



## HOW MANY SNPS ARE NEEDED TO PROVIDE AN ACCURATE ESTIMATE OF LINEAGE C INTROGRESSION INTO BLACK HONEY BEES?

Henriques D<sup>1</sup>, Chávez-Galarza J<sup>1</sup>, Johnston JS<sup>2</sup>, Rufino J<sup>1</sup>, Pinto MA<sup>1</sup>

<sup>1</sup>Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, *Campus de Sta. Apolónia*, Apartado 1172, 5301-855 Bragança, Portugal.

<sup>2</sup>Dept. of Entomology, Texas A&M University, College Station, Texas 77843-2475, USA.

Correspondence: Maria Alice Pinto, CIMO, Instituto Politécnico de Bragança, Portugal.

E-mail: [apinto@ipb.pt](mailto:apinto@ipb.pt)

Beekeeping activities, especially queen trading, have shaped the distribution of honey bees subspecies in Europe, which have resulted in extensive introductions of two lineage C subspecies, *A. m. ligustica* and *A. m. carnica*, into western Europe. As a consequence, replacement and gene flow between native and commercial populations have been occurring at varying levels across western European populations. Previous studies have monitored introgression by using microsatellite and PCR-RFLP markers. However, single nucleotide polymorphism (SNP) markers are more advantageous as they provide a genome-wide coverage and higher quality data. In addition, SNPs are suitable for automated high throughput technologies allowing genotyping of hundreds to thousands of loci in many individuals. Here we determined how many SNPs are needed to provide an accurate estimate of introgression of lineage C honey bees into black honey bees. We started the analysis using the full SNP dataset of over 1183 loci, which were genotyped using the GoldenGate assay and the GenomeStudio software of Illumina. Then, we examined 10 additional datasets that were generated by sequentially eliminating SNPs within a range of 1 to 10 cM. For each dataset we run the software STRUCTURE using the admixture model and correlated allele frequencies to estimate the hybrid proportion of each individual. The analyses showed similar introgression proportions produced by SNP combinations between 1183 and 81 loci, suggesting that a low number of loci is sufficient to produce accurate estimates.