Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms (SNPs)

JULIO CHÁVEZ-GALARZA,[*] DORA HENRIQUES,[*] J. SPENCER JOHNSTON,[†] JOÃO C. AZEVEDO,[*] JOHN C. PATTON,[‡] IRENE MUÑOZ,[§] PILAR DE LA RÚA[§] AND M. ALICE PINTO[*]

[*]Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, *Campus* de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, [†]Dept. of Entomology, Texas A&M University, College Station, Texas 77843-2475, USA, [‡]Dept. of Forestry and Natural Resources, Purdue University, 715 W State St., West Lafayette, Indiana 4797-2061, USA, [§]Área de Biología Animal, Dpto. De Zoología y Antropología Física, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain.

Running title: Selection in Iberian honey bees

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

Abstract

Understanding the genetic mechanisms of adaptive population divergence is one of the most fundamental endeavors in evolutionary biology and is becoming increasingly important as it will allow predictions about how organisms will respond to global environmental crisis. This is particularly important for the honey bee, a species of unquestionable ecological and economical importance that has been exposed to increasing human-mediated selection pressures. Here, we conducted a SNP-based

genome scan in honey bees collected across an environmental gradient in Iberia, and used four F_{ST}-based outlier tests to identify genomic regions exhibiting signatures of selection. Additionally, we analyzed associations between genetic and environmental data for the identification of factors that might be correlated or act as selective pressures. With these approaches, 4.4 % (17 of 383) of outlier loci were cross-validated by four F_{ST}-based methods and 8.9 % (34 of 383) were cross-validated by at least three methods. Of the 34 outliers 15 were found to be strongly associated with one or more environmental variables. Further support for selection, provided by functional genomic information, was particularly compelling for SNP outliers mapped to different genes putatively involved in the same function such as vision, xenobiotic detoxification, and innate immune response. This study enabled a more rigorous consideration of selection as the underlying cause of diversity patterns in Iberian honey bees, representing an important first step towards the identification of polymorphisms implicated in local adaptation and possibly in response to recent human-mediated environmental changes.

Keywords: SNP, F_{ST} outlier tests, directional selection, balancing selection, genome scan, *Apis mellifera iberiensis*

Introduction

Patterns of genetic diversity are shaped by genome-wide (e.g. random genetic drift, gene flow) and locus-specific (e.g. natural selection) processes. While the former details population demography and phylogenetic history, the latter helps identify genes that are important for fitness and adaptation (Luikart *et al.* 2003). Accordingly, disentangling genome-wide (neutral) from locus-specific (selected) variation is a fundamental goal in evolutionary biology because the outcome not only leads to more robust inferences of demographic history but also to identification of ecologically relevant genetic variation involved in local adaptation. In the context of rapid human-induced environmental change (e.g. habitat fragmentation, climate change, introduction of novel parasites and diseases) such knowledge is becoming increasingly important for better managing and preserving genetic diversity (Allendorf & Luikart 2007).

Recent advances in computer technology and statistical genetic methods provide the tools for addressing that goal (reviewed by Luikart *et al.* 2003; Nielsen 2005; Storz 2005; Vasemägi & Primmer 2005; Helyar *et al.* 2011). Inspired by the original idea of Lewontin & Krakauer (1973) a family of statistical methods, known as the F_{ST}-based outlier tests, identify loci that exhibit frequencies significantly different than expected under neutrality for a given demographic model in genome scan data (Beaumont & Nichols 1996; Vitalis *et al.* 2001; Schlötterer 2002; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009), thereby allowing separation of neutral from selected variation. The basic rationale is that genetic differentiation is higher for loci affected by directional selection (directional outliers) and lower for loci under balancing selection (balancing outliers) as compared to neutral variation. Those loci are usually called directional outliers and balancing outliers, respectively. Selection can be further validated as the cause of outlier behavior by combining genetic and non-genetic data for

the identification of environmental factors that might act as selective pressures (Foll & Gaggiotti 2006; Joost *et al.* 2007; Nielsen *et al.* 2009; Coop *et al.* 2010; Gomez-Uchida *et al.* 2011; Nunes *et al.* 2011; Prunier *et al.* 2011; Shimada *et al.* 2011; Tsumura *et al.* 2012) and by illuminating functional roles of selected loci (Luikart *et al.* 2003; Vasemägi & Primmer 2005; Prunier *et al.* 2011; Shimada *et al.* 2011; Lehtonen *et al.* 2012; Tsumura *et al.* 2012), a task that is facilitated for organisms with annotated genomes.

Owing to increasingly affordable and quicker genotyping of numerous loci scattered in the genome of numerous individuals, it is becoming increasingly popular to implement outlier tests at the genome and population-wide scales. This approach has revealed signatures of selection and, whenever supported by environmental and genomic information, has provided unprecedented insights into the ecological and molecular basis of adaptation in numerous taxa including plants (Prunier *et al.* 2011; Tsumura *et al.* 2012), fishes (Renaut *et al.* 2011; Shimada *et al.* 2011), birds (Lehtonen *et al.* 2012), pigs (Ai *et al.* 2013), and insects (De Jong *et al.* 2013), among others. Adding to the list, in the present study a genome-wide scan using single nucleotide polymorphisms (SNPs) was conducted in the honey bee subspecies that is native to the Iberian Peninsula: the Iberian honey bee (*Apis mellifera iberiensis*).

A. m. iberiensis is one of 30 currently recognized subspecies of honey bees, which occur naturally in the Middle East, Africa, and Europe (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al.* 2011). Contrary to the African continent, where most honey bee subspecies are predominantly wild (Dietemann *et al.* 2009), in Europe, following arrival of the invasive mite *Varroa destructor* in early 1980s, wild colonies have virtually disappeared (Moritz *et al.* 2007, Jaffe *et al.* 2010).

Native European honey bee subspecies (including the Iberian honey bee) are now mostly confined to apiaries.

The Iberian honey bee has been the subject of numerous population genetic surveys and thus represents one of the best studied, and yet controversial, among all subspecies. Maternal and biparental genetic markers have revealed highly complex and incongruent patterns of variation (Smith *et al.* 1991; Garnery *et al.* 1995; Smith & Glenn 1995; Franck *et al.* 1998; Garnery *et al.* 1998a, b; Cánovas *et al.* 2008, 2011), which have led to competing hypotheses for the origin of the Iberian honey bee.

Early phylogeographical studies of morphology (Ruttner *et al.* 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al.* 1994; Smith & Glenn 1995; Arias *et al.* 2006) revealed the existence of a smooth gradient extending from North Africa to France with Iberian honey bees showing intermediate phenotypes. This pattern raised the hypothesis of primary intergradation and an African origin for this subspecies (Ruttner *et al.* 1978). However, mitochondrial polymorphisms showed the co-occurrence of highly divergent African-derived (lineage A) and western European-derived (lineage M) haplotypes forming not a smooth but a steep south-north cline in the Iberian Peninsula (Garnery *et al.* 1995; Franck *et al.* 1998; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008; Pinto *et al.* 2013), a pattern that was more consistent with a secondary contact scenario (Smith *et al.* 1991). Adding to the complexity, microsatellites (Franck *et al.* 1998; Garnery *et al.* 1998b; Miguel *et al.* 2011; Cánovas *et al.* 2011) and recent geometric morphometric data (Miguel *et al.* 2011) exhibited virtually no differentiation and no traces of African genes in Iberian honey bee populations and revealed a sharp break between Iberian and northern African populations thereby supporting neither hypothesis. The difference between maternal and bi-parental variation led Franck *et al.* (1998) to reject the secondary contact hypothesis

and propose historical human-assisted introductions of African colonies with selection the best explanation for the reported morphological and allozymic clines and the diffusion and maintenance of African haplotypes in the southwestern half of the Iberian Peninsula. A recent study using SNPs suggests, however, that while selection may have shaped the genome of lineage-M honey bees, which includes *A. m. iberiensis* and *A. m. mellifera*, the process occurred during ancient expansions from Africa into Western Europe, resurrecting the primary intergradation hypothesis (Zayed & Whitfield 2008).

Selection has repeatedly been invoked to explain cytonuclear latitudinal patterns (Franck *et al.* 1998; Garnery *et al.* 1998a, b), yet no previous effort has attempted to evaluate its relative importance in structuring Iberian honey bee populations. In this study, populations sampled across three north-south Iberian transects were subjected to SNP-based genome scans to evaluate the importance of selection in shaping diversity patterns of Iberian honey bees. The ecological and molecular bases of the genomic regions exhibiting signatures of selection were further supported and investigated by spatial analysis, through non-random associations between locus specific variation and environmental variables, and by functional annotations of genes marked by outlier SNPs. The approach followed here revealed unprecedented insights into Iberian honey bee diversity patterns, which might be helpful for future management and conservation of honey bees, an increasingly important endeavor given the current worldwide concern regarding honey bee health.

Methods

Sampling

A total of 711 honey bee haploid males, representing 23 sites (Fig. 1) and 237 apiaries, were collected in 2010 across three north-south transects in the Iberian Peninsula. The

sites were selected to represent both the natural distribution of *A. m. iberiensis* and a wide variety of climates ranging from the semi-arid in southeastern to oceanic in northwestern Iberia. One transect extended along the Atlantic coast (AT, 8 sites), one through the center (CT, 9 sites) and another along the Mediterranean coast (MT, 6 sites). The number of apiaries sampled per site varied between 8 and 13, with most sites having 10. Accordingly, sample size per site varied between 24 (8 apiaries per site x 3 hives per apiary) and 39, with most sites having 30 individuals. In each apiary samples were taken from the inner part of three different hives and placed into absolute ethanol. Samples were stored at -20 ºC until molecular analysis. Global positioning system (GPS) coordinates were recorded in the field for each apiary.

DNA extraction and SNP genotyping

Total DNA was extracted using a phenol-chloroform isoamyl alcohol (25:24:1) protocol (Sambrook et al. 1989) from the thorax of 711 individuals, each representing a single colony. A total of 1536 SNP loci were genotyped for those individuals 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.

