How many SNPs are needed to provide an accurate estimate of lineage C introgression into black honey bees?

Henriques D¹, Chávez-Galarza J¹, Johnston JS², Rufino J¹, Pinto MA¹

¹Mountain Research Centre (CIMO), Politecnico Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal.
²Dept. 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

Introduction

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 honey bees 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 depart from a set of 1183 SNPs to determine the minimum number of SNPs that provide an estimate of introgression of lineage C honey bees into black honey bees as accurate as that generated by the 1183 loci.

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.

SNP genotyping

1536 SNP loci were scored using Illumina’s BeadArray Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay following manufacturer’s protocols. The 77 individuals were genotyped using Illumina’s Genome Studio software.

Table 1. The 14 data SNP data sets

<table>
<thead>
<tr>
<th>Data set</th>
<th>cm</th>
<th>Nº of SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Full</td>
<td>1183</td>
</tr>
<tr>
<td>2</td>
<td>&lt;1 cm</td>
<td>921</td>
</tr>
<tr>
<td>3</td>
<td>&lt;2 cm</td>
<td>743</td>
</tr>
<tr>
<td>4</td>
<td>&lt;3 cm</td>
<td>574</td>
</tr>
<tr>
<td>5</td>
<td>&lt;4 cm</td>
<td>446</td>
</tr>
<tr>
<td>6</td>
<td>&lt;5 cm</td>
<td>327</td>
</tr>
<tr>
<td>7</td>
<td>&lt;6 cm</td>
<td>238</td>
</tr>
<tr>
<td>8</td>
<td>&lt;7 cm</td>
<td>183</td>
</tr>
<tr>
<td>9</td>
<td>&lt;8 cm</td>
<td>138</td>
</tr>
<tr>
<td>10</td>
<td>&lt;9 cm</td>
<td>104</td>
</tr>
<tr>
<td>11</td>
<td>&lt;10 cm</td>
<td>81</td>
</tr>
<tr>
<td>12</td>
<td>&lt;11 cm</td>
<td>61</td>
</tr>
<tr>
<td>13</td>
<td>&lt;12 cm</td>
<td>40</td>
</tr>
<tr>
<td>14</td>
<td>&lt;13 cm</td>
<td>16</td>
</tr>
</tbody>
</table>

Results

Introgression of C-lineage was inferred for each black honey bee individual by running STRUCTURE 2.3.3 (Pritchard et al. 2000) for the 14 SNP data sets. The probabilistic estimations of the admixture coefficient (Q) was generated by STRUCTURE using the following settings: admixture model and 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).

Conclusions

1. The analysis at the population level suggests that 40 SNPs provide introgression estimates similar to those generated by the 1183-SNP data set.
2. However, a more detailed analysis at the individual level suggests that a larger number of SNPs is needed to provide introgression estimates as accurate as those generated by the full data set.
3. According to the preliminary analysis conducted herein the minimum number of SNPs is 238.

References


Acknowledgements

We thank Per Nygren, Bjørn Dahlbe, Lionel Garnery, Pilar de la Rúa, Raffaele Dall’Olio, Norman Carreck, Gabriele Soland-Reckeweg and Romée van der Zee for providing the honey bee samples. DNA extractions and SNP genotyping was performed by Colette Abbey, who was deeply committed to this project. 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/PO01/EU through the project PTDC/BAE-EC/098445/2008.