Introgression levels of the Italian and carniolan honey bee subspecies into the black honey bee: a comparison between microsatellite and single nucleotide polymorphism (SNP) markers

Dora Henriques[1], Laura Jara [2], Julio Chávez-Galarza[1,3], José Rufino Amaro[1], Pilar De la Rúa[2], Maria Alice Pinto[1]

[1]Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Apartado 1172, Bragança 3201-855, Portugal.
[2]Área de Biología Animal Departamento de Zoología y Antropología Física, Universidad de Murcia, Campus de Espinardo, Murcia 30010, Spain.
[3]Centro de Biologia Molecular e Ambiental (CBMA), Universidade do Minho, Campus de Gualtar, Braga 4710-057, Portugal.
Correspondence: Dora Henriques, CIMO, Instituto Politécnico de Bragança, Portugal. E-mail: dorasmh@gmail.com, Phone: 963677522

Introduction

Human activities have been shaping the distribution of honey bee subspecies in Europe. In fact, during the last decades there has been an extensive introduction of the beekeepers' favorite eastern European (lineage C) subspecies *A. m. ligustica* (Italian) and *A. m. carnica* (carniolan) into western Europe. Because of these introductions, there has been gene flow, and even replacement, of the native western European subspecies *A. m. mellifera* (black honey bee), which belongs to lineage M. Assessing levels of introgression is an important activity in breeding programs, especially when conservation of native subspecies is a major concern. Previous surveys of the *A. m. mellifera* populations estimated the introgression of lineage C into lineage M honey bees by using mtDNA and microsatellite markers (STR). Others markers, such as SNPs, have some advantages as they provide a genome wide coverage, higher quality data, and at the same time they are suitable for automatic and standardization in high throughput technologies. Previous studies indicate that the discriminatory power of SNPs to detect population structure is lower than microsatellites; about 100 SNPs are needed to provide the same power of 10-20 microsatellites. In this study we will compare introgression levels between microsatellites and SNPs in a black honey bee collection originating from several countries across western Europe.

Samples

A total of 77 *A. m. mellifera* individuals from France (18), Denmark (10), the Netherlands (15), Switzerland (6), Scotland (10), Norway (10) and England (8) were collected. Samples of *A. m. carnica* (19) from Croatia and Serbia, and *A. m. ligustica* (17), from Italy, were included as reference populations of C-lineage.

Genotyping

The 113 individuals were genotyped using SNPs and microsatellite. The 1536 SNP loci were scored using Illumina's BeadArray Technology and the Illumina GoldenGate® Assay. The genotyping was made using Illumina's Genome Studio software.

Two multiplex PCRs reactions with a total number of 12 microsatellite loci were done. PCR products were visualized using ABI-3730. Alleles were subsequently scored using GeneMapper v3.7 software.

Data sets

Of the 1536 a total of 1183 SNPs were available for analysis after removing monomorphic loci (cutoff 2%) and non-calls.

To obtain genomic position for both kind of markers, sequences were mapped to the honey bee Assembly 4.5 using BLAST in NCBI. Genomic position was ascertained using the Map Viewer tool available in NCBI.

The introgression levels were first estimated using both full datasets. Then, to have similar discriminatory power between the 12 microsatellite and SNPs, we used between 60 and 120 SNP loci from the initial 1183 SNP dataset by selecting SNPs located nearby the microsatellite loci.

Introgression analysis

Introgression of C-lineage was inferred for each black honey bee individual by running STRUCTURE 2.3.3 (Pritchard *et al.* 2000) for the 4 data sets. The probabilistic estimations of the admixture coefficient (Q) was generated by STRUCTURE using the following settings:

- :Admixture model
- Correlated allele frequency
- 250 000 burn in steps
- 750 000 MCMC iterations
- 20 runs
- K=2 clusters

CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) was used to compute the pairwise "symmetric similarity coefficient" between pairs of runs and to align the 20 runs. The means of the permuted results were plotted using DISTRUCT 1.1 (Rosenberg 2004).

Results and Discussion

STRUCTURE analysis performed using different dataset, show that introgression level are variable across the 77 individuals sampled in the black honey bee range (Fig. 1).