The Oligo Pool consisted of the 768 most informative SNPs used previously to study honey bee population structure and evolution (Whitfield *et al.* 2006) combined with a newly developed set of 768 SNPs. Both sets of 768 SNPs were drawn from over 1.1 million SNPs defined by single base differences between (i) the reference genome of *A. mellifera* (Assembly 3.0; sequenced from the North American DH4 strain, which was primarily *A. m. ligustica*) and genome sequence traces of Africanized honey bees (largely *A. m. scutellata* admixed with the genomes of both western and eastern European honey bees) and (ii) observed polymorphisms in ESTs. In the first case SNPs

were named “ahb” and “AMB” whereas in the second they were named “est” (see Tables 1 and S1). The new 768 SNPs were selected with the goal of obtaining markers that were evenly spaced across the honey bee genome (D. Weaver, pers. comm.).

Genotype calling was performed using Illumina’s GenomeStudio® Data Analysis software. For each sample, intensity clusters generated automatically by the software were manually verified, and edited when necessary. SNPs with poorly separated clusters or low signals (110) were excluded from the dataset. For the remaining 1426 SNPs, most honey bee samples (695 out of 711) exhibited a call rate between 95 to 100 %. The rest of the samples (16) had a call rate lower than 95 % but above 90 %.

Environmental data

Publicly available environmental data were obtained for the location of each apiary. Altitude was estimated using 30 arc-second (~ 1 km) spatial resolution data from the WorldClim database (http://www.worldclim.org). Climatic data were extracted from two data sets. The first data set (spatial resolution of 0.5º, representing ~ 50 km), which covered the period 1901-2009, was obtained from the Climatic Research Unit (www.cru.uea.ac.uk), Norwich, UK, and consisted of precipitation (Prec), minimum temperature (Tmin), mean temperature (Tmean), maximum temperature (Tmax), and cloud cover (Cld). The second data set (spatial resolution of 1º, representing ~ 100 km), which covered the period 1983-2005 (http://eosweb.larc.nasa.gov), was downloaded from OPENEI (http://en.openei.org), and consisted of relative humidity (Rh) and insolation (Ins), which is the amount of radiation reaching the Earth’s surface per day (insolation on horizontal surface in $kWh/m^2/day$). All climatic data were integrated into a geographic information system (ArcGIS 9.3 from ESRI) to extract yearly, seasonal

and monthly data. Land use/land cover data for the Iberian Peninsula was extracted from the CORINE Land Cover 2006 vector data from the European Environment Agency (http://www.eea.europa.eu). Land cover was described for each apiary by calculating the percentage of level 3 land cover classes (Heymann *et al.* 1994) within a 3 km radius circular area (28.3 km^2). In order to remove redundant environmental variables, i.e. variables that were correlated at $|r| > 0.8$ (Manel *et al.* 2010), a principal component analysis was performed using the *ade4* package (Thioulouse *et al.* 1997). Using this procedure we kept 80 environmental variables (Table S2) for further analysis, from an initial data set of 123.

Detection of outlier loci by F_{ST}-based methods

Outliers were detected using four multiple-population F_{ST}-based methods. These methods assume varying demographic models to identify loci under selection as outliers in the extreme tails of theoretical null distributions of F_{ST}. The purpose of employing conceptually different approaches was to identify potential false positives. The first method was FDIST2 (Beaumont & Nichols 1996) as implemented in LOSITAN (Antão *et al.* 2008). This coalescence-based method uses an island model to identify as outlier loci those that present unusually low or high F_{ST} values compared to neutral expectations. Approximation to the mean neutral F_{ST} in the data set was accomplished by choosing the neutral mean F_{ST} option (99% confidence interval) and running 1,000,000 simulations in LOSITAN. The second method, implemented in ARLEQUIN 3.5.1.3, is a modification of FDIST2 that overcomes possible false positives in the presence of strong population structure by using a hierarchical island model (Excoffier *et al.* 2009). The hierarchical outlier analysis was performed by pooling the sites into an Atlantic group (all AT sites), an M-lineage group (CT1-CT3, MT1-MT3), and an A-

lineage group (CT4-CT9, MT4-MT6), based on maternal patterns (Fig. S1) and analysis of molecular variance (Table 2, Fig. S2). The parameters of the run included the presence of 50 groups of 100 demes with 100,000 iterations simulated. The third method, implemented in BAYESFST, uses Markov Chain Monte Carlo simulations to assess the significance of a locus-specific parameter that indicates selection in a model of F_{ST} (Beaumont & Balding 2004). This Bayesian method was performed using 3,000,000 iterations. The fourth method, implemented in BAYESCAN 2.01, uses a Bayesian approach and a reversible-jump Markov Chain Monte Carlo method to estimate the posterior probability that a given locus is under selection (Foll & Gaggiotti 2008). It tests two alternative models: one that includes the effect of selection and another that excludes it. The BAYESCAN analysis was conducted using 20 pilot runs of 100,000 iterations, 1,500,000 iterations (sample size of 75,000 and thinning interval of 20) and an additional burn-in of 500,000 iterations.

Introgression and ascertainment bias

Because introgression may mimic selection, prior to the F_{ST}-based tests the 711 Iberian individuals were assessed for introgression using STRUCTURE 2.3.3 (Pritchard *et al.* 2000) and by implementing a principal components analysis with ADEGENET 1.3-7 (Jombart 2008). Over 1075 polymorphic SNPs and a reference collection consisting of the two eastern European beekeepers-favorite honey bee subspecies, *A. m. ligustica* (17 individuals) and *A. m. carnica* (19 individuals) and the northern African subspecies *A. m. intermissa* (31 individuals) were used. These analyses revealed virtually no signs of contemporaneous or historical introgression from the two eastern European or the northern African subspecies (see Figs. S3 and S4 for the parameters settings and

results), suggesting that introgression should not be a confounding factor in the F_{ST}-based outlier detection.

Retaining uninformative monomorphic loci in the data set tends to increase dramatically the rate of false positives when searching for selection using F_{ST}-based methods (Nielsen *et al.* 2009; Gomez-Uchida *et al.* 2011). Accordingly, the Iberian data set was screened for monomorphic SNPs, as defined by a cut-off criterion of > 0.98 for the most common allele. This filtering process produced a final data set of 383 polymorphic loci (288 genomic and 95 EST-derived) for the Iberian honey bee, which was used in all subsequent analyses performed in this study. The high proportion of monomorphic SNPs in Iberian honey bees might be explained by ascertainment bias. While the SNP panel was relatively diverse it did not include *A. m. iberiensis*. Accordingly, when SNPs were genotyped on Iberian honey bees, an ascertainment bias was introduced. However, ascertainment bias is expected to affect every genotyped Iberian individual equally and thus not systematically bias any particular individual or population, and thus evidence for selection is not caused by a SNP discovery artifact.

Genomic information

Each outlier SNP's 100 bp flanking sequence was mapped to the Honey Bee Assembly 4.5 using BLAST in BEEBASE (hymenopteragenome.org/beebase) and NCBI (www.ncbi.nlm.nih.gov). Genomic position was ascertained using the Map Viewer tool available in NCBI. SNPs were classified as belonging to exons, introns, 3' or 5' untranslated region (UTR), or intergenic regions. Genes marked by SNPs were identified using the Official Gene Set 3.2 (BEEBASE) and Entrez Gene (NCBI). As functional annotation of the honey bee genome is incomplete, putative Gene Ontology classifications were ascribed to as many genes as possible, based on homology to

Drosophila melanogaster, using best-BLASTP hit and e-value cutoff 0.01 in FLYBASE (www.flybase.org) complemented by NCBI annotation.

Differentiation of neutral and directional outlier loci

Differentiation of putatively neutral and the strongest directional outliers was investigated by analysis of molecular variance (AMOVA). Two and three-level AMOVAs were performed using ARLEQUIN (Excoffier & Lischer 2010) with 10,000 permutations to assess whether levels of differentiation were significantly greater than zero. Two-level AMOVAs (a single group) were conducted to test the prediction of a higher variance attributed to sites for directional outliers as compared to putatively neutral loci. Three-level AMOVAs (multiple groups) were conducted to assess (i) whether putatively neutral and outlier variation were similarly structured and, if there were any structure, (ii) whether neutral and/or outlier structure were oriented predominantly east-west or north-south. To that end, each neutral and directional outlier data set was partitioned by transect (longitudinal grouping) and mtDNA lineage (latitudinal grouping) into different grouping combinations (see Fig. S2 for details) in an effort to search for structures that generated the highest variance fraction attributed to groups.

Associations between environmental variables and outlier loci

A spatial analysis was performed to identify associations between allelic frequencies and environmental variables by using the software matSAM (Joost *et al.* 2008). This approach may provide insightful clues about selective forces acting upon outlier loci. Multiple univariate logistic regression models were computed by matSAM at the individual level (representing a single hive). The software matSAM assesses the

significance of the coefficients calculated by the logistic regression function by implementing likelihood ratio (G) and Wald statistical tests. A model is considered significant only if the null hypothesis is rejected by both tests, after Bonferroni correction (Joost *et al.* 2007). The 383 SNP data set was tested against 80 environmental variables described above (see Table S2). The significance threshold level was set to 1.632E-7, corresponding to a 99% confidence interval following Bonferroni correction. In addition to the matSAM analysis performed at the apiary level, associations between land cover and the strongest directional outlier loci were further examined for the 23 sites through linear regressions of allelic frequencies and percentage of level 3 land cover classes within a circular area of 1963.5 km^2 (25 km radius) around each site's centroid. In this analysis different combinations of agriculture land cover classes were tested and only the best models were selected.

