* 2009; Pinho *et al.* 2009; Miraldo *et al.* 2011; Carneiro *et al.* 2013; among others). First, while a deeper structure was retrieved by the complete SNP data set, excluding the 74 SNPs with signatures of selection (Chávez-Galarza *et al.* 2013) did not qualitatively change the clinal pattern of variation. A similar historical signal emerged when selected loci were removed from the genome-wide SNP data set.

Second, the geographic cline analysis revealed that a large proportion (9 of 17) of the neutral individual SNPs and both SNP data sets share a common cline

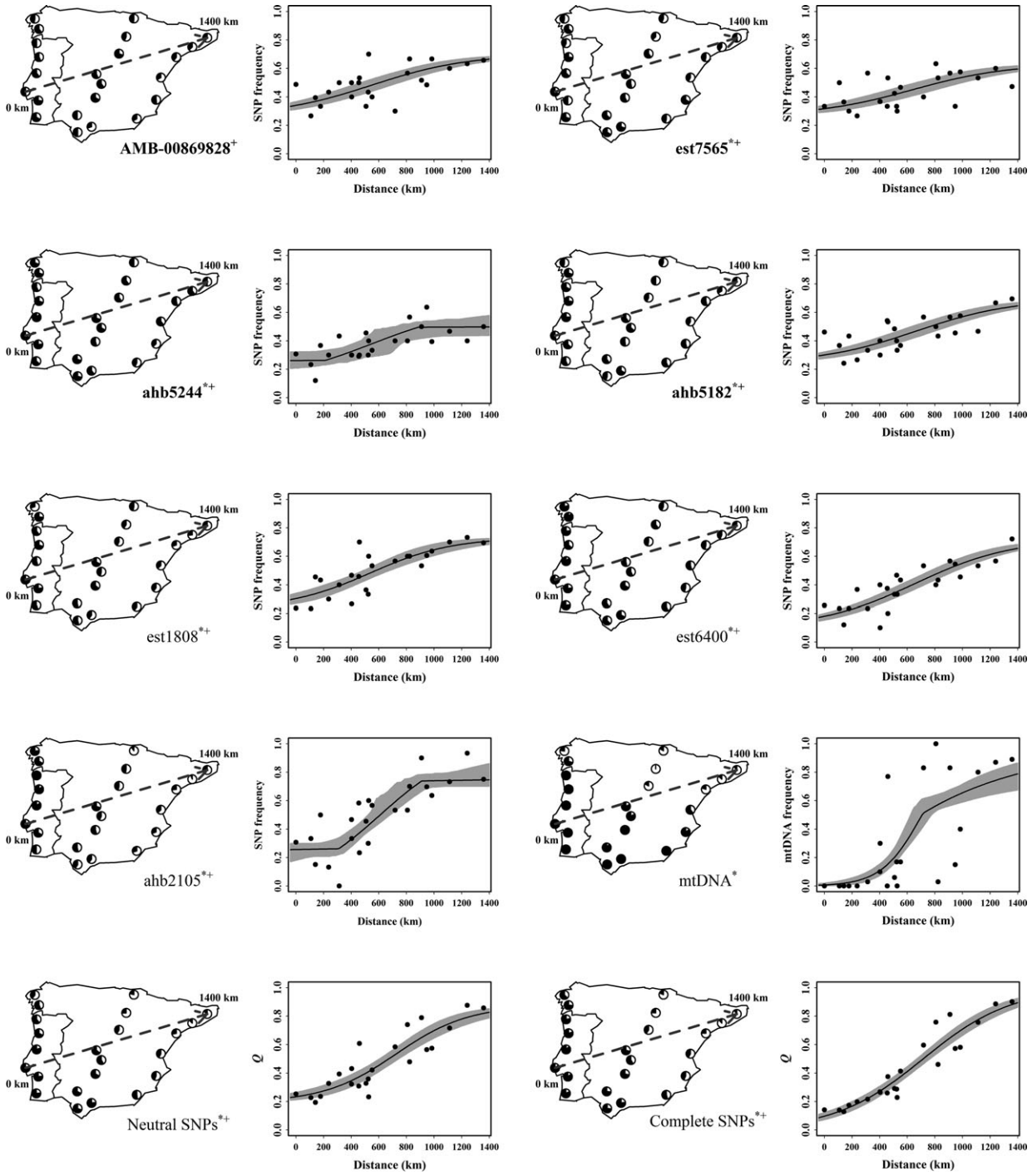

Fig. 6 Map of the Iberian Peninsula with pie charts summarizing frequency data for each sampling site and plot of maximum-likelihood geographic cline for four neutral SNP loci (marked in bold), three selected SNP loci (see Fig. S11, Supporting information for the remaining 26 SNPs), mtDNA, and the *Q* values estimated with STRUCTURE from the neutral and the complete SNP data sets. The symbols * and + indicate the loci or data sets with coincident centre and concordant width, respectively (see Table S3, Supporting information). The dashed line placed in each map represents the transect traced from Lisbon (0 km) to Girona (1400 km) for the geographic cline analysis.

