Genetic integrity of the Dark European honey bee (Apis mellifera mellifera) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data

Pinto, M Alice, Henriques, Dora, Chavez-Galarza, Julio, Kryger, Per, Garnery, Lionel, van der Zee, Romee, Dahle, Bjorn, Soland-Reckeweg, Gabriele, de la Rua, Pilar, Dall'Olio, Raffaele, Carreck, Norman L and Johnson, J Spencer (2014) Genetic integrity of the Dark European honey bee (Apis mellifera mellifera) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. Journal of Apicultural Research, 53 (2). pp. 269-278. ISSN 0021-8839

This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/48863/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk
Genetic integrity of the Dark European honey bee 
(*Apis mellifera mellifera*) from protected populations: 
a genome-wide assessment using SNPs and mtDNA 
sequence data

M Alice Pinto*1, Dora Henriques1, Julio Chávez-Galarza1, Per Kryger2, Lionel Garnery3, Romée van der Zee4, Bjørn Dahl5,6, Gabriele Soland-Reckeweg7, Pilar de la Rúa8, Raffaele Dall’Olio9, Norman L Carreck10,11 and J Spencer Johnston12

1Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal. 
2Århus University, Flakkebjerg, Forsøgsvej 1, 4200 Slagelse, Denmark. 
3Université de Versailles Saint-Quentin-en-Yvelines, 45 Ave. des Etats-Unis, 78035 Versailles, France. 
4Netherlands Centre Bee Research, NL-9014 CC Tersoal, Netherlands. 
5Norwegian Beekeepers Association, NO-2040 Klatfa, Norway. 
6Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway. 
7Apigenix, Gaicht 19, 2513 Twann, Switzerland. 
8Área de Biología Animal, Dpto.de Zoología y Antropología Física, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain. 
9Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Via di Saliceto 80, 40128 Bologna, Italy. 
10Laboratory of Apiculture and Social Insects, School of Life Sciences, University of Sussex, Falmer, Brighton, East Sussex, BN1 9QG, UK. 
11International Bee Research Association, 6, Centre Court, Main Avenue, Treforest, CF37 5YR, UK. 
12Dept. of Entomology, Texas A&M University, College Station, Texas 77843-2475, USA.

*Corresponding author: Email: apinto@ipb.pt

Summary

The recognition that the Dark European honey bee, *Apis mellifera mellifera*, is increasingly threatened in its native range has led to the establishment of conservation programmes and protected areas throughout western Europe. Previous molecular surveys showed that, despite management strategies to preserve the genetic integrity of *A. m. mellifera*, protected populations had a measurable component of their gene pool derived from commercial C-lineage honey bees. Here we used both sequence data from the tRNAleu-cox2 intergenic mtDNA region and a genome-wide scan, with over 1183 single nucleotide polymorphisms (SNPs), to assess genetic diversity and introgression levels in several protected populations of *A. m. mellifera*, which were then compared with samples collected from unprotected populations. MtDNA analysis of the protected populations revealed a single colony bearing a foreign haplotype, whereas SNPs showed varying levels of introgression ranging from virtually zero in Norway to about 14% in Denmark. Introgression overall was higher in unprotected (30%) than in protected populations (8%), and is reflected in larger SNP diversity levels of the former, although opposite diversity levels were observed for mtDNA. These results suggest that, despite controlled breeding, some protected populations still require adjustments to the management strategies to further purge foreign alleles, which can be identified by SNPs.
Integridad genética de la abeja negra de la miel (*Apis mellifera mellifera*) en poblaciones protegidas: evaluación amplia del genoma utilizando datos de SNPs y de la secuencia de ADN mitochondrial