Results

Detection of outlier loci by F_{ST}-based methods

The genome scan approach implemented in this study identified, by at least one of the four F_{ST}-based methods, a total of 69 outlier loci (of 383; 18.0 %) at a 95 % confidence level (Tables 1 and S1). The detection rate of outliers varied among the four methods, with the highest number of loci obtained by LOSITAN (57 of 383; 14.9 %) and the lowest by BAYESCAN (17 of 383; 4.4 %). The approaches implemented by ARLEQUIN and BAYESFST detected 49 (12.8 %) and 41 loci (10.7 %), respectively. All loci detected by BAYESCAN were also identified by the other three methods, whereas 15 loci were exclusive to LOSITAN, six to ARLEQUIN, and four to BAYESFST. As predicted from other studies (Wilding *et al.* 2001; Bonin *et al.* 2006; Shimada *et al.* 2011; Wang *et al.* 2012), the number of outlier loci exhibiting a signal of

directional selection (50) was higher than that of balancing selection (19; Tables 1 and S1) and varied according to the F_{ST}-based method employed. Specifically, the number of outlier loci under directional and balancing selection was as follows, respectively: LOSITAN 43/14, ARLEQUIN 33/16, BAYESFST 27/14, and BAYESCAN 12/5 (Tables 1 and S1, Fig. S5).

The 69 outliers were dispersed throughout the 16 honey bee linkage groups (LG) with three loci still unplaced in a chromosome (Fig. 2). Linkage groups one, two, and six harbored the highest number of outlier loci with seven, 10, and seven, respectively. The lowest physical distance (linkage) between outlier loci was observed for pairs ahb1245/ahb1232 (LG1) and ahb10181/ahb10154 (LG9) with 37.9 kb and 8.3 kb, respectively. The remaining 65 loci were further apart, with physical distances varying between 65.9 kb and 26,304 kb for pairs est9938/est9912 (LG13) and est1222/ahb226 (LG1), respectively. Therefore, except for the two pairs above, given the honey bee's exceptionally high recombination rate of 23.2 cM/Mb (Beye *et al.* 2006) most outlier loci might be considered unlinked. Indeed, in an earlier study using 1,136 SNPs (many employed herein) across 14 honey bee subspecies and Africanized honey bees, Whitfield *et al.* (2006) observed a rapid decay of linkage disequilibrium over a distance of 5-10 kb.

Among the 69 outlier loci, 34 (23 directional and 11 balancing) were detected by at least three F_{ST}-based methods (Table 1) whereas 35 were detected by two or fewer F_{ST}-based methods (Table S1) and might be false positives. In contrast, 17 loci are strong candidates for selection as they were simultaneously detected by the four methods (Table 1). Among these 17, 10 loci (nine directional and one balancing) exhibited the strongest signal (P-value ≤ 0.005 and posterior probability ≥ 0.99) being therefore the best candidates (marked in bold in Table 1). The nine best directional

outliers exhibited large differences in allele frequencies among sites, contrasting with the sole candidate (AMB-00963630) for balancing selection with nearly even frequencies across the geographical range (Fig S1). In spite of the weaker signal, 17 additional loci can still be considered good candidates as they were detected by three F_{ST}-based methods (Table 1).

Genomic information shows that of the 34 outliers, 32 loci were located in or near genes (Table 1) that code for proteins (see Table S3 for accession numbers) involved in a diverse array of putative functions including signaling, structural, metabolism, regulation, transport and immunity. Among the 27 outliers that mapped to genes, 12 were located in introns, six in untranslated regions (3' or 5' UTR) and nine in exons, although none was predicted to induce amino acid changes.

Differentiation of neutral and directional outlier loci

In order to assess whether putatively neutral and directional outlier variation were similarly structured, two and three-level hierarchical AMOVAs were performed for different loci combinations (neutral *versus* the strongest outliers). For the two-level AMOVA, the percentage of the total variance among sites was substantially higher for outliers (15.97 %; *P-value* <0.0001) than for neutral loci (3.77 %; *P-value* <0.0001). As observed for the two-level AMOVA, when multiple groups were considered (three-level AMOVA), the variance component attributed to groups was higher for outlier loci than for neutral loci, regardless of the groupings tested (Table 2).

For neutral loci, the variance due to groups was low (0.75-1.06 %) and always smaller than the variance among sites (2.89-3.35 %). The best variance partitioning was obtained for six groups (1.06), although slightly better (1.03) than that obtained for three groups formed by the Atlantic sites, the M-lineage sites of central and

Mediterranean transects, and the A-lineage sites of central and Mediterranean transects (Table 2, Fig. S2). Separating sites in the central and Mediterranean transects, using the same criteria applied to the Atlantic sites, generated poorer variance partitioning (data not shown). Overall, these results indicate a very weak neutral substructure across the Iberian Peninsula, although the Atlantic populations seem to be slightly more differentiated.

In contrast to the result with neutral loci, when the strongest directional outliers were analyzed, the variance component attributed to groups was elevated (10.51-12.70 %) and was higher than the variance among sites (5.53-8.17 %), as long as the Atlantic transect was kept as a separate group. Grouping the Atlantic with central and/or Mediterranean sites always inflated the variance attributed to sites and lowered the among groups component (Table 2 and data not shown). In summary, these results indicate that while neutral structure is very weak, directional outlier structure is pronounced and stronger longitudinally than latitudinally.

Associations between environmental variables and outlier loci

In addition to the F_{ST}-based outlier approaches, genomic regions exhibiting a signal of selection were further confirmed by the spatial analysis implemented by matSAM at the individual level. Significant associations (≥ 99 % confidence level with Bonferroni correction), with at least one environmental variable, were detected for 33 (8.62 %) out of the 383 screened loci (Table S2). Of the 33 loci, five were exclusive to matSAM and 28 had previously been identified by at least one F_{ST}-based method as directional outliers (Tables 1 and S1).

The environmental variables that were more frequently associated with SNPs were precipitation and longitude, with 21 and 17 such associations, respectively (Tables

1, S1 and S2). The variables altitude and land cover were on the opposite side of the spectrum with no significant (after Bonferroni correction) associations detected. Most loci (20) were associated with a single or two variables. A few loci (four) were associated with more than four variables. The greatest number and strongest associations were found for locus ahb8266, which was correlated with latitude, precipitation, minimum temperature, mean temperature, maximum temperature, cloud cover, and insolation (Table 1).

Among the 34 loci that were detected by at least three and four F_{ST}-based methods, 15 and nine loci, respectively, showed significant (≥ 99 %) associations with at least one environmental variable (Table 1). Overall, only nine loci were simultaneously detected by the five methods, with confidence levels above 99.99%. These loci were the strongest candidates for directional selection and were mainly associated with latitude, longitude, precipitation, and insolation.

Regression analyses of allele frequencies and land cover, implemented at the site level, revealed significant associations for six of the nine strongest directional outliers (Table 3). The strongest correlations were detected for loci est10016, est11018, and est5112. Loci ahb1245, ahb2123, and est7297 were not associated with any of the land cover classes tested. The land classes that, individually or combined, produced the best models were all agricultural and included “non-irrigated arable land”, “permanently irrigated land”, “fruit trees and berry plantations”, “olive groves”, “annual crops associated with permanent crops”, “complex cultivation patterns”, and “agro-forestry areas”. None of the forest classes tested was found to be associated with allele frequencies (data not shown).

Discussion

Candidate SNP loci for selection

Determining whether an outlier is a marker of a selected locus or a false positive is a major concern when searching for adaptive molecular variation. A common strategy has been to seek for confirmatory evidence from multiple outlier approaches that generate the expected neutral distribution of F_{ST} estimates under different demographic scenarios (Luikart *et al.* 2003; Storz 2005; Vasemägi & Primmer 2005). In this study, of the 69 detected outliers, only 17 (all identified by BAYESCAN) were cross-validated by the four conceptually different methods, whereas 34 and 44 (Tables 1 and S1) were simultaneously detected by at least three and two methods, respectively. While discrepant results among outlier approaches have been repeatedly reported (Bonin *et al.* 2006; Shikano *et al.* 2010; Nunes *et al.* 2011; Shimada *et al.* 2011; De Jong *et al.* 2013), comparative simulations have shown that BAYESCAN outperforms under a wide range of scenarios, exhibiting the lowest rate of false positives (Pérez-Figueroa *et al.* 2010; Narum & Hess 2011). Accordingly, these 17 outliers are the best candidates for selection. At the same time, the proportion of outlier loci detected by at least four (4.4 %), three (8.9 %) or two (11.5 %) methods is within the range reported for other taxa (extensively reviewed by Nosil *et al.* 2009; Shikano *et al.* 2010; Nunes *et al.* 2011; Shimada *et al.* 2011; Buckley *et al.* 2012; among others), including honey bees (Zayed & Whitfield 2008), and all are potentially important markers, including 10 loci identified by only two methods that map to functional variation, which deserve further attention to avoid the risk of losing interesting candidates.

A common feature of outlier approaches is their higher power in detecting directional selection compared to balancing selection, a disparity that is particularly pronounced when levels of genetic differentiation are low (Beaumont & Balding 2004; Foll & Gaggiotti 2008). While this would explain a higher proportion of directional than

balancing outliers detected herein and in most studies (Nielsen *et al.* 2009; Cooke *et al.* 2012; Limborg *et al.* 2012), the fact that five loci were identified by every single method in a scenario of low differentiation (F_{ST} values ranged between 0.008 and 0.093 for the 383 loci; data not shown) across the Iberian honey bee range indicates a strong signal of balancing selection, especially for locus AMB-00963630, which was detected with a confidence level above 99 %.

Biological relevance of candidate genes and association with possible selection pressures

Support for selection comes from functional annotations of genes carrying outlier SNPs that relate directly to colony fitness, and their association with possible selection pressures identified by the spatial analysis (Table 1). Support is particularly compelling for outlier loci mapped to different genes putatively involved in the same function, which is the case of balancing and directional outliers that are mapped to genes encoding proteins that are related to vision, xenobiotic detoxification, and immune response.