centre, indicating considerable genome-wide coincidence. Existence of multiple coincident clines argues for secondary contact (Barton & Hewitt 1981), especially if some of the clines reflect changes in selectively neutral loci (Durrett *et al.* 2000). Simulations on the origin of contact zones show that a signature of secondary

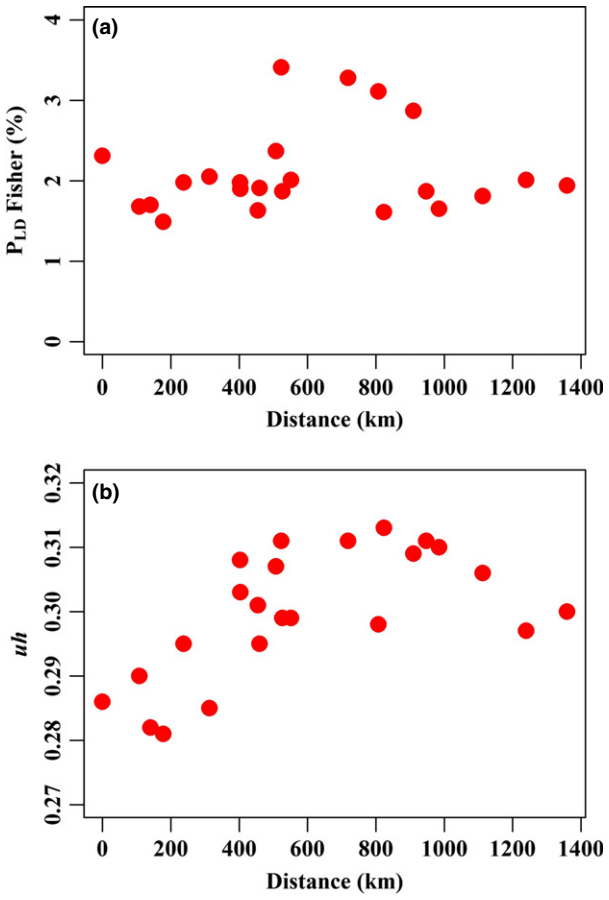

Fig. 7 (a) Percentage of pairs of neutral SNP loci showing significant linkage disequilibrium (LD) with Fisher's exact test (P_{LD} Fisher), before Bonferroni correction, and (b) unbiased haploid genetic diversity (*uh*) estimated from the neutral SNPs, both projected along the geographic cline.

contact, which is characterized by clinal variation at neutral loci and extensive disequilibrium at the centre of the contact zone, can persist for thousands of generations if neutral loci were tightly linked to selected loci (Durrett *et al.* 2000). If neutral loci were not tightly linked to selected loci, the initially steep clines, formed at the moment the two divergent groups meet, will gradually widen as the intermixing proceeds. In contrast, in the presence of primary intergradation, neutral loci will not vary clinally and disequilibrium between a neutral locus and a closely linked locus under selection will decay quickly (Durrett *et al.* 2000).

Further support for secondary contact is provided by LD and diversity patterns. Both parameters show a trend of elevated values towards the centre of the cline and overlapping the consensus centre location, as expected when two divergent populations meet. It should be noted, however, that the LD levels at the centre of the contact zone were unexpectedly low for recent contact, which may simply reflect the sparse distribution of the SNPs and the short scale of LD in honey bees. Indeed, the exceptionally high recombination rate in honey bees (Beye *et al.* 2006) would lead to a rapid decay of LD after admixture, as observed in the Africanization process in the New World (Pinto *et al.* 2005).

