Effect of Linkage Disequilibrium on inferences of population Structure and Introgression of Iberian and Black Honey Bees

P6.2
00351932805831
jchavez@ipb.pt

Julio Chávez-Galarza, Dora Henriques, Per Kryger, Pilar de la Rúa, J. Spencer Johnston, José Rufino, M. Alice Pinto

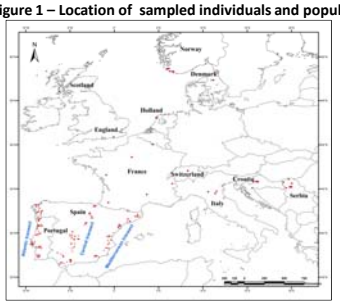
Introduction

Identification of population structure, a primary goal in population genetics, is easily performed because there is a number of methods available, implemented by user-friendly software packages. However, the user must be cautious when inferring population structure because spurious results may be obtained when there is strong linkage disequilibrium. With recent development of high-density SNPs we have now more power to interrogate the honey bee genome. However, the greater the number of loci genotyped the greater the chance of scoring loci that are linked. In addition, events such as population bottleneck, small effective population size, genetic drift, and admixture may also generate strong linkage disequilibrium. According to Kaeuffer et al. (2007), correlation r_{LD} is the best way to deal with linkage disequilibrium. These authors recommend removing loci with r_{LD} higher than 0.5 when inferring structure. In this study we used the GoldenGate Assay of Illumina to genotype over 1221 loci in individuals sampled from populations of *A.m. iberiensis* and *A.m. mellifera*. In this dataset we used the genetic distance between SNPs and r_{LD} to test the effect of linkage in the number of clusters and the introgression level inferred by the clustering method implemented in the software STRUCTURE.

Sampling

A total of 824 drone samples (each representing a single colony) was collected as following: 711 *A. m. iberiensis*, 77 *A. m. mellifera* (France, Holland, England, Scotland, Denmark, Norway), 17 *A. m. ligustica* (Italy) and 19 *A. m. carnica* (Serbia, Croatia) (Figure 1).

Figure 1 – Location of sampled individuals and populations



Genotyping

A panel of 1536 SNP's was genotyped using Illumina BeadStation 500G and a custom Oligo Pool Assay. Individuals were scored using Illumina's Genome Studio software.

A final dataset of 367 SNP loci for *A.m. iberiensis* and 1183 for *A. m. mellifera* was obtained after excluding monomorphic loci (2% cutoff) and non-calls.

The SNP position in the honey bee genome was identified using NCBI database.

Analyses

Linkage disequilibrium between pairs of loci was assessed using the correlation coefficient (r_{LD}) calculated by PLINK software (Purcell et al., 2007).

Population structure and introgression was inferred using STRUCTURE (Pritchard et al. 2000).

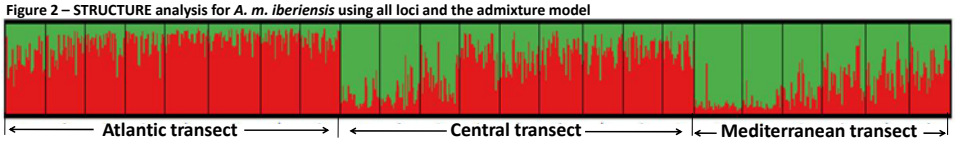
STRUCTURE was run using the following settings: admixture and linkage models, correlated allele frequency, 125000 burnin steps, 375000 MCMC iterations, number of clusters K=2 to 4, and 3 repetitions per K.

The analyses were run using four different loci combination: (1) all loci, (2) all loci after excluding those tightly linked (within 50000 bp distance), (3) all loci after excluding those exhibiting an r_{LD}>0.50 (as recommended by Kaeuffer et al. 2007), (4) only loci exhibiting an r_{LD}>0.50.

Results and Discussion

• STRUCTURE analyses of *A. m. iberiensis* identified two clusters (best K=2 according to Evano's test, data not shown) forming a southwestern-northeastern cline (Figure 2), a pattern which is congruent with mtDNA data (Cánovas et al. 2008). Introgression levels of *A. m. carnica*/*A. m. ligustica* into *A. m. iberiensis* were negligible (data not shown).

Figure 2 – STRUCTURE analysis for *A. m. iberiensis* using all loci and the admixture model

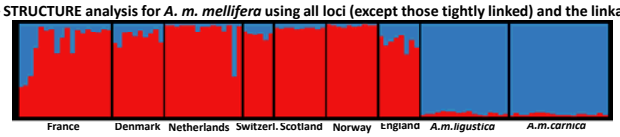


• The assignment probabilities (Q) of *A. m. iberiensis* generated by admixture and linkage models and using the different loci combinations (all loci with and without tightly linked SNPs).were very similar (Figure 3).

• Linkage disequilibrium may increase the probability of detecting spurious clustering (Falush et al. 2003; Kaeuffer et al. 2007). To deal with this potential clustering bias the authors of STRUCTURE recommend removing tightly linked loci (<1 cM) whereas Kaeuffer et al. (2007) suggest using the correlation r_{LD}. These authors recommend removing loci with r_{LD}>0.5 when inferring structure. In this study, r_{LD} values obtained for *A.m. iberiensis* were < 0.5, in contrast with those of *A.m. mellifera* (Table 1).