Outlier SNPs mapped to genes related to vision. Five outlier SNPs mark genes associated with vision. The strong directional outlier ahb8266 is mapped to an intron of a gene that encodes a transmembrane glycoprotein named teneurin 3-like isoform 1 (GB12816). Its putative ortholog in *D. melanogaster,* tenascin major, has been shown to accomplish an important role during neural development and in the process of vision (Kinel-Tahan *et al.* 2007). Balancing outlier ahb142 is also located in an intronic region of a transmembrane protein named Semaphorin 1A (GB11468), which is required for the synapse formation and axon guidance (Godenschwege *et al.* 2002). This protein has

been shown to participate in regulating the photoreceptor axon guidance in the visual system of *D. melanogaster* establishing an appropriate topographic termination pattern in the optic lobe (Cafferty *et al.* 2006). Three additional outliers, two directional (est7297, est2423) and one balancing (ahb4188), are putatively related to vision through their participation in the visual cycle, which has been shown to be a major pathway contributing to the maintenance of rhodopsin levels in *D. melanogaster* (Wang *et al.* 2010). The strong outlier est7297 is located in the intronic region of a gene that encodes a 15-hydroxyprostaglandin dehydrogenase [NAD+]-like protein (GB11685) belonging to NADB-Rossman superfamily, based on sequence comparison. Its putative ortholog in *D. melanogaster* is a photoreceptor dehydrogenase that participates in regenerating the chromophore for the production of rhodopsin (Wang *et al.* 2010). Outlier est2423 is located in the exon of a gene that encodes a retinol dehydrogenase 11-like protein (GB11195), which might be implicated in the production of the chromophore by catalyzing the *cis*-retinol to *cis*-retinal (Kim *et al.* 2005; Belyaeva *et al.* 2009; Wang *et al.* 2010). Finally, balancing outlier ahb4188 is mapped to the exon of a gene encoding for a blop (blue sensitive) opsin (GB13493), which accomplishes a receptor function of the chromophore and forms a rhodopsin responsible for absorbing the 440 nm (blue spectra) wavelength (Townson *et al.* 1998; Earl & Britt 2006). Fixation of an alternative allele could confer better protein activity or structural conformation (Camps *et al.* 2007) for the production and regeneration of the chromophore, while the maintenance of polymorphism could provide an adaptive advantage favoring the absorbance of different wavelengths within blue spectra, as recently reported for titi monkeys (Bunce *et al.* 2011).

The three directional outliers (ahb8266, est7297, and est2423) are associated with insolation, among other environmental variables, suggesting that these loci mark

genomic regions that are involved in local vision-associated adaptation. Vision is a key component of foraging behavior (Winston 1987). Flight departures from the colony for foraging have been both positively and negatively correlated with solar radiation intensity (Burrill & Dietz 1981). Once outside, foragers rely on their highly developed trichromatic visual system and use color discrimination for finding food sources and homing to the hive (reviewed by Menzel & Müller 1996). It is therefore possible that selection favors alleles that enable a more efficient light perception at contrasting climates such as the Mediterranean with long periods of sunny skies in southern Iberia, and the Atlantic with long periods of rain and cloudy skies, particularly in northwestern Iberia.

Outlier SNPs mapped to genes related to xenobiotic detoxification. Three directional SNPs mark genes involved with detoxification of xenobiotics. Outlier est10016 is located in the exon of a gene that encodes CYP6AS7 protein (GB18052), which belongs to the cytochrome P450s monooxygenases superfamily. This protein superfamily plays a major role in the protection against xenobiotics and has been implicated in tolerance to plant toxins (Scott & Wen 2001; Mao *et al.* 2009) and evolved resistance to pesticides in many insects (Feyereisen 1999; Ffrench-Constant *et al.* 2004; Li *et al.* 2007), including tolerance of pyrethroid insecticides in honey bees (Pilling *et al.* 1995; Johnson *et al.* 2006; Mao *et al.* 2011). Outlier ahb1245 is located in an intron of a putative gene that codes for a microsomal glutathione S–transferase 2 (GB10566), which belongs to another major superfamily associated with detoxification, the glutathione-S-transferases (GSTs). This gene has been implicated in the detoxification of metabolites formed by cytochrome P450 enzymes (Yu 2002; Claudianos *et al.* 2006). Finally, outlier est5302 maps to the exon of a gene (GB10566) that codes for a protein belonging to the UDP-

glycosyltransferases (UGT) superfamily whose probable function is related to olfaction and detoxification mechanisms, based on its putative *D. melanogaster* UDP-glucosyltransferase 35b ortholog (Bull & Whitten 1972; Wang *et al.* 1999). In addition to the xenobiotic metabolism, these genes might also play a role in the defense against pathogens. A study in *D. melanogaster* suggested that genes encoding GSTs and cytochromes P450 likely participate in the detoxification of reactive oxygen species produced during microbial killing to protect the gut epithelium (Buchon *et al.* 2009). A similar mechanism might be involved in honey bees, as suggested by an elevated GST activity in the honey bee gut infected by one of its most important pathogens, the microsporidium *Nosema ceranae* (Dussaubat *et al.* 2012).

It makes perfect sense that three outlier SNPs map to loci involved with detoxification. During their life time honey bees are exposed to a wide range of natural (e.g. phytochemicals present in nectar, pollen and propolis) and synthetic (agricultural pesticides) xenobiotics. While honey bees have evolved detoxification mechanisms to metabolize natural phytochemicals, they are known for being unusually sensitive to a range of insecticides (Hardstone & Scott 2010). Agricultural insecticides used for crop protection are increasingly recognized as major drivers of recent worldwide honey bee losses. In 2012 alone, more than 100 papers and reports were published on the impact of insecticides, particularly neonicotinoids, on bees (mentioned by Osborne 2012). In addition to acute poisoning accidents (e.g. the case in Germany described by vanEngelsdorp & Meixner 2010), neonicotinoids seem to have a pervasive impact on honey bees through short and long-term chronical exposure to sub-lethal doses that can result in reduction of disease resistance and breeding success as well as behavioral disturbances including problems in flying and navigation, impaired memory and learning, and reduced foraging ability (extensively reviewed by Blacquière *et al.* 2012).

Accordingly, honey bees are potentially under strong selection pressure when they forage in intensive farming areas. In this study, we observed significant correlations between agricultural land cover (a surrogate of agrochemicals use) and outlier SNPs est10016 and est5302 (Table 3). Given increasing evidence of sub-lethal effects of pesticides on colony survival, this finding deserves further investigation.

Outlier SNPs mapped to genes related to innate immune response. Two strong outlier SNPs, which showed contrasting selection signatures, mark genes implicated in immune response. While est11018 exhibits a signal of directional selection and ahb6903 is seemingly under balancing selection, both modes of selection have been shown to be associated with immune response in fruit fly (Jiggins & Hurst 2003; Sackton *et al.* 2007), cod (Gomez-Uchida *et al.* 2011) and humans (Hollox & Armour 2008). Directional outlier est11018 maps to the 5'UTR region of a gene that encodes a putative Nimrod C2 protein (GB13979). Nimrod C2 belongs to a diverse class of transmembrane proteins that have been shown to function as phagocytosis receptors and/or microbial binding factors, suggesting an important role in the cellular immunity and elimination of apoptotic cells (Kurucz *et al.* 2007; Sackton *et al.* 2007; Somogyi *et al.* 2008). Balancing outlier ahb6903 is located in the intron of a gene encoding a Dscam-Down syndrome cell-adhesion molecule (GB15141). *DSCAM*, an orthologous gene in *D. melanogaster*, encodes for a receptor that has long been implicated in neuronal development, although more recently it has also been shown to participate in the phagocytosis of infectious non-self (Watson *et al.* 2005). As expected, the two outliers were not associated with any of the environmental variables tested in this study. Despite potential direct influence of climate on both host and pathogen, selection pressure on genomic regions related to immune response is imposed by parasites and pathogens,

which in a managed population such as the honey bee, will be more dependent on beekeeping activities.

As was true with detoxification, selection for alleles associated with immunity makes biological sense. Honey bees face a multitude of enemies which, in addition to pesticides, are recognized as key contributors to worldwide population declines (reviewed by Potts *et al.* 2010 and vanEngelsdorp & Meixner 2010). While honey bees have evolved both group (e.g. grooming, nest hygiene, necrophoric behavior, behavioral fever) and individual (mechanical, physiological and immune responses) defense mechanisms, their health has been challenged in unprecedented ways by an increasing number of emerging enemies including parasites, pathogens, and predators. Among the most significant are the parasitic mite *V. destructor*, the gram-positive bacterium *Paenibacillus larvae* and the microsporidia *Nosema apis* and *N. ceranae* (vanEngelsdorp & Meixner 2010). The latter has been particularly worrisome in Spain, where it has been reported as a highly virulent pathogen and a potential culprit in colony collapse disorder (Higes *et al.* 2009; Higes *et al.* 2010), although epidemiological studies in the US (Cox-Foster *et al.* 2007; vanEngelsdorp *et al.* 2009) and Germany (Gisder *et al.* 2010) failed to associate this emerging enemy to colony losses. While the impact of *V. destructor* (on its own and as a vector of viruses) has been alleviated by in-hive acaricides (although mite-resistance has been reported for an increasing number of chemicals), colonies are rarely treated against the pathogens *P. larvae* and *Nosema spp.* because antibiotics are interdicted in the European Union due to honey contamination. Accordingly, Iberian honey bees have likely been under pathogen-driven selective pressure, although only two strong outliers, both mapped to genes putatively involved in phagocytosis immune response, were identified in this study. In *D. melanogaster* genes implicated in phagocytosis have been found to have an important role in deterring

bacteria (Kurucz *et al.* 2007) and fungi (Jin *et al.* 2008). While little is known about the importance of phagocytosis in honey bees (Evans & Spivak 2010), gene-expression studies have reported differential regulation of *NIMC2* or *DSCAM* genes in colonies exposed to different parasites and pathogens (Evans 2006; Navajas *et al.* 2008; Bull *et al.* 2012; Nazzi *et al.* 2012) suggesting that these genes might participate in the immune response.