Finally, while the observed patterns can also be explained by alternative processes involving isolation by distance with some migration or by divergence in parapatry, the ultimate support for secondary contact comes from mtDNA. Congruent with previous surveys (Smith *et al.* 1991; Garnery *et al.* 1992, 1995; Franck *et al.* 1998; Cánovas *et al.* 2008; Miguel *et al.* 2007), two highly divergent mtDNA lineages (African and western European) were identified and these lineages form a cline that closely parallels that of genome-wide SNPs, as revealed by both the clustering and the cline analyses. The mtDNA cline centre coincided with that of both the complete and the neutral SNP data sets and with most individual SNPs (18 of 33). In contrast, the mtDNA cline width was not concordant with any of the SNPs (either individual or data sets), which largely formed wider clines. Narrower maternal clines could arise from stronger drift on the smaller N_e of the haploid marker (Polechova & Barton 2011), from divergent selection on mtDNA types (Yuri *et al.* 2009), or from the greater gene flow expected of nuclear loci (Endler 1977). Our results of narrower maternal clines compared to nuclear clines add to a body of research showing discordant cytonuclear transmission across contact zones (Gowen *et al.* 2014 and many references therein).

Whether secondary contact resulted from ancient range expansions from North Africa following climate amelioration of the last postglacial period (De la Rúa *et al.* 2002; Pinto *et al.* 2013) or from recent introductions of the North African subspecies *A. m. intermissa* by the Arabs during Muslim occupation (Franck *et al.* 1998) is a matter of debate. A STRUCTURE analysis found no signs of *A. m. intermissa* genes, belonging to the African lineage, in Iberian populations, excepting for a residual component detected in sampling sites CT8 and CT9 nearby the Strait of Gibraltar (see Fig. S3, Supporting information in Chávez-Galarza *et al.* 2013). This observation together with a report of deep differentiation between *A. m. iberiensis* and *A. m. intermissa* SNPs (Whitfield *et al.* 2006) means that a recent colonization event would have to be accompanied by a complete replacement of the nuclear, but not the mitochondrial, genomes of colonizers. This hypothesis assumes long-term male-biased gene flow, which would erode a signal of subdivision at the nuclear but not at the maternal level. The problem is that the latter scenario is not consistent with honey bee reproductive biology, as females have an important role in long-distance dispersal (Winston 1987). A much earlier event is

therefore more likely to have been responsible for the patterns we see today.

The complexities and incongruences of Iberian honey bee patterns revealed by distinct genetic markers suggest an ancient history of allopatric divergence in Iberian refugia followed by postglacial range expansions and secondary contact. Iberia served as an important refuge during the cold periods of the Pleistocene in Europe (reviewed by Weiss & Ferrand 2007). During this epoch, repeated cycles of contraction into and expansion out of multiple refugia shaped diversity patterns of great complexity in a variety of Iberian animal taxa (see Miraldo *et al.* 2011 and references therein), among which the honey bee is seemingly no exception. Estimates of genetic divergence from mtDNA (Arias & Sheppard 1996) and whole-genome nuclear DNA (Wallberg *et al.* 2014) suggest that the split among the four honey bee evolutionary lineages occurred between 670 000 and 300 000 years ago, respectively. Accordingly, colonization of Iberia across the Strait of Gibraltar (Ruttner *et al.* 1978; Whitfield *et al.* 2006; Han *et al.* 2012), from an origin in either Africa (Whitfield *et al.* 2006) or western Asia (Wallberg *et al.* 2014), likely occurred during Middle Pleistocene. Given dispersal abilities of the honey bees, it is plausible that they dispersed across the Iberian territory during interglacial periods and retreated to refugia during the glacial periods. Evidence from comparative phylogeography suggests that multiple refugia existed in Iberia ('refugia within refugia' paradigm of Gómez & Lunt 2007). Two such refugia, one in the Mediterranean coast of northeastern Spain, possibly close to the Ebro valley, and another in the Betic ranges of southern Spain, were inferred from overlapping subdivision patterns exhibited by several Iberian taxa (Gómez & Lunt 2007; see Fig. 1). The blue and the red clusters identified in Fig. 3 are consistent with the existence of these two putative refugia. Although the presence of multiple honey bee refugia is a tentative result, it is an idea that can be further explored and tested using the power of multiple gene genealogies analysis.

