Unraveling the local adaptation and population structure of Iberian honey bee (*Apis mellifera iberiensis*)

Julio César Chávez-Galarza[1,2], Filipe Oliveira Costa[1], Maria Alice Pinto[2]

[1]Centre of Environmental and Molecular Biology (CBMA), University of Minho

[2]Centre of Mountain Research (CIMO), Polytechnic Institute of Bragança

School of Sciences
University of Minho

Introduction

The Iberian Peninsula is considered as one of the most important Pleistocene glacial refugia in the European continent. Several phylogeographical studies of flora and fauna describe the Iberian Peninsula as a cradle of genetic differentiation and species repository, explained by its diverse habitats with varied microclimates, which have influenced the demographic processes and local adaptation of many species, and Iberian honey bee is no exception. Several studies have been carried out in the Iberian honey bee to explain its origin, suggesting two possible hypotheses: a primary intergradation process based on morphology and allozymes (Ruttner 1988, Smith and Glenn 1995), or a secondary contact process based on mtDNA (Garnery *et al.* 1992) supported by an abrupt southwestern-northeastern cline formed by two divergent lineages. Surveys with microsatellites support neither hypothesis (Franck *et al.* 1998). Currently, population genomics has made possible the study of local adaptation and demographic processes by using genome-wide scan approaches and single nucleotide polymorphisms (SNPs) are seemingly the most appropriate marker for that endeavor. The main objective of this work is to unravel the evolutionary history of the Iberian honey bee using SNPs.

Sampling

A total of 711 Iberian honey bee individuals (each representing a single colony) was collected in 2010 across three North-South transects in the Iberian Peninsula (Fig. 1).

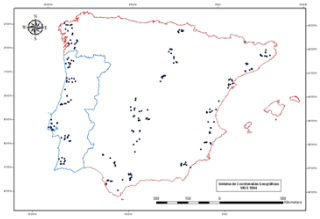

Fig. 1 Sampling transects. Each dot represents an apiary. Three colonies were sampled in each apiary.

Analysis of selection

Candidate loci under selection were identified using four F_{st}-based outlier approaches: LOSITAN (Antão *et al.* 2008), ARLEQUIN (Excoffier *et al.* 2009), BAYESFST (Beaumont and Balding 2004), and BAYESCAN (Foll and Gaggiotti 2008). In addition to F_{st}-based outlier approaches, univariate logistic regression analysis was performed by the matSAM Program (Joost *et al.* 2007) to determine the degree of association between the frequencies of each allele and the values of the environmental variables.

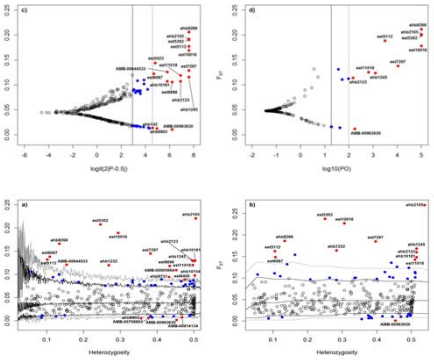

Fig. 2 (a) LOSITAN, (b) ARLEQUIN, (c) BAYESFST, and (d) BAYESCAN plots showing loci under selection. Solid and dashed lines represent 95% and 99% confidence intervals, respectively. The middle line in (a) and (b) depicts the median value. Outlier directional (upper) and balancing (lower) SNP loci with *P-value*<0.005 and posterior probability>0.99 and are labeled in each plot (red).

Genotyping

Individuals were genotyped for a panel of 1536 SNP's with Illumina BeadStation 500G using a custom Oligo Pool Assay. Individuals were scored using Illumina's Genome Studio software. The final number of loci was 383, after removing monomorphic loci (cutoff 2%) and failed genotype calls.

Environmental data

Environmental data were obtained for each apiary. This dataset consisted of Altitude (WorldClim database), Precipitation, Minimum temperature, Mean temperature, Maximum temperature and Cloud clover (Climatic Research Unit), Relative humidity and Insolation (OPENEI), and Land cover (European Environment Agency). All climatic data were extracted yearly, seasonally and monthly.