The SNP outliers identified in this study are only a small fraction of the total SNP variation in the honey bee. As such, they are markers of selection, and unlikely to be direct targets of selection. Many of the outlier SNPs were located in introns and UTRs, and likely have little or no impact on the polypeptide, although these same regions may be involved in regulation or alternative splicing (Ladd & Cooper 2002). Those outlier SNPs that were located in exons were all silent substitutions, which have widely been assumed to have no effect on protein fitness, although comparison of all dipteran and hymenopteran sequenced genomes, shows that codon usage bias is highest in the honey bee (Behura & Severson 2012), and is associated with differential translation rates, mRNA stability and modification of protein structure and activity (Tsai *et al.* 2008). A few SNPs were located outside coding regions. Accordingly, it is likely that many SNPs, if not all, have hitchhiked with linked polymorphisms that are the causative mutations under selection.

Differentiation of neutral and directional outlier loci

Partition of neutral variation was negligible across the Iberian honey bee range, barely reflecting the steep north-south maternal cline formed by the two highly divergent western European (M) and African (A) lineages extensively reported in the literature (Garnery *et al.* 1995; Franck *et al.* 1998; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas

et al. 2008; Pinto *et al.* 2013) and further confirmed in this study (Fig. S1). A pattern of virtually no differentiation was also revealed by geometric morphometric data (Miguel *et al.* 2011) and microsatellites (Franck *et al.* 1998; Garnery *et al.* 1998b; Cánovas *et al.* 2011) but not by traditional morphometry (Ruttner *et al.* 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al.* 1994; Smith & Glenn 1995; Arias *et al.* 2006), which have been reported to exhibit a latitudinal gradient along the Mediterranean coast of the Iberian Peninsula.

Contrasting with neutral loci, geographical partitioning recovered by directional outliers argues for selection as an underlying force shaping a latitudinal gradient, as suggested by others (Franck *et al.* 1998; Garnery *et al.* 1998a, b). The spatial analysis identified precipitation, and to a lesser extent temperature as associated with the north-south (latitudinal) selection (Table 1). Precipitation was associated along with latitude for nine outlier SNPs, while mean and maximum temperatures were also associated with the precipitation and latitude for four of these nine. What was not expected was that several directional outliers were found even more strongly correlated with longitude than latitude (see AMOVA and matSAM analysis). Further, while matSAM identified precipitation associated with both longitude and latitude, insolation was identified as uniquely associated with longitude. Insolation is thus identified as a novel parameter shaping population structure east to west. The observation of east-west outlier structuring further adds to the complexity of Iberian honey bees suggesting that the Atlantic portion of the Iberian Peninsula harbors an additional important component of the Iberian honey bee diversity.

Concluding remarks

In this study genome-wide scans using 383 polymorphic SNP loci were conducted in searching for the footprints of selection in the environmentally heterogeneous Iberian honey bee range. A combination of methods was employed leading to detection of several SNPs marking genomic regions that are promising candidates for adaptation. SNP loci exhibiting strong outlier behavior were cross-validated by at least three conceptually different F_{ST}-based methods, and a subset was found to be associated with environmental variables that may be causal or correlated selection pressures.

Further support for the role of selection in shaping variation in Iberian honey bees was provided by the functional relevance of the genes carrying outlier SNPs. This is particularly compelling when independent SNPs are mapped to genes that are functionally related. While genome-wide scans provide a powerful way of highlighting candidate genes for selection, experimental support for selection from functional and expression studies (Schluter *et al.* 2010; Renaut *et al.* 2011; Riveron *et al.* 2013) and also indirect evidence from sequence variation analysis (Low *et al.* 2007; Wood *et al.* 2008; Kent *et al.* 2011; Oliveira *et al.* 2012) are ultimately required to make causal inferences about the molecular basis of adaptation of Iberian honey bees. Nevertheless, the approach pursued here enabled, for the first time, a more rigorous consideration of selection as the underlying cause of diversity patterns in Iberian honey bees, whereas in previous studies (Franck *et al.* 1998; Garnery *et al.* 1998a, b) evocation of this evolutionary force to explain clinal variation in the Iberian Peninsula was speculative. Furthermore, our findings represent an important first step towards the identification of polymorphisms implicated in local adaptation and possibly in response to recent human-mediated environmental changes.

Acknowledgments

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 Gomez Pajuelo for providing numerous contacts of beekeepers in Spain; to Andreia Brandão, Inês Moura, and Margarida Neto for helping in the field work; 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 for providing the facilities at Texas A&M University for SNP genotyping; to Miguel Vaz-Pinto and João Paulo Castro for GIS and GPS assistance; to José Rufino for providing computer power for running the simulations. GoldenGate SNP data were produced with support from the Whole Systems Genomics Initiative at Texas A&M University in cooperation with Texas A&M AgriLife Research. Five reviewers made constructive and insightful comments on an earlier draft of the manuscript. Julio Chávez-Galarza was supported by Fundação para a Ciência e Tecnologia through the PhD scholarship SFRH/BD/68682/2010. Irene Muñoz was supported by the Spanish Ministry of Education, Culture and Sports. Financial support for this research was provided by Fundação para a Ciência e Tecnologia and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008 to M. A. Pinto.

Author contributions

M.A.P., J.C.P., and J.S.J. designed the study. J.C.-G. performed all the analyses and searched for the GO information. D. H. assisted J.C.-G. with files' preparation, STRUCTURE and AMOVA analyses. M.A.P., I.M. and P.D.L.R. coordinated sampling across Iberia and performed sampling across the Atlantic transect. J.C.A. produced the sampling maps, conducted the land use-allele frequency analysis, and assisted with GIS.

J.S.J. coordinated SNP genotyping. M.A.P. validated SNP genotypes. J.C.-G, M.A.P., J.S.J and D. H. interpreted the results. J.C.-G. and M.A.P. wrote the paper with input from the other authors.

References

Ai H, Huang L, Ren J (2013) Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. *PLoS ONE* **8**, e56001.

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations*, 1st edn. Blackwell Publishing.

Antão T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics* **9**, 323.

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.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* **263**, 1619-1626.

Behura SK, Severson DW (2012) Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS ONE* **7**, e43111.

Belyaeva OV, Lee SA, Kolupaev OV, Kedishvili NY (2009) Identification and characterization of retinoid-active short-chain dehydrogenases/reductases in *Drosophila melanogaster*. *Biochimica Et Biophysica Acta-General Subjects* **1790**, 1266-1273.

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

Blacquière T, Smagghe G, van Gestel CAM, Mommaerts V (2012) Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology* **21**, 1581-1581.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution* **23**, 773-783.

Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B (2009) *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host and Microbe* **5**, 200-211.

Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology* **21**, 267-280.

Bull DL, Whitten CJ (1972) Factors influencing organophosphorus insecticide resistance in tobacco budworms. *Journal of Agricultural and Food Chemistry* **20**, 561-&.

Bull JC, Ryabov EV, Prince G *et al.* (2012) A strong immune response in young adult honeybees masks their increased susceptibility to infection compared to older bees. *PLoS Pathogens* **8**, e1003083.

Bunce JA, Isbell LA, Neitz M *et al.*(2011) Characterization of opsin gene alleles affecting color vision in a wild population of titi monkeys (*Callicebus brunneus*). *American Journal of Primatology* **73**, 189-196.

Burrill RM, Dietz A (1981) The response of honey bees to variations in solar radiation and temperature. *Apidologie* **12**, 319-328.

Cafferty P, Yu L, Long H, Rao Y (2006) Semaphorin-1a functions as a guidance receptor in the *Drosophila* visual system. *Journal of Neuroscience* **26**, 3999-4003.

Camps M, Herman A, Loh E, Loeb LA (2007) Genetic constraints on protein evolution. *Critical Reviews in Biochemistry and Molecular Biology* **42**, 313-326.

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.

Claudianos C, Ranson H, Johnson RM *et al.* (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Molecular Biology* **15**, 615-636.

Cooke GM, Chao NL, Beheregaray LB (2012) Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. *Molecular Ecology* **21**, 2410-2427.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**, 1411-1423.

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

Cox-Foster DL, Conlan S, Holmes EC *et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283-287.

De Jong MA, Collins S, Beldade P, Brakefield PM, Zwaan BJ (2013) Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. *Molecular Ecology* **22**, 341-353.

Dietemann V, Pirk CWW, Crewe R (2009) Is there a need for conservation of honeybees in Africa? *Apidologie* **40**, 285-295.

Dussaubat C, Brunet JL, Higes M *et al.* (2012) Gut pathology and responses to the microsporidium *Nosema ceranae* in the honey bee *Apis mellifera*. *PLoS ONE* **7**, e37017.

Earl JB, Britt SG (2006) Expression of *Drosophila* rhodopsins during photoreceptor cell differentiation: Insights into R7 and R8 cell subtype commitment. *Gene Expression Patterns* **6**, 687-694.

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

Evans JD (2006) Beepath: An ordered quantitative-PCR array for exploring honey bee immunity and disease. *Journal of Invertebrate Pathology* **93**, 135-139.

Evans JD, Spivak M (2010) Socialized medicine: Individual and communal disease barriers in honey bees. *Journal of Invertebrate Pathology* **103**, S62-S72.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb)* **103**, 285-298.

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.

Feyereisen R (1999) Insect P450 Enzymes. *Annual Review of Entomology* **44**, 507–533.

Ffrench-Constant RH, Daborn PJ, Le Goff G (2004) The genetics and genomics of insecticide resistance. *Trends in Genetics* **20**, 163-170.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**, 875-891.

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.

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, 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 B, P., Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology* **4**, 465-471.

Gisder S, Hedtke K, Mockel N *et al.* (2010) Five-year cohort study of *Nosema spp*. in Germany: Does climate shape virulence and assertiveness of *Nosema ceranae*? *Applied and Environmental Microbiology* **76**, 3032-3038.