Influence of human-mediated processes in shaping variation

Complicating the interpretation of diversity patterns in honey bees are contemporary human-mediated processes. Honey bees native to Europe have long been subjected to human manipulation (Crane 1999), with a variable impact in their genetic composition (reviewed by De la Rúa *et al.* 2009). In western Europe north of the Pyrenees, human-mediated movements of colonies between lineages (introduction of commercial queens) promoted variable levels of C-lineage introgression (Jensen *et al.* 2005; De la Rúa *et al.* 2009; Soland-Reckeweg *et al.* 2009; Oleksa *et al.* 2011; Pinto *et al.* 2014) and even replacement of the native *A. m. mellifera* subspecies in some areas (Jensen *et al.* 2005). In contrast, the Iberian honey bee is relatively free of C-lineage genes (Miguel *et al.* 2007, 2011; Cánovas *et al.* 2008, 2011; Pinto *et al.* 2013). A single colony harbouring a C-derived mitotype was scored in this study, and no signs of introgression were detected at the nuclear level in that colony and in the remaining 710 (but see Figs S3 and S4, Supporting information in Chávez-Galarza *et al.* 2013).

While movements of colonies between lineages have not yet seriously threatened the Iberian honey bee genetic integrity, the lack of microsatellite structure (Franck *et al.* 1998; Cánovas *et al.* 2011; Miguel *et al.* 2011) has been interpreted as an indication of high levels of gene flow aided by within-lineage movements associated with transhumance (Cánovas *et al.* 2011). This interpretation, however, is inconsistent with our results that show congruent cytonuclear subdivision, local structure detected by the sPCA, and relatively low levels of LD, none of which support the large-scale influence of transhumance in the Iberian honey bee gene pool. A possible explanation for microsatellite patterns is that saturation of the mutation spectrum homogenized allele size distributions (Nauta & Weissing 1996). Such homogenization has been suggested before to explain a similar pattern in the European rabbit (Queney *et al.* 2001). Almost identical to what is observed for the Iberian honey bee, the European rabbit exhibits a northeastern–southwestern cline for mtDNA (Branco *et al.* 2000), allozymes (Campos *et al.* 2007; Ferrand & Branco 2007), and nuclear sequence data (Branco *et al.* 2002; Geraldes *et al.* 2008; Carneiro *et al.* 2013), but no clinal pattern for microsatellites (Queney *et al.* 2001).

The cytonuclear structure in Iberian honey bees is noteworthy given that in Spain, over one million colonies, representing ~50% of existing colonies, have been yearly involved in wide-range movements, in the last decades (A. G. Pajuelo, personal communication). The fact that a marked clinal pattern in both the nuclear and mitochondrial genomes still persists indicates that human-mediated movements play a minor role in shaping Iberian honey bee genetic structure. Nonmutually exclusive explanations can be accounted for the observed pattern. Either transhumance takes place after the reproductive season, or some kind of reproductive barrier or local adaptation is preventing gene flow and long-term establishment of translocated colonies.

Concluding remarks

In this study, a well-defined northeastern–southwestern clinal pattern, revealed simultaneously by nuclear and

maternal markers, provided support for the hypothesis of secondary contact proposed by earlier mtDNA studies. This finding, together with putative signatures of selection detected in a previous study (Chávez-Galarza *et al.* 2013), suggests a complex interplay between adaptation and demography in shaping the Iberian honey bee patterns that we see today. Contemporary human-mediated processes do not seem to be dramatically changing these patterns, a scenario that might change if Spanish and Portuguese beekeepers adopt a strategy of using commercial C-lineage strains, as is occurring in several countries of western Europe (reviewed by De la Rúa *et al.* 2009; Pinto *et al.* 2014). Iberian honey bees are providers of important environmental services through pollination and are number one honey producers in the European Union (European Commission 2013). More importantly, Iberian honey bees represent an important reservoir of diversity that not long ago colonized a broad territory in western Europe (Franck *et al.* 1998; Garnery *et al.* 1998a; Miguel *et al.* 2007). Understanding patterns and underlying processes shaping Iberian honey bee's diversity is an important first step towards preserving this subspecies and thereby the species *Apis mellifera*, an effort of unquestionable value as we face a worldwide honey bee crisis.