• Structure analyses revealed varying levels of *A. m. carnica*/*A. m. ligustica* introgression into *A. m. mellifera* across its geographical distribution (Figure 4).

Figure 4 – STRUCTURE analysis for *A. m. mellifera* using all loci (except those tightly linked) and the linkage model



•The average values of Q for the *A. m. mellifera* population, which can be interpreted in this case as the introgression level of *A. m. ligustica*/*A. m. carnica into A. m. mellifera,* generated by different models and loci combinations are shown in Figure 5. While there are virtually no differences between admixture and linkage models, the proportion of introgression varies across loci combinations. Exclusion of loci with r_{LD}>0.5 yields a lower level of introgression whereas inference using only loci with r_{LD}>0.5 generates higher introgression values.

•The proportion of *A. m. ligustica*/*A. m. carnica* introgression (Q) at the A. m. mellifera individual level generated using the linkage model for different loci combinations is shown in Figure 6. Introgression estimates is very similar across different loci combinations for *A.m. mellifera* individuals of greater purity. However, when the introgression level increases, the estimates vary with the loci combination used. In general, using only loci with r_{LD}>0.5 tends to overestimate the proportion of introgression.

Figure 3 – Assignment probabilities of belonging to the red cluster (Figure 2) generated by STRUCTURE using different models and loci combinations for *A. m. iberiensis*

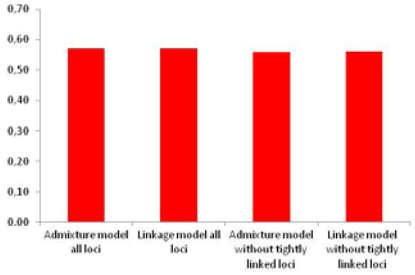


Table 1. Statistics of r_{LD} for *A. m. iberiensis* and *A. m. mellifera*

	Nº SNP loci	Average	Maximum	Minimum
A.m. iberiensis	**367**	**-0.00032**	**0.46**	**-0.19**
A. m. mellifera	**1183**	**0.17**	**1**	**-048**
***A. m. mellifera*[a]**	**367**	**0.036**	**0.86**	**-0.39**

[a]Values of r_{LD} were calculated using only the SNP loci shared with *A.m. iberiensis*

Figure 5 – Proportion of introgression generated by STRUCTURE for *A. m. mellifera*

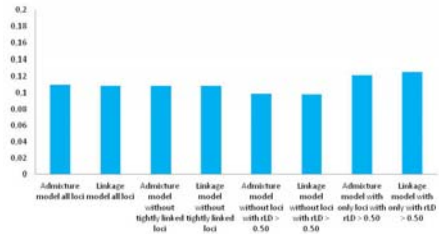
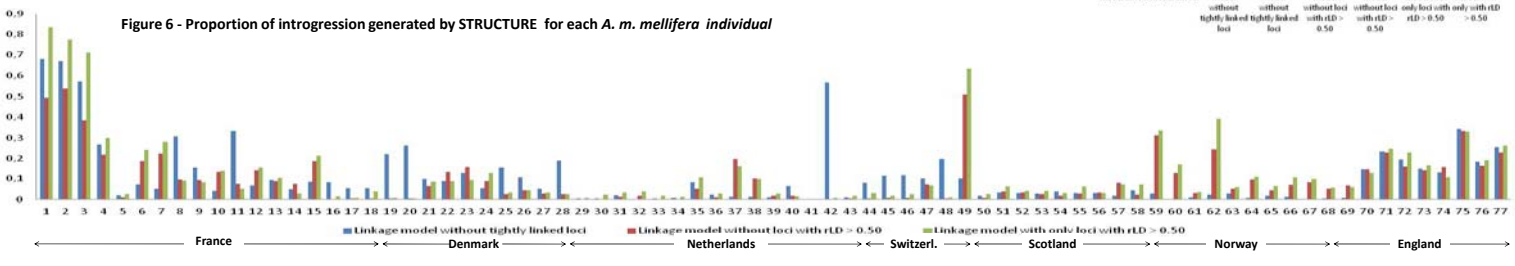


Figure 6 - Proportion of introgression generated by STRUCTURE for each *A. m. mellifera individual*



Literature

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

Chávez-Galarza, J., Johnston, J. S., Azevedo, J. C., Muñoz, I., de la Rúa, P., Patton, J. C., Costa, F., Pinto, M. A. 2011. Population structure as revealed by SNPs in the Iberian honey bee (Apis mellifera iberiensis). Trends in Biodiversity and Evolution. CIBIO. Vairão, Portugal.

Kaeuffer, R., Réale, D., Coltman, D.W., Pontier, D. 2007. Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. Heredity. 99: 374-380.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3): 559-575.

Acknowledgments

We owe special thanks to 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. Special thanks to Dr Clare Gill for providing the facilities for SNPs genotyping and for assistance with 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/099640/2008.