Godenschwege TA, Hu HL, Shan-Crofts X, Goodman CS, Murphey RK (2002) Bi-directional signaling by Semaphorin 1a during central synapse formation in *Drosophila*. *Nature Neuroscience* **5**, 1294-1301.

Gomez-Uchida D, Seeb JE, Smith MJ *et al.* (2011) Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evolutionary Biology* **11**, 48.

Hardstone MC, Scott JG (2010) Is *Apis mellifera* more sensitive to insecticides than other insects? *Pest Management Science* **66**, 1171-1180.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.

Heymann Y, Steenmans C, G Croissille G, Bossard M (1994) CORINE Land Cover. Office for Official Publications of the European Communities, Luxembourg.

Higes M, Martin-Hernandez R, Garrido-Bailon E *et al.* (2009) Honeybee colony collapse due to *Nosema ceranae* in professional apiaries. *Environmental Microbiology Reports* **1**, 110-113.

Higes M, Martin-Hernandez R, Meana A (2010) *Nosema ceranae* in Europe: an emergent type C nosemosis. *Apidologie* **41**, 375-392.

Hollox EJ, Armour JAL (2008) Directional and balancing selection in human beta-defensins. *BMC Evolutionary Biology* **8**, 113.

Jaffe R, Dietemann V, Allsopp MH *et al.* (2010) Estimating the density of honeybee colonies across their natural range to fill the gap in pollinator decline censuses. *Conservation Biology* **24**, 583-593.

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.

Jiggins FM, Hurst GDD (2003) The evolution of parasite recognition genes in the innate immune system: Purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *Journal of Molecular Evolution* **57**, 598-605.

Jin LH, Shim J, Yoon JS *et al.* (2008) Identification and functional analysis of antifungal immune response genes in *Drosophila*. *PLoS Pathogens* **4**, e1000168.

Johnson RM, Wen Z, Schuler MA, Berenbaum MR (2006) Mediation of pyrethroid insecticide toxicity to honey bees (Hymenoptera : Apidae) by cytochrome P450 monooxygenases. *Journal of Economic Entomology* **99**, 1046-1050.

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

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955-3969.

Joost S, Kalbermatten M, Bonin A (2008) Spatial analysis method(SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources* **8**, 957-960.

Kent CF, Issa A, Bunting AC, Zayed A (2011) Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*. *Molecular Ecology* **20**, 5226-5235.

Kim TS, Maeda A, Maeda T *et al.*(2005) Delayed dark adaptation in 11-cis-retinol dehydrogenase-deficient mice - A role of RDH11 in visual processes in vivo. *Journal of Biological Chemistry* **280**, 8694-8704.

Kinel-Tahan Y, Weiss H, Dgany O, Levine A, Wides R (2007) *Drosophila* odz gene is required for multiple cell types in the compound retina. *Developmental Dynamics* **236**, 2541-2554.

Kurucz E, Markus R, Zsamboki J *et al.* (2007) Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Current Biology* **17**, 649-654.

Ladd AN, Cooper TA (2002) Finding signals that regulate alternative splicing in the postgenomic era. *Genome Biology* **3**, reviews0008.

Lehtonen PK, Laaksonen T, Artemyev AV *et al.* (2012) Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity* **108**, 431-440.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175-195.

Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, **52**, 231-253.

Limborg MT, Helyar SJ, de Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology* **21**, 3686-3703.

Low WY, Ng HL, Morton CJ *et al.* (2007) Molecular evolution of glutathione S-transferases in the genus *Drosophila*. *Genetics* **177**, 1363-1375.

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.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology* **19**, 3824-3835.

Mao W, Rupasinghe SG, Johnson RM *et al.* (2009) Quercetin-metabolizing CYP6AS enzymes of the pollinator *Apis mellifera* (Hymenoptera: Apidae). *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **154**, 427-434.

Mao W, Schuler MA, Berenbaum MR (2011) CYP9Q-mediated detoxification of acaricides in the honey bee (*Apis mellifera*). *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12657-12662.

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.

Menzel R, Müller U (1996) Learning and memory in honey bees: from behavior to neural substrates. *Annual Review of Neuroscience* **19**, 379–404.

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 post-glacial re-colonization routes in the western Europe. *Apidologie* **38**, 141-155.

Moritz RFA, Kraus FB, Kryger P, Crewe RM (2007) The size of wild honeybee populations (*Apis mellifera*) and its implications for the conservation of honeybees. *Journal of Insect Conservation* **11**, 391-397.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.

Navajas M, Migeon A, Alaux C *et al.*(2008) Differential gene expression of the honey bee *Apis mellifera* associated with *Varroa destructor* infection. *BMC Genomics* **9**, 301.

Nazzi F, Brown SP, Annoscia D *et al.* (2012) Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. *PLoS Pathogens* **8**, e1002735.

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

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.*(2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197-218.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology* **20**, 193-205.

Oliveira TYK, Harris EE, Meyer D, Jue CK, Silva WA (2012) Molecular evolution of a malaria resistance gene (DARC) in primates. *Immunogenetics* **64**, 497-505.

Osborne JL (2012) ECOLOGY Bumblebees and pesticides. *Nature* **491**, 43-45.

Pérez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* **23**, 2267-2276.

Pilling ED, Bromleychallenor KAC, Walker CH, Jepson PC (1995) Mechanism of synergism between the pyrethroid insecticide lambda-cyhalothrin and the imidazole fungicide prochloraz, in the honeybee (*Apis mellifera* L.). *Pesticide Biochemistry and Physiology* **51**, 1-11.

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.

Potts SG, Biesmeijer JC, Kremen C *et al.* (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25**, 345-353.

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

Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology* **20**, 1702-1716.

Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus spp*.). *Molecular Ecology* **20**, 545-559.

Riveron JM, Irving H, Ndula M *et al.* (2013) Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector

Anopheles funestus. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 252-257.

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

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.

Sackton TB, Lazzaro BP, Schlenke TA *et al.*(2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics* **39**, 1461-1468.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**, 753-763.

Schluter D, Marchinko KB, Barrett RDH, Rogers SM (2010) Natural selection and the genetics of adaptation in threespine stickleback. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 2479-2486.

Scott JG, Wen ZM (2001) Cytochromes P450 of insects: the tip of the iceberg. *Pest Management Science* **57**, 958-967.

Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie* **34**, 367-375.

Shikano T, Ramadevi J, Merilä J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution* **27**, 2775-2789.

Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution* **28**, 181-193.

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.

Somogyi K, Sipos B, Penzes Z *et al.* (2008) Evolution of genes and repeats in the Nimrod superfamily. *Molecular Biology and Evolution* **25**, 2337-2347.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: A multivariate analysis and graphical display software. *Statistics and Computing* **7**, 75-83.

Townson SM, Chang BSW, Salcedo E *et al.* (1998) Honeybee blue-and ultraviolet-sensitive opsins: Cloning, heterologous expression in *Drosophila*, and physiological characterization. *Journal of Neuroscience* **18**, 2412-2422.

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C *et al.* (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *Journal of Molecular Biology* **383**, 281-291.

Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T (2012) Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. *Heredity* **109**, 349-360.

vanEngelsdorp D, Evans JD, Saegerman C *et al.* (2009) Colony collapse disorder: a descriptive study. *PLoS ONE* **4**, e6481.

vanEngelsdorp D, Meixner MD (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* **103**, S80-S95.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology* **14**, 3623-3642.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811-1823.

Wang Q, Hasan G, Pikielny CW (1999) Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. *Journal of Biological Chemistry* **274**, 10309-10315.

Wang T, Chen GP, Zan QJ, Wang CB, Su YJ (2012) AFLP genome scan to detect genetic structure and candidate loci under selection for local adaptation of the invasive weed *Mikania micrantha*. *PLoS ONE* **7**, e41310.

Wang XY, Wang T, Jiao YC, von Lintig J, Montell C (2010) Requirement for an enzymatic visual cycle in *Drosophila*. *Current Biology* **20**, 93-102.

Watson FL, Puttmann-Holgado R, Thomas F *et al.* (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* **309**, 1874-1878.

Whitfield CW, Behura SK *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera. Science* **314**, 642-645.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611-619.

Winston M (1987) *The biology of the honey bee*, London.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology* **17**, 3123-3135.

Yu SJ (2002) Biochemical characteristics of microsomal and cytosolic glutathione S-transferases in larvae of the fall armyworm, *Spodoptera frugiperda* (J. E. Smith). *Pesticide Biochemistry and Physiology* **72**, 100-110.

Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3421-3426.