Acknowledgements

We are deeply indebted to numerous beekeepers and technicians of beekeepers' associations and of government institutions for inestimable assistance with the honey bee collection. Special appreciation is due to Antonio G. Pajuelo for providing numerous contacts of beekeepers in Spain and answering all the questions related with beekeeping in Spain; to Irene Muñoz, Andreia Brandão, Inês Moura and Margarida Neto for helping in the field work; to Pilar de la Rúa and Irene Muñoz for helping contacting beekeepers in Spain; to Danny Weaver for permission to use the OPA; to Colette Abbey for DNA extractions and running GoldenGate assays; to Clare Gill for assistance with the GenomeStudio software, and Penny Riggs for providing the facilities at Texas A&M University for SNP genotyping; to João C. Azevedo, Miguel Vaz-Pinto and João Paulo Castro for GIS and GPS assistance; to Helena M. Ferreira for searching genomic information of the SNP pairs in LD; and to Liz and Graham Derryberry for providing a script for obtaining the likelihood profiles. GoldenGate SNP data were produced with support from the Texas A&M Institute for Genome Sciences and Society at Texas A&M University in cooperation with Texas A&M AgriLife Research. An earlier version of the manuscript was improved by the constructive comments made by the editor N. Barton and three anonymous reviewers. Julio Chávez-Galarza and Dora Henriques were supported by PhD scholarships from the Fundação para a Ciência e Tecnologia (FCT) (SFRH/BD/68682/2010 and SFRH/BD/84195/2012, respectively). Miguel Carneiro was supported by the same institution through FCT Investigator (IF/00283/2014) and post-doc grants (SFRH/BPD/72343/2010). Financial support for this research was provided by FCT and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008 to M. A. Pinto.

References

Arias MC, Sheppard WS (1996) Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution*, **5**, 557–566.

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, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 13–42. Oxford University Press, Oxford.

Barton NH, Hewitt GM (1981) The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. *Heredity*, **47**, 367–383.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.

Beye M, Gattermeier I, Hasselmann M *et al.* (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.

Branco M, Ferrand N, Monnerot M (2000) Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity*, **85**, 307–317.

Branco M, Monnerot M, Ferrand N, Templeton AR (2002) Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution*, **56**, 792–803.

Campana MG, Hunt HV, Jones H, White J (2011) CorrSieve: software for summarizing and evaluating Structure output. *Molecular Ecology Resources*, **11**, 349–352.

Campos R, Branco M, Weiss S, Ferrand N (2007) Patterns of hemoglobin polymorphism (α-globin (HBA) and β-globin (HBB)) across the contact zone of two distinct phylogeographic lineages of the European rabbit (*Oryctolagus cuniculus*). In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Carneiro M, Stuart JEB, Afonso S *et al.* (2013) Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. *Molecular Ecology*, **22**, 2511–2525.

Carranza S, Arnold EN, Wade E, Fahd S (2004) Phylogeography of the false smooth snakes, *Macroprotodon* (Serpentes,

Colubridae): mitochondrial DNA sequences show European populations arrived recently from Northwest Africa. *Molecular Phylogenetics and Evolution*, **33**, 523–532.

Chávez-Galarza J, Henriques D, Johnston JS *et al.* (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 5890–5907.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie*, **20**, 93–101.

Cosson JF, Hutterer R, Libois R *et al.* (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, **14**, 1151–1162.

Crane E (1999) *The World History of Beekeeping and Honey Hunting*. Routledge, New York.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2001) Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. *Molecular Ecology*, **10**, 1733–1742.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2002) Microsatellite analysis of non-migratory colonies of *Apis mellifera iberica* from south-eastern Spain. *Journal of Zoological Systematics and Evolutionary Research*, **40**, 164–168.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2003) Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. *Genetics Selection Evolution*, **35**, 339–350.

De la Rúa P, Hernandez-Garcia R, Jiménez Y, Galián J, Serrano J (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Jaffé R, Dall'Olio R, Muñoz I, Serrano J (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie*, **40**, 263–284.

De la Rúa P, Jiménez Y, Galián J, Serrano J (2004) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. *Journal of Apicultural Research*, **43**, 162–166.

Derryberry EP, Derryberry G, Maley J, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14**, 652–663.

Durrett R, Buttel L, Harrison R (2000) Spatial models for hybrid zones. *Heredity*, **84**, 9–19.

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, NJ.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

European Commission (2013) EU Market Situation for Honey (ed. Agriculture and Rural Development), pp. 1–20.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Ferrand N, Branco M (2007) The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 207–235. Springer, The Netherlands.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet JM (2000) Hybrid origins of honey bees from italy (*Apis mellifera ligustica*) and sicily (*A. m. sicula*). *Molecular Ecology*, **9**, 907–921.

Franck P, Garnery L, Loiseau A *et al.* (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**, 145–154.

Garnery L, Franck P, Baudry E *et al.* (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution*, **30**, S31–S47.

Garnery L, Franck P, Baudry E *et al.* (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, S49–S74.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Gay L, Crochet P-A, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, **62**, 2789–2806.

Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman MW, Ferrand N (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology*, **17**, 4489–4499.

Godinho R, Crespo EG, Ferrand N (2008) The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670–4683.