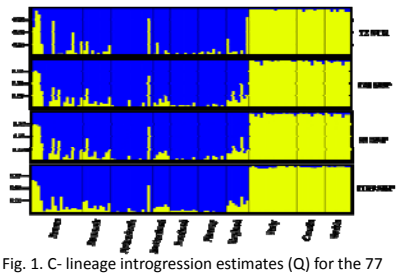

Fig. 1. C- lineage introgression estimates (Q) for the 77 individuals sampled in western Europe using 12 STRs, 120 SNPs, 60 SNPs and 1183 SNPs datasets

The populations that show individuals with more differentiated membership proportion between datasets are France and England (Fig 1; Table 2).

The outside membership proportion (Q) in the yellow cluster is very variable dataset to dataset. While in the 12 STR dataset the membership proportion vary between 0.009 and 0.951 in the 1183 SNP dataset vary between 0.001 and 0.6902 (Table 1).

Table 1- The outside membership proportion (Q) in the yellow cluster for each dataset

	12 STR	1183 SNP	120 SNP	60 SNP
High	0.951	0.6902	0.778	0.747
Low	0.009	0.001	0.003	0.005

Taking the 12STR dataset as reference, deviation between the datasets was calculate and the dataset that had the highest deviation was 1183 SNPs (0.08), while the 60 SNPs dataset display the lowest value (0.076) (Fig. 2).

While over than 42 individuals for each dataset have deviation values lower than 0.05, there are some individuals that have values of deviation higher than 0.20 (Table 2). The maximum deviation was found for an individual from England that have a deviation over 0.57 for all dataset (Table 2).

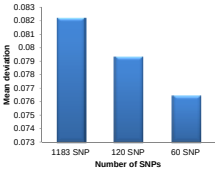

Fig.2. Mean introgression deviation from the Q value estimated, taking the 12 STR data set as reference for the 77 individuals

Table 2- Individuals with a deviation higher of membership proportion (Q) higher than 0.20.

	ind	1183 SNP	120 SNP	60 SNP
England	1	0.2167	0.2107	0.2096
England	2	0.2116		
England	3		0.2814	0.2372
England	6	0.2592	0.4162	0.3922
England	8	0.5702	0.5722	0.5852
France	2	0.2744		0.2077
France	3	0.3062		
France	8	0.4113	0.3193	0.2945
France	10	0.3034	0.3504	0.3684
France	15	0.4337	0.2932	0.224
Netherlands	14	0.2027		
Switzerland	2		0.2155	

At the population level the introgression estimates are more similar between datasets than at individual level. The highest deviation is also for the 1183 SNPs datasets with a value of 0.03 (Fig.3).

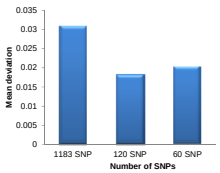

Fig. 3 Mean introgression deviation from the Q value estimated, taking the 12 STR data set as reference for the 7 *A. m. mellifera* populations

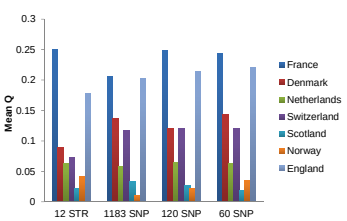

Fig. 4. Mean Q for the 7 *A.mellifera* populations, using different datasets

When analyzing the mean Q for the 7 *A. m. mellifera* populations for different datasets, it is possible observe that the highest differences are between the 12 STR and 1183 SNPs datasets. The populations that show higher differences are France, Denmark and Switzerland. While the 12 STR dataset has higher mean Q values for France, the 1183 SNPs dataset has higher values for Denmark and Switzerland populations (Fig. 4).

Conclusion

Results obtained by microsatellites are different from the results obtained by SNPs.

While there are more than 50% of individuals that have a deviation lower than 0.05, there are over 10 individuals that have values of deviation higher than 0.20, with an individual from England with a deviation higher than 0.5.

In spite of the differences found between all datasets, the 12STR are more similar with 120 SNP and 60 SNPs than with 1183 SNP dataset.

References

•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.
•Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155:945-959.
•Rosenberg NA. 2004. DISTRUCT.: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4:137-138.

Acknowledgements

We thank to Helena Ferreira to help us with some of the analysis. 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. 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.