Data Accessibility

SNP genotypes for the 711 Iberian honey bee individuals: DRYAD entry doi:

Geographical coordinates of sampling locations and environmental data for the 711 Iberian honey bee individuals: DRYAD entry doi:

Supporting information

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

Table S1. Genomic information obtained from NCBI (http://www.ncbi.nlm.nih.gov/), BEEBASE (http://hymenopteragenome.org/beebase), and FLYBASE (www.flybase.org) for the remaining directional and balancing outlier SNP loci as detected by two and one F_{ST}-based method (*P*-values ≤ 0.05), and for SNP loci that were associated with at least one environmental variable identified by the spatial analysis method (matSAM)

Table S2. Environmental variables associated with SNP loci (CI ≥ 99% with Bonferroni correction), as detected by the spatial analysis method (matSAM). Variables are annual, monthly and seasonal. Altitude and land cover are not included in this table as they were not significantly associated with any locus after Bonferroni correction

Table S3. Accession numbers of genes marked by SNPs that were detected by at least one of the five methods employed listed in Tables 1 and S1. Loci exhibiting the strongest signal of selection (see Table 1) are marked in bold

Figure S1. Individual bar charts indicate allele frequencies of mtDNA (M-lineage haplotypes), the strongest directional outliers (from ahb1245 to est11018), the strongest balancing outlier (AMB-00963630), and a representative neutral locus, in each of the 23 sampling sites. For each plot, sampling sites from left to right are AT1 to AT8, CT1 to CT9, and MT1 to MT6

Figure S2. Graphical representation of the five groupings used in the AMOVA (see Table 2). Letters "a)" to "e)" in this Figure correspond to the letters "a)" to "e)" in the footnotes of Table 2

Figure S3. Population structure and admixture levels obtained with the software STRUCTURE based on 1075 SNP loci. Each individual is represented by a bar, which is partitioned into K colored segments that represent the individual's estimated membership proportions in K clusters. Black lines separate individuals of the focal subspecies *A. m. iberiensis* (AT1 to MT6) and of the reference subspecies *A. m. intermissa, A. m. ligustica*, and *A. m. carnica*. The number of ancestral clusters (K) was estimated using the admixture ancestry and correlated allele frequency models with the

unsupervised option. STRUCTURE was set up for 750,000 Markov chain Monte Carlo iterations after an initial burn-in of 250,000. Over 20 independent runs for each K (from 2 to 4) 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)

Figure S4. Principal component analysis (PCA) based on 1075 SNP loci. PCA was performed on a normalized matrix of individuals *versus* SNP loci. As PCA is sensitive to missing data, genotypes were imputed for missing values using the mean allele frequency through the function ScaleGen available in ADEGENET. Principal components and variances were calculated from the singular value decomposition. PC1 separates *A. m. iberiensis* from the two eastern European beekeepers-favorite subspecies *A. m. ligustica* and *A. m. carnica* whereas PC2 separates the northern African *A. m. intermissa*. PCA 1 and PCA2 explain 48.1 % and 5.6 % of the variance, respectively.

Figure S5. a) LOSITAN, b) ARLEQUIN, c) BAYESFST, and d) BAYESCAN plots showing loci under selection. Solid and dashed lines represent 95 % and 99 % confidence intervals, respectively. The middle line in a) and b) depicts the median value. Outlier directional (upper) and balancing (lower) SNP loci with *P-value* ≤ 0.005 and posterior probability ≥ 0.99 and are labeled in each plot

Figure captions

Fig. 1 Map of the Iberian Peninsula showing the centroids of the sampling sites, sample size per site, and site codes. The number of apiaries sampled per site varied between 8 and 13, with most sites having 10

Fig. 2 Physical map of the 16 honey bee linkage groups (LG1 to LG16) showing the genomic positions of the 69 outlier SNP loci detected by at least one F_{ST}-based method and that are associated with at least one environmental variable. The 34 outliers detected by at least three F_{ST}-based methods are marked in bold (Table 1) whereas those detected by two or fewer F_{ST}-based methods (Table S1) are marked in gray. The 10 strongest candidates for selection (Table 1) are marked with an asterisk. Three SNPs have no position assigned yet. The map was depicted from the honey bee genome sequence available at http://www.ncbi.nlm.nih.gov/projects/mapview using the Map Viewer tool

Fig. 1

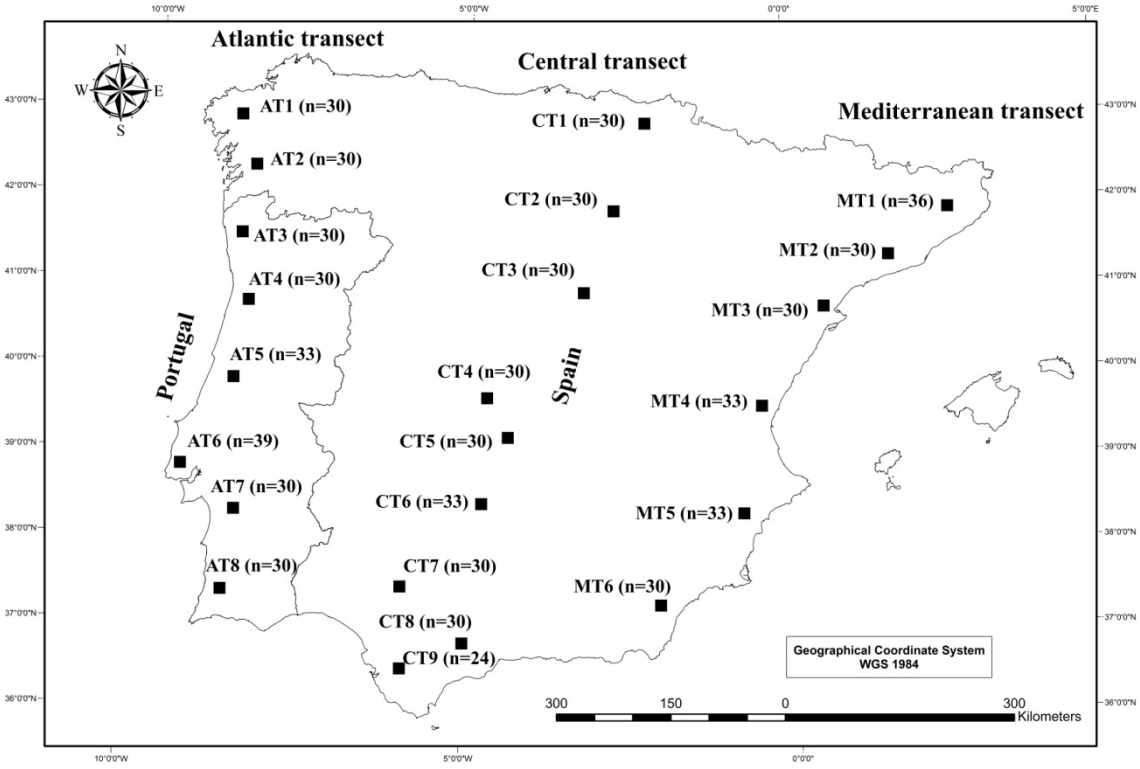
Atlantic transect
Central transect
Mediterranean transect
AT1 (n=30)
AT2 (n=30)
AT3 (n=30)
AT4 (n=30)
AT5 (n=33)
AT6 (n=39)
AT7 (n=30)
AT8 (n=30)
CT1 (n=30)
CT2 (n=30)
CT3 (n=30)
CT4 (n=30)
CT5 (n=30)
CT6 (n=33)
CT7 (n=30)
CT8 (n=30)
CT9 (n=24)
MT1 (n=36)
MT2 (n=30)
MT3 (n=30)
MT4 (n=33)
MT5 (n=33)
MT6 (n=30)
Portugal
Spain
N
S
W
E
Geographical Coordinate System
WGS 1984
300
150
0
300
Kilometers
10°0'0"W
5°0'0"W
0°0'0"
5°0'0"E
43°0'0"N
42°0'0"N
41°0'0"N
40°0'0"N
39°0'0"N
38°0'0"N
37°0'0"N
36°0'0"N

Fig. 2

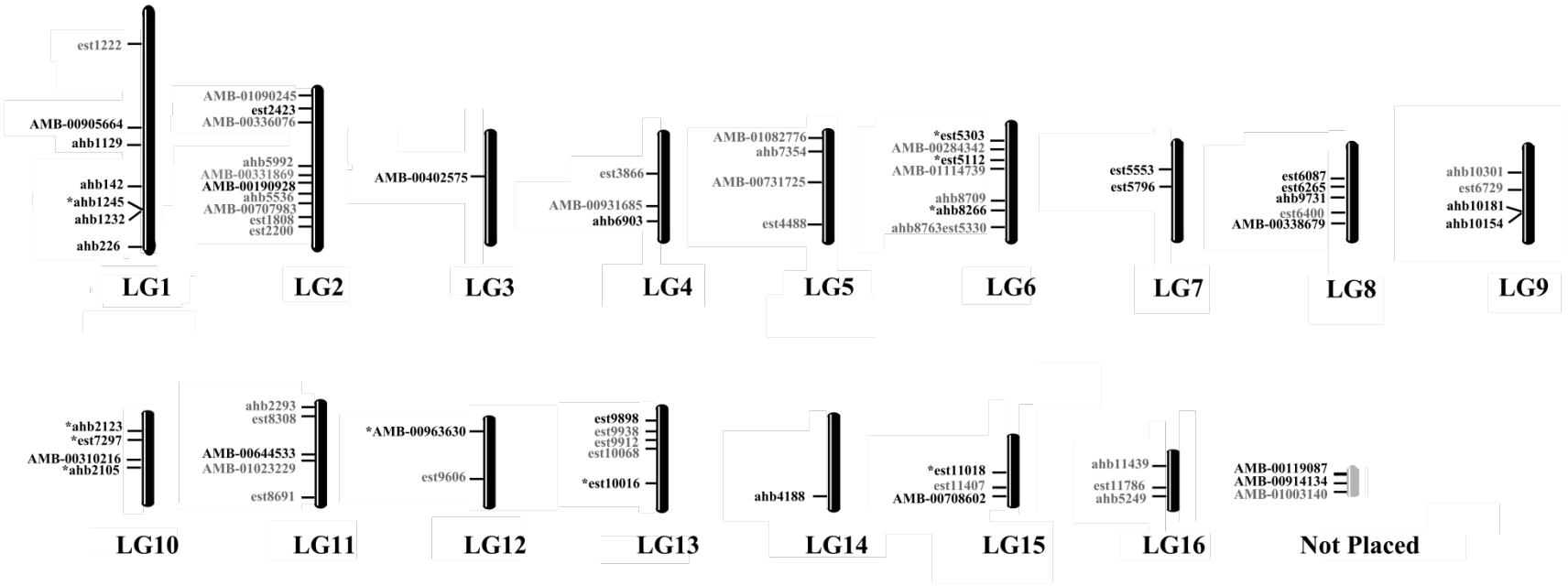
est1222
AMB-00905664
ahb1129
ahb142
*ahb1245
ahb1232
ahb226
LG1
AMB-01090245
est2423
AMB-00336076
ahb5992
AMB-00331869
AMB-00190928
ahb5536
AMB-00707983
est1808
est2200
LG2
AMB-00402575
LG3
est3866
AMB-00931685
ahb6903
LG4
AMB-01082776
ahb7354
AMB-00731725
est4488
LG5
*est5303
AMB-00284342
*est5112
AMB-01114739
ahb8709
*ahb8266
ahb8763est5330
LG6
est5553
est5796
LG7
est6087
est6265
ahb9731
est6400
AMB-00338679
LG8
ahb10301
est6729
ahb10181
ahb10154
LG9
*ahb2123
*est7297
AMB-00310216
*ahb2105
LG10
ahb2293
est8308
AMB-00644533
AMB-01023229
est8691
LG11
*AMB-00963630
est9606
LG12
est9898
est9938
est9912
est10068
*est10016
LG13
ahb4188
LG14
*est11018
est11407
AMB-00708602
LG15
ahb11439
est11786
ahb5249
LG16
AMB-00119087
AMB-00914134
AMB-01003140
Not Placed