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

Gonçalves H, Marínez-Solano I, Pereira RJ, Carvalho B, García-París M, Ferrand N (2009) High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent, multiple refugia: evidence from mtDNA, microsatellites and nuclear genealogies. *Molecular Ecology*, **18**, 5143–5160.

Gowen FC, Maley JM, Cicero C *et al.* (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC Evolutionary Biology*, **14**, 135.

Guillaumet A, Pons JM, Godelle B, Crochet PA (2006) History of the Crested Lark in the Mediterranean region as revealed

by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, **39**, 645–656.

Han F, Wallberg A, Webster MT (2012) From where did the Western honeybee (*Apis mellifera*) originate? *Ecology and Evolution*, **2**, 1949–1957.

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

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *TAG. Theoretical and Applied Genetics*, **38**, 226–231.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, **14**, 93–106.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kandemir I, Kence M, Sheppard WS, Kence A (2006) Mitochondrial DNA variation in honey bee (*Apis mellifera* L.) populations from Turkey. *Journal of Apicultural Research*, **45**, 33–38.

Klaassen CH, Gibbons JG, Fedorova ND, Meis JF, Rokas A (2012) Evidence for genetic differentiation and variable recombination rates among Dutch populations of the opportunistic human pathogen *Aspergillus fumigatus*. *Molecular Ecology*, **21**, 57–70.

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

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, **4**, 981–994.

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, **18**, 59–60.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera*-*Apis mellifera simensis* n. ssp. *Apidologie*, **42**, 425–437.

Miguel I, Baylac M, Iriondo M *et al.* (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie*, **42**, 150–161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and postglacial re-colonization routes in the western Europe. *Apidologie*, **38**, 141–155.

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.

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics*, **143**, 1021–1032.

Nielsen D, Page RE Jr, Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia*, **50**, 867–871.

Oleksa A, Chybicki I, Tofilski A, Burczyk J (2011) Nuclear and mitochondrial patterns of introgression into native dark bees (*Apis mellifera mellifera*) in Poland. *Journal of Apicultural Research*, **50**, 116–129.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Phillips BL, Baird SJ, Moritz C (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536–1548.

Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N (2009) Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 368–377.

Pinto MA, Henriques D, Chávez-Galarza J *et al.* (2014) Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, **53**, 269–278.

Pinto MA, Henriques D, Neto M *et al.* (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie*, **44**, 430–439.

Pinto MA, Muñoz I, Chávez-Galarza J, De la Rúa P (2012) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity. *Apidologie*, **43**, 663–673.

Pinto MA, Rubink WL, Patton JC, Coulson RN, Johnston JS (2005) Africanization in the United States: replacement of Feral European Honeybees (*Apis mellifera* L.) by an African Hybrid Swarm. *Genetics*, **170**, 1653–1665.

Polechova J, Barton N (2011) Genetic drift widens the expected cline but narrows the expected cline width. *Genetics*, **189**, 227–235.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, **18**, 2169–2178.

R Development Core Team (2013) *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rortais A, Arnold G, Alburaki M, Legout H, Garnery L (2011) Review of the *Dra*I COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). *Conservation Genetics Resources*, **3**, 383–391.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rosenberg NA, Mahajan S, Ramachandran S *et al.* (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, **1**, e70.

Ruttner F (1988) *Biogeography and Taxonomy of Honey Bees*. Springer, Berlin.
Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, **9**, 363–381.
Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, **14**, 1679–1685.
Shaibi T, Muñoz I, Dall'Olio R *et al.* (2009) *Apis mellifera* evolutionary lineages in Northern Africa: Libya, where orient meets occident. *Insectes Sociaux*, **56**, 293–300.
Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, **34**, 367–375.
Smith DR, Glenn TC (1995) Allozyme polymorphisms in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **86**, 12–16.
Smith DR, Palopoli MF, Taylor BR *et al.* (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **82**, 96–100.
Soland-Reckeweg G, Heckel G, Neumann P, Fluri P, Excoffier L (2009) Gene flow in admixed populations and implications for the conservation of the Western honeybee, *Apis mellifera*. *Journal of Insect Conservation*, **13**, 317–328.
Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.
Szymura JM, Barton NH (1991) The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, **45**, 237–261.
Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
Yuri T, Jernigan RW, Brumfield RT, Bhagabati NK, Braun MJ (2009) The effect of marker choice on estimated levels of introgression across an avian (Pipridae: *Manacus*) hybrid zone. *Molecular Ecology*, **18**, 4888–4903.
Wallberg A, Han F, Wellhagen G *et al.* (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.
Waples RS, Gaggiotti O (2006) Invited review: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.
Weir B (1996) *Genetic Data Analysis*. Sinauer Associates, Sunderland.
Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.
Whitfield CW, Behura SK, Berlocher SH *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.
Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin*, **1**, 80–83.
Winston M (1987) *The biology of the Honey Bee*. Harvard University Press, London.