**Resumen**

El reconocimiento de que la abeja negra de la miel, *Apis mellifera mellifera*, está cada vez más amenazada en su área de distribución natural, ha promovido el establecimiento de programas de conservación y de áreas de protección en toda Europa occidental. Los estudios moleculares previos mostraron que a pesar de las estrategias de gestión para preservar la integridad genética de *A. m. mellifera*, las poblaciones protegidas tenían un componente commensurable de su acervo genético derivado de abejas comerciales del linaje C. Aquí hemos utilizado datos tanto de la secuencia de la región intergénica trnA*-cox2 del ADNmt como del genoma nuclear, con más de 1.183 polimorfismos de nucleótido único (SNP), para evaluar la diversidad genética y los niveles de introgresión en varias poblaciones conservadas de *A. m. mellifera*, que luego se compararon con una muestra recolectada en poblaciones no protegidas. El análisis del ADNmt de las poblaciones conservadas reveló una única colonia con un haplotipo foráneo, mientras que los SNP mostraron niveles variables de introgresión que van desde prácticamente cero en Noruega a aproximadamente 14% en Dinamarca. La introgresión global fue mayor en las poblaciones sin protección (30%) que en las protegidas (8%), lo cual se refleja en mayores niveles de diversidad de SNP en las primeras, en contraste con los niveles de diversidad de ADNmt observados que fueron más bajos. Estos resultados sugieren que, a pesar de la cima controlada, algunas poblaciones protegidas todavía requieren ajustes en las estrategias de gestión para eliminar más alelos foráneos, que puedan ser identificados mediante el uso de SNPs.

**Keywords:** *Apis mellifera mellifera*, Dark European honey bee, introgression, conservation, diversity, SNPs, trnA*-cox2 intergenic region

**Introduction**

Honey bee diversity is the single most important legacy that we can leave to future generations of beekeepers, as it constitutes the raw material upon which natural and artificial selection operates. Loss of genetic diversity can only hamper honey bees’ adaptive response to modern beekeeping and increasingly demanding environmental conditions, and might be linked to worldwide colony declines (vanEngelsdorp and Meixner, 2010), although the latter is a matter of debate (Harpur et al., 2012, 2013; De la Rúa et al., 2013). Due to admixture of divergent honey bee subspecies (commonly promoted by beekeepers when using commercial foreign queens) there is an emerging movement to protect native genetic diversity (De la Rúa et al., 2009; Dietemann et al., 2009; Meixner et al., 2010). The organisation Societas Internationalis pro Conservatione Apis melliferae melliferae (SICAMM), established in 1995 for protecting the dark European honey bee, *Apis mellifera mellifera* (Ruttner et al., 1990) is an example of such a movement.

Europe has been a cradle for honey bee differentiation, which led to the evolution of 10 subspecies among the 30 currently recognized (Ruttner, 1988; Hepburn and Radloff, 1998; Engel, 1999; Sheppard and Meixner, 2003; Meixner et al., 2011), thereby representing a substantial component of total honey bee diversity. These 10 European subspecies have been grouped into two evolutionary lineages: the western European (lineage M) and the eastern European (lineage C).

Lineage M stretches across a broad territory ranging from northern Iberian Peninsula in the south to Scandinavia in the north, and from the British Isles in the west to the Ural Mountains in the east (Ruttner, 1988). This vast area is occupied by only two subspecies, although most of it is home to *A. m. mellifera*. Lineage C occurs in a smaller geographical area comprising the Apennine and Balkan peninsulas, bordered at the north by the Alps and at the south by Sicily and the west Aegean islands (Ruttner, 1988). Yet, this latter lineage comprises a larger number of subspecies, including the two most frequently used in commercial beekeeping worldwide: the Italian honey bee *A. m. ligustica* and the Carniolan honey bee *A. m. carnica*.