Table 1 Outlier SNP loci as detected by four and three F_{ST}-based methods (BAYESFST, ARLEQUIN, and LOSITAN with *P*-values ≤ 0.05; BAYESCAN with posterior probability ≥ 0.95) and that are associated with environmental variables identified by the spatial analysis method (matSAM; CI ≥ 99%). SNP loci exhibiting the strongest signal (*P*-values ≤ 0.005 and posterior probability ≥ 0.99) are marked in bold. Genomic information for the SNP loci listed was obtained from NCBI (http://www.ncbi.nlm.nih.gov/), BEEBASE (http://hymenopteragenome.org/beebase), and FLYBASE (www.flybase.org)

SNP code	Linkage group	Position	Gene product	Putative function	SNP location	matSAM
Loci detected by BAYESFST, ARLEQUIN, LOSITAN, and BAYESCAN						
Directional						
ahb1245	1	25131116	Gst-mic2 - microsomal glutathione S-transferase 2	Metabolism	Intron	Long, Prec, Tmin, Ins
est5302	6	3199025	UDP-glucosyltransferase (UDP-glucosyltransferase 35b, Ugt35b) [§]	Metabolism	Exon	Lat, Prec, Tmean, Tmax, Cld
est5112	6	6341643	Vha16 - Vacuolar H+ ATP synthase 16 kDa proteolipid subunit (Vacuolar H^+ATPase subunit 16-1, Vha16-1)[§]	Transport	3'-UTR	
ahb8266	6	13484985	Teneurin 3 – like isoform 1 (Tenascin major, Ten-m)[§]	Structural	Intron	Lat, Prec, Tmean, Tmax, Cld, Ins
ahb10181*	9	7162173	Hypothetical protein LOC726750/Hydrocephalus- inducing protein-like	Unknown/Structural	4110/1297	
ahb2123	10	2249429	Choline transporter-like protein 1-like	Transport	3'-UTR	Long, Lat, Prec, Ins
est7297	10	3151832	15-hydroxyprostaglandin dehydrogenase [NAD+]-like (Photoreceptor dehydrogenase, Pdh)[§]	Metabolism	Intron	Long, Lat, Prec, Ins
ahb2105	10	6427742	Hypothetical protein LOC100577401	Unknown	Exon	Long, Prec, Rh
AMB-00644533	11	8036193	bs-Serum response factor homolog (blistered, bs)[§]	Regulation	Intron	Long, Prec, Ins
est9898	13	1328979	GTP-binding protein CG1354 isoform 1	Signaling	Exon	Lat, Prec, Tmean, Tmax, Ins

est10016	13	9554011	Cytochrome P450-CYP6AS7 (Cyp6a14)[§]	Metabolism	Exon	Lat, Prec, Tmean, Tmax, Cld, Ins
est11018	15	5299405	NimC2-nimrod C2 (nimrod C2, nimC2)[§]	Immunity	5'-UTR	
Balancing						
ahb142	1	21550552	Sema 1-Semaphorin 1A (Sema-1a)[§]	Structural	Intron	
ahb6903	4	9847583	Dscam-Down syndrome cell adhesion molecule (Down syndrome cell adhesion molecule, Dscam)[§]	Immunity	Intron	
AMB-00963630*	12	1214828	Hypothetical protein LOC100576488/Collagen alpha-2(IX) chain-like	Unknown/Structural	18439/25703	
AMB-00708602	15	8856798	Cubilin-like	Signaling	Intron	
AMB-00914134***	0	78294	Protein lin10-like (X11Lβ)[§]	Regulation	6219	
Loci detected by BAYESFST, ARLEQUIN, and LOSITAN						
Directional						
AMB-00905664	1	15234278	Hypothetical protein LOC100578389	Unknown	Intron	
ahb1232*	1	25169032	Notum pectinacetylesterase homolog (Notum)[§]/ NMDA kainate 2 sensitive receptor	Signaling/Signaling	21596/10461	
ahb226**	1	29654719				
est2423	2	2983426	Retinol dehydrogenase 11-like	Metabolism	Exon	Ins
AMB-00190928	2	7469407	Ubc-E2H-Ubiquitin-conjugating enzyme E2 H (UbcE2H)[§]	Signaling	Intron	Lat, Prec, Ins
AMB-00402575	3	5547433	Protein outspread-like (outspread, osp)[§]	Regulation	Intron	

est5553	7	4067968	Aldh-Aldehyde dehydrogenase isoform 1 (Aldehyde dehydrogenase, Aldh)[§]	Metabolism	5'-UTR	Long
est6087	8	5157656	Rfabg-Retinoid and fatty acid-binding glycoprotein isoform1 (Retinoid- and fatty acid-binding glycoprotein, Rfabg)[§]	Transport	Exon	
ahb9731	8	8266264	Hypothetical protein LOC411273	Unknown	Exon	Lat, Prec, Ins
ahb10154	9	7170480	Hydrocephalus- inducing protein-like	Structural	Exon	Lat, Prec, Tmean, Ins
AMB-00119087**	0	30896				Long, Prec
Balancing						
ahb1129	1	18156863	Hypothetical protein LOC413562	Regulation	Intron	
est5796	7	5788241	Hypothetical LOC100578906	Unknown	5'-UTR	
est6265	8	6383127	PHD finger and CXXC domain-containing protein CG17446-like isoform1 (Cfp1)[§]	Regulation	3'-UTR	
AMB-00338679	8	11725179	Midasin-like	Structural	Intron	
AMB-00310216***	10	5583063	5 HT2 beta-Serotonin receptor	Signaling	22473	
ahb4188	14	9203355	Blop-blue-sensitive opsin (Rhodopsin, Rh5)[§]	Signaling	Exon	

*SNP located between two genes (or putative genes). Function of both genes and physical distances (bp) to the 3' or 5' ends are indicated
**SNP located far away from genes (> 167.8 kb).
***SNP located close to a gene. Physical distance (bp) to the 3' or 5' end of the gene is indicated in the column "SNP location"
[§]Names and/or symbols within parentheses correspond to orthologous genes of *Drosophila melanogaster* as in FLYBASE
Lat – latitude; Long – longitude; Prec – precipitation; Tmin – minimum temperature; Tmean – mean temperature; Tmax – maximum temperature; Ins – insolation; Cld – cloud cover; Rh – relative humidity

Table 2 Variance components (%) of the three-level AMOVAs obtained with the 309 neutral loci and the nine strongest directional outliers for different grouping combinations (all *P-values* < 0.0001). Grouping by transects and mtDNA lineage were the main criteria (see Fig. S2 for a graphical visualization of the five groupings)

Grouping criteria	Nº of Groups	Within sites		Among sites		Among groups	
		Neutral	Outlier	Neutral	Outlier	Neutral	Outlier
By mtDNA lineage[a)]	2	95.75	80.14	3.35	11.56	0.89	8.29
By transect[b)]	3	96.01	81.24	3.24	8.05	0.75	10.71
By AT & mtDNA lineage[c)]	3	95.92	80.70	3.05	6.59	1.03	12.70
By AT & mtDNA lineage[d)]	3	95.95	81.32	3.09	8.17	0.96	10.51
By transect & mtDNA lineage[e)]	6	96.05	82.19	2.89	5.53	1.06	12.28

[a)] Group 1: all M-lineage sites across transects; Group 2: all A-lineage sites across transects

[b)] Group 1: all AT sites; Group 2: all CT sites; Group 3: all MT sites

[c)] Group 1: all AT sites; Group 2: M-lineage sites of CT and MT; Group 3: A-lineage sites of CT and MT

[d)] Group 1: only A-lineage AT sites; Group 2: M-lineage sites of AT, CT and MT; Group 3: A-lineage sites of CT and MT

[e)] Group 1: M-lineage site of AT; Group 2: A-lineage sites of AT; Group 3: M-lineage sites of CT; Group 4: A-lineage sites of CT; Group 5: M-lineage sites of MT; Group 6: A-lineage sites of MT

Table 3 Best linear regression models between the nine strongest directional outliers (see Table 1) and land cover at the site scale

SNP code	Land cover classes[1]	r	r^2	*P*-value
ahb1245	None	-	-	-
ahb2123	None	-	-	-
est7297	None	-	-	-
ahb2105	222	0.4330	0.1875	0.03902
est11018	222	0.6415	0.4116	0.00097
est5112	211	0.5846	0.3417	0.00339
est5302	211+212+222+223+241	0.5391	0.2907	0.00794
ahb8266	211+212+222+223+241+242	0.5567	0.3099	0.00580
est10016	211+212+222+223+241+242+244	0.6591	0.4344	0.00062

[1]Corine land cover 2006 codes: (211) non-irrigated arable land, (212) permanently irrigated land, (222) fruit trees and berry plantations, (223) olive groves, (241) annual crops associated with permanent crops, (242) complex cultivation patterns, and (244) agro-forestry areas. For each site, land cover classes were considered individually and combined within a 20 km radius circle. Regression analyses tested for associations between allele frequencies and the land cover proportion for the 23 sites.