M.A.P., J.C.P. and J.S.J. conceived the study. J.C.-G. performed all the analyses. D.H. assisted J.C.-G. with preparation of SNP files and STRUCTURE analyses. M.C. assisted J.C.-G. with cline analyses. M.A.P. coordinated sampling across Iberia and performed sampling across the Atlantic transect. J.S.J. coordinated SNP genotyping. M.A.P. validated SNP genotypes. D.H. assisted M.A.P. with mtDNA PCRs. J.C.-G. analysed the mtDNA sequences. M.A.P. and J.C.-G. interpreted the results with input from J.S.J., M.C. and D.H. J.R. provided the computing environment (software and hardware) for the simulations and analyses. M.A.P. and J.C.-G. wrote the manuscript with input from J.S.J. and M.C.

Data accessibility

SNP genotypes for the 711 Iberian honey bee individuals ordered by sampling location, from AT1 to MT6, in GenePop format: DRYAD entry doi 10.5061/dryad.1kk2k.

Geographic coordinates of sampling locations and environmental data for the 711 Iberian honey bee individuals ordered by sampling location, from AT1 to MT6, in Excel format: DRYAD entry doi 10.5061/dryad.1kk2k.

Aligned mtDNA sequences for 652 Iberian honey bee individuals ordered by sampling location, from AT1 to MT6, in Fasta format: DRYAD entry doi 10.5061/dryad.21s3t.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Statistics of physical distances (bp) of the 309 neutral SNPs used in the genetic analysis of the Iberian honey bee.

Table S2 Pairwise Φ_{PT} values among sampling sites estimated from the neutral SNP data set with GENALEX 6.5 (Peakall & Smouse 2012). Significance of φ_{PT} estimates was assessed using 10,000 permutations. Global Φ_{PT} value was 0.020 (*P*-value=0.001). Pairwise Φ_{PT} values ranged from 0.000 to 0.046. Φ_{PT} values marked in bold were significantly different from zero following Bonferroni correction. Sampling site codes are specified in Fig. 1.

Table S3 Cline parameter estimates for the best-fitting model of 33 SNP loci, complete SNP data set, neutral SNP data set and mtDNA. Cline width is presented as 1/maximum slope. Cline centre and width are measured in km, *P*min and *P*max are the allele frequencies at the ends of the cline, and δ and τ are the shape parameters for the mirror (M), left (L) and right (R) tails. Two log-likelihood unit support limits are presented in parentheses. The symbol * indicates coincident centre and the symbol $^{+}$ indicates concordant width, based on LRTs (*P*-value > 0.05). The left side AICc corresponds to the best-fitting model, and the right side AICc corresponds to the null model. Neutral SNP loci are marked in bold.

Table S4 Linkage disequilibrium (LD) and unbiased haploid genetic diversity (*uh*) estimated from the 309 neutral SNPs for the 23 sampling sites in the Iberian Peninsula (see Fig. 1 for location of sampling sites), as measured by r^2 and percentage of pairs of loci showing significant LD with Fisher's exact test [P_{LD} (Fisher)] before Bonferroni correction (only a single pair remained significant after Bonferroni correction). None of the SNP pairs exhibiting significant LD before Bonferroni correction, for which there is genomic information, are physically linked as they are located in different chromosomes.

Fig. S1 Estimated population structure of *A. m. iberiensis* inferred from the neutral SNP data set (309 loci) by DAPC (top) and STRUCTURE (bottom) at $K = 3$ to 5 clusters. Each of the 711 individuals included in the analyses is represented by a vertical bar partitioned into coloured segments, the size of each corresponding to the individuals' estimated membership proportions in each of the *K* clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT – Atlantic, CT – central, MT – Mediterranean), as indicated at the top bar. Black lines separate individuals from the 23 sampling sites, which are arranged from high *Q* (left) to low *Q* (right) in the blue cluster.

Fig. S2 Graphical display of the three methods (Δ*K*, ΔF_{st} and BIC) to predict the optimal *K* for the analysis of *A. m. iberiensis* population structure using the neutral SNP data set (309 loci) and the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013).