The native distribution of European honey bees has faced increasing challenges imposed by factors external to beekeeping activity (e.g. agrochemicals, habitat loss and fragmentation) and by beekeeping-related factors, of which accidental introduction of pests and pathogens and deliberate introduction of foreign queens are amongst the most detrimental. These factors may lead to losses of local genetic diversity through reductions in effective population size and through disruption of co-evolved gene complexes, as a consequence of matings with foreign subspecies eventually leading to introgressive hybridization (Muñoz et al., 2012). The Dark European honey bee is probably the European subspecies most threatened by the above human-mediated factors, among which introgression has a major role (Jensen et al., 2005; Soland-Reckeweg et al., 2009; Oleksa et al., 2011).
Emerging recognition of the importance of using native subspecies as a source of genetic material for sustainable beekeeping has led to establishment of protected areas across northern Europe aimed at conserving the genetic integrity of the Dark European honey bee (De la Rúa et al., 2009; Soland-Reckeweg et al., 2009; Meixner et al., 2010; Oleksa et al., 2011). In these protected areas selected breeding stocks are mated at isolated mating stations in order to prevent gene flow from undesired sources, mainly derived from foreign queens of C-lineage ancestry. Assessing levels of introgression in breeding stocks is an important activity in these programmes. Herein, we assessed diversity and introgression levels of A. m. mellifera honey bee colonies sampled from several protected populations in northern Europe. To that end, we analysed both mitochondrial DNA (mtDNA) sequence data of the tRNA<sub>leu</sub>-cox2 intergenic region and a genome-wide scan of 1183 polymorphic single nucleotide polymorphisms (SNPs) (Meixner et al., 2013), which represents the finest coverage, ever performed, of the nuclear genome of A. m. mellifera populations included in conservation programmes. We found that some protected populations still hold a significant component of C-lineage ancestry suggesting that management strategies of some conservation programmes need to be refined to achieve programme goals.

**Material and methods**

A total of 114 drone samples, each representing a single colony, were collected from randomly selected colonies between 2010 and 2011. Seventy-seven colonies were sampled in the native range of the M-lineage subspecies A. m. mellifera, including England (8), France (15), Belgium (3), Denmark (10), the Netherlands (15), Switzerland (6), Scotland (10), and Norway (10). The eight samples from England and five from France were collected from unprotected populations (hereafter named “unprotected group”). The remaining samples represent protected pure breeding populations (hereafter named “protected group”), which have mated on islands (Læsø, Denmark; Texel, The Netherlands; Colonsay, Scotland) or in isolated mating stations in the continent (France, Belgium, Switzerland, and Norway) maintained to preserve A. m. mellifera genetic identity (Fig. 1). A reference collection of 37 samples, representing C-lineage diversity (hereafter named “reference group”), was obtained in the native range of A. m. carnica and A. m. ligustica from Serbia (9), Croatia (11) and Italy (17), respectively. Samples were collected from within the hives and stored into absolute ethanol at -20ºC until molecular analysis.

**DNA extraction and mitochondrial DNA analysis**

Total DNA was extracted using a phenol-chloroform isooamyl alcohol (25:24:1) protocol (Sambrook et al., 1989) from the thorax of the 114 samples. MtDNA was analysed using the highly polymorphic tRNA<sub>leu</sub>-cox2 intergenic region, which was amplified using the primers E2 and H2 and the PCR reaction and conditions detailed by Garnery et al. (1993), with minor modifications. Following quantification in a routine agarose gel, PCR products were sent to Macrogen for direct sequencing in both directions. The sequences were manually checked for base calling and aligned with published sequences available in GenBank using MEGA version 5.03 (Tamura et al., 2011) to allow identification of haplotypes. The novel haplotypes and variants were named following the convention established earlier (Garnery et al., 1998) and recently reviewed for lineage M (Rortais et al., 2011).

**Single nucleotide polymorphism analysis**

A total of 1536 SNP loci were genotyped for the 114 drone samples using Illumina’s Bead Array Technology and the Illumina GoldenGate<sup>®</sup> allele-specific extension assay (Illumina, San Diego; CA, USA) with a custom Oligo Pool Assay (OPA), following manufacturer’s protocols. Further details about this highly multiplexed genotyping assay technique can be found in Shen et al. (2005). The OPA was modified from that described in Whitfield et al. (2006) by the replacement of 401 invariant SNPs from the expressed sequence tag (EST)-derived set with polymorphic genome-derived SNPs selected to produce a more uniform coverage of the genome (Chávez-Galarza et al., 2013). The modified OPA included 376 SNPs that were EST-derived and thus located within coding regions. The remaining 1160 were selected to be approximately equipollent without regard to position within or near coding regions. Genotype calling was performed using Illumina’s Genome Studio<sup>®</sup> Data Analysis software. For each sample, intensity clusters generated automatically by the software were manually
verified, and edited when necessary. SNPs with poorly separated clusters or low signals were excluded from the data set.