Analysis of population structure

Fst-pairwise values for each sampling sites was obtained with ARLEQUIN (Excoffier *et al.* 2009). Population structure was inferred using the Bayesian- clustering, multivariate and spatial analysis. Bayesian methods was performed using STRUCTURE (Pritchard et al. 2000), multivariate approaches using Principal Components (PCA), Principal Coordinates (PCoA), Discriminant analysis of Principal Components (DAPC) implemented in ADEGENET package (Jombart *et al* 2008) and spatial analysis with TESS (François and Durand 2010) and spatial PCA (sPCA) (Jombart *et al* 2008). STRUCTURE model: admixture, correlated allele frequencies. TESS model: BYM with linear trend degree. Run parameters: 250000 burnin steps, 750000 MCMC iterations.

Results

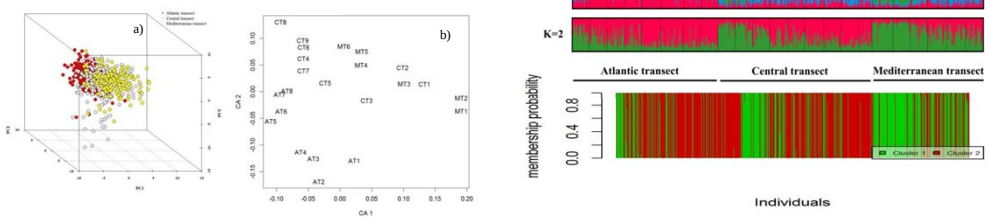

Fig.3 (a) PCA of individuals, based on alleles (b) PCoA based on Fst-pairwise of sampling sites (c) Proportions of ancestry inferred by STRUCTURE software for each individual (represented by vertical lines), when the nº of groups is K=2 to 4, and (d) DAPC of individuals, when the nº of groups is K=2

•Over 74 loci under selection were identified by Fst-based outlier approaches (Fig 2). Analysis with MatSAM identified 33 loci associated to environmental variables. Precipitation was the environmental variable with more associations with outlier SNP (Table 1).

• Outlier SNPs were located in the 16 linkage groups from honey bee genome marking genes with putative functions of signaling, structural, metabolism, regulation, transport, and immunity.

• Processes of xenobiotic detoxification, vision and immunity were represented by genes exhibiting a strongest signature of selection (Chávez-Galarza *et al.* 2003) .

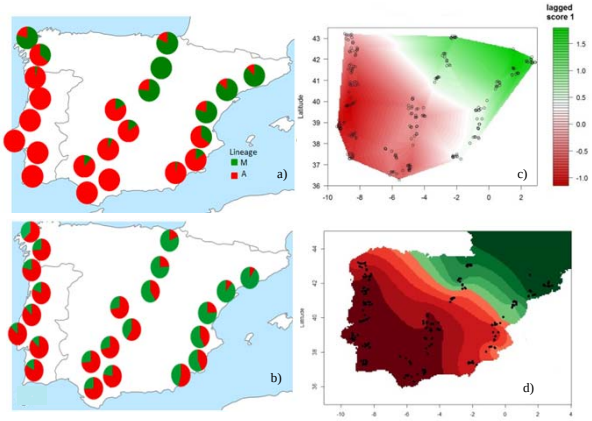

Fig 4 (a) Proportions of honey bees of western European (M) and African (A) lineage ancestry taken from Chávez – Galarza *et al.* unpublished, (b) The individual probabilities displayed in Fig 3(c) K=2 , (c) Spatial PCA representing the first principal component, and (d) The genetic clustering of individuals in the presence of a spatial Bayesian model inferred with TESS.

Table 1 Genomic information for the SNP loci with the strongest signal of selection related to xenobiotic detoxification, vision, immunity obtained from NCBI, BEEBASE, and FLYBASE.