Fig. S3 Heat map of pairwise Φ_{PT} values between Iberian sampling sites estimated from the neutral SNP data set (309 loci) using GENALEX 6.5 (Peakall & Smouse 2012). The heat map clearly highlights northeastern populations (CT1-3, MT1-2) and CT8 as the most differentiated across Iberia. (a) The 23 sampling sites are arranged from north to south in each of the three transects (see Fig. 1). (b) The 23 sampling sites are arranged along the west–east transect traced for the geographic cline analysis (see dashed line in Fig. 6).

Fig. S4 Population structure of *A. m. iberiensis* estimated by (a) DAPC and (b) STRUCTURE from the neutral SNP data set (309 loci) at $K = 2$ clusters. The 23 sampling sites are arranged along the west–east transect traced for the geographic cline analysis (see dashed line in Fig. 6). Plots represent each of the 711 individuals by a vertical bar partitioned into two coloured segments (blue and red) corresponding to membership proportions (*Q*) in each of the two clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from high *Q* (left) to low *Q* (right) in the blue cluster.

Fig. S5 Mean membership proportion (±SE) in the blue cluster inferred from the neutral (309 loci) and the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013) with DAPC for each sampling site. Sampling sites are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT – Atlantic, CT – central, MT – Mediterranean).

Fig. S6 Maternal pattern, obtained from the $tRNA^{leu}$-cox2 intergenic mitochondrial region, and estimated structure inferred from the neutral SNP data set (309 loci) with DAPC at $K = 2$ clusters. Patterns of variation are displayed at both individual and sampling site level for each transect (AT – Atlantic, CT – central, MT – Mediterranean). Vertical plots display the mitotype (A in red; M in blue; C in orange, one single individual in AT3) and the membership proportions (*Q*) for each of the 711 individuals. Horizontal bar plots show mitotype frequencies (top) and the mean *Q* in blue and red clusters (bottom) at each sampling site.

Fig. S7 Estimated population structure of *A. m. iberiensis* inferred by the spatially explicit algorithm implemented by TESS for the neutral SNP data set (309 loci). Each of the 711 individuals included in the analyses is represented by a vertical bar partitioned into coloured segments, the size of each corresponding to the individuals' estimated *Q* in each of the *K* (from 2 to 5) clusters. Vertical black lines separate individuals from the 23 sampling sites, which are arranged by M (blue), C (orange, one single individual in AT3), and A (red) maternal lineages indicated by colour at the bottom. Sampling sites are arranged from north to south in each transect (AT – Atlantic, CT – central, MT – Mediterranean). Sampling site codes (from AT1 to MT6) are shown at the top bar.

Fig. S8 Plot of DIC values (*Y*-axis) against K_{max} (*X*-axis) obtained with TESS analysis under the admixture model BYM for **(a)** the neutral SNP data set and **(b)** the complete SNP data set.

Fig. S9 Analysis of global and local genetic structure, among 711 individuals of *A. m. iberiensis* from 23 sampling sites, by spatial principal component analysis (sPCA) using the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). Each square represents the score of an individual, which is positioned by its spatial coordinates. (a-d) The first four global scores of sPCA. (e-f) The first two local scores of sPCA.

Fig. S10 Spatially explicit analysis implemented by the software TESS for the 711 individuals of *A. m. iberiensis* using the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). (a) Plot of individuals' *Q* at the optimal $K = 3$ clusters. Each of the 711 individuals included in the analysis is represented by a vertical bar partitioned into three coloured segments (blue, red, and white) corresponding to *Q* in each of the three clusters. Maternal data (M lineage in blue, C lineage in orange, and A lineage in red) are shown at the bottom. Sampling sites and individuals within sampling sites are arranged as in Fig. 3. (b) Map of the Iberian Peninsula showing the two major clusters ($Q \geq 0.5$)

interpolated by TESS. Dots represent the locations of sampled apiaries across the Atlantic (AT1-8), central (CT1-9) and Mediterranean (MT1-6) transects.

Fig. S11 Map of the Iberian Peninsula with pie charts summarizing frequency data for each sampling site and plot of maximum-likelihood geographic cline for neutral (marked in bold) and selected SNP loci. The symbols * and $^{+}$ indicate the loci or data sets with coincident centre and concordant width, respectively (see Table S3). The dashed line placed in each map represents the transect traced from Lisbon (0 km) to Girona (1400 km) for the geographic cline analysis.