SNP code	Gene product	Selection	matSAM
Detoxification			
ahb1245	Gst-mic2 - microsomal glutathione S-transferase 2	Directional	Long, Prec, Tmin, Ins
est5302	UDP-glucosyltransferase (UDP-glucosyltransferase 35b, Ugt35b) §	Directional	Lat, Prec, Tmean, Tmax, Cld
est10016	Cytochrome P450 - CYP6AS7 (Cyp6a14)§	Directional	Lat, Prec, Tmean, Cld, Ins
Vision			
ahb8266	Teneurin 3 – like isoform 1 (Tenascin major, Ten-m)§	Directional	Lat, Prec, Tmean, Tmax, Cld, Ins
est7297	15-hydroxyprostaglandin dehydrogenase [NAD+]-like (Photoreceptor dehydrogenase, Pdh)§	Directional	Long, Lat, Prec, Ins
est2423	Retinol dehydrogenase 11-like	Directional	Ins
ahb142	Sema 1 – Semaphorin 1A (Sema-1a)§	Balancing	
ahb4188	Blop - blue-sensitive opsin (Rhodopsin, Rh5)§	Balancing	
Immunity			
est11018	NimC2 - nimrod C2 (nimrod C2, nimC2)§	Directional	
ahb6903	Dscam - Down syndrome cell adhesion molecule (Down syndrome cell adhesion molecule, Dscam)§	Balancing	

§Names and/or symbols within parentheses correspond to orthologous genes of *Drosophila melanogaster* as in FLYBASE Longitude(Long), Latitude (Lat), precipitation (Prec), Minimum temperature (Tmin), Mean temperature (Tmean), Maximum temperature (Tmax), Cloud clover (Cld) and Insolation (Ins).

•PCA reveals low differentiation between individuals (Fig 3a), DAPC and STRUCTURE recover two substructures (K=2, Fig 3 c, d) and visualize a clinal pattern. While, PCoA visualizes more differentiation between sampling sites from Atlantic Transect with other transects, which is corroborated with STRUCTURE K=3 and 4 (Fig 3 b, c).

•The spatial representation of mtDNA lineage and proportion of ancestry obtained with STRUCTURE confirm a concordant northeastern-southwestern clinal pattern of variation (Fig 4 a, b), which contrasts with the absence of structure revealed by microsatellite markers (Franck et al. 1998).

•The spatial analysis performed with sPCA and TESS also found that clinal pattern, suggesting that the Iberian Peninsula has been a secundary contact zone (Fig 4 c,d), but it is necessary to carry out further analysis to corrobate this first results.

References

•Antão T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: A workbench to detect molecular adaptation based on Fst-outlier method. BMC Bioinformatics. 9: 323.
•Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology. 13: 969-980.
•Chávez-Galarza JC, 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 (SNPs). Molecular Ecology. *Accepted*.
•Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity. 103: 285-298.
•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.
•François O, Durand E. 2010. The state of the art- Spatially explicit Bayesian clustering models in population genetics. Molecular Ecology Resources. 10: 773-784.
•Foll M, Gaggiotti OE. 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics. 180: 977-993.
•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.
•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.
•Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155: 945–959.
•Ruttner F. 1988. Biogeography and taxonomy of honey bees. Springer, Berlin.
•Smith DR, Glenn TC. 1995. Allozyme polymorphisms in Spanish honeybees (*A.m. iberica*). Journal of Heredity. 86: 12-16.

Acknowlegdments

We owe special thanks to numerous beekeepers from Spain and Portugal for their valuable help in obtaining samples. DNA extractions and SNP genotyping was performed by Colette Abbey, who was deeply committed to this project. Thanks to Dr Clare Gill for providing the facilities for SNPs genotyping and for assistance with analysis. Special thanks to Dora Henriques for her assistance with genetic structure analysis. Julio Chávez-Galarza is supported by Fundação para a Ciência e Tecnologia through the scholarship SFRH/BD/68682/2010. This research was funded by Fundação para a Ciência e Tecnologia and COMPETE/QREN/EU through the project PTDC/BIA-BEC/098402/2008.