To obtain the genomic position, each SNP’s 100 bp flanking sequence was mapped to the Honey bee Assembly 4.5 using BLAST in NCBI (www.ncbi.nlm.nih.gov). Only SNPs that perfectly matched a unique position in Assembly 4.5 were retained. Genomic position was ascertained using the Map Viewer tool available in NCBI.

Genetic diversity was assessed for each SNP and each population using unbiased estimates of gene diversity (Nei, 1987) and allelic richness, a measure of the number of alleles independent of sample size (Petit et al., 1998). The mean number of alleles (Na), number of effective alleles (Ne) and unbiased diversity (Hs) were estimated using the admixture ancestry and correlated allele frequency models with the unsupervised option. The program was set up for 750,000 Markov chain Monte Carlo iterations after an initial burn-in of 250,000. Over 20 independent runs for each K (from 1 to 5) were performed to confirm consistency across runs. The output was exported into STRUCTURE HARVESTER v0.6.93 (Earl and Von-Holdt, 2012) and the estimation of the most probable K was calculated as described by Evanno et al. (2005). The Greedy algorithm, implemented in the software 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 for each K. The means of the permuted results were plotted using the software DISTRUCT 1.1 (Rosenberg, 2004).

Population structure was also examined using principal components analysis (PCA) implemented in the R package ADEGENET 1.3-7 (Jombart, 2008). PCA was performed on a normalized matrix of individuals versus SNP loci. As PCA is sensitive to missing data, genotypes were imputed for missing values using the mean allele frequency through the function ScaleGen available in ADEGENET. Principal components and variances were calculated from the singular value decomposition.

Results

Mitochondrial DNA

Sequence analysis of the trNAAsp-cox2 intergenic region of the 114 colonies produced a total of seven different haplotypes all belonging to lineages M and C (Table 1). Colonies of the reference group

Table 1. Number of M and C haplotypes per sampled location and group. Letters below haplotypes refer to variants. The protected and unprotected groups refer to colonies sampled from conservation and unprotected areas in the native range of A. m. mellifera, respectively. The reference group was sampled in the native range of A. m. ligustica (Italy) and A. m. carnica (Serbia and Croatia). Excepting for M8, C1, C1b, C2c-k, the remaining variants are novel (see sequencing data in Fig. S1 and GenBank under accession numbers KF274625 – KF274641).

Letter N denotes sample size.

<table>
<thead>
<tr>
<th>Location</th>
<th>N</th>
<th>M4</th>
<th>M4'</th>
<th>M7</th>
<th>M8</th>
<th>M64</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a</td>
<td>b</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>Protected</td>
<td></td>
<td>i</td>
<td>j</td>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
<td>a</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>15</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>6</td>
<td></td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>4</td>
<td>13</td>
<td>15</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Unprotected</td>
<td></td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>France</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>1</td>
<td>3</td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>17</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Serbia</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td></td>
<td>5</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Grand total</td>
<td>114</td>
<td>5</td>
<td>13</td>
<td>16</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

The individual-based Bayesian clustering algorithm implemented in STRUCTURE 2.3.3 (Pritchard et al., 2000) was employed to infer admixture proportions (Q) in population samples collected in the native range of A. m. mellifera. The number of ancestral clusters (K) was estimated using the admixture ancestry and correlated allele frequency models with the unsupervised option. The program was set up for 750,000 Markov chain Monte Carlo iterations after an initial burn-in of 250,000. Over 20 independent runs for each K (from 1 to 5) were performed to confirm consistency across runs. The output was
belonged to C-lineage, except five colonies from Italy that harboured a single novel variant of the M7 haplotype (M7a, accession number KF274639; Fig. S1). While colonies from Serbia and Croatia were predominantly of C2 ancestry, colonies from Italy were C1, although four and two variants were identified for both haplotypes, respectively. Colonies representing the *A. m. mellifera* protected group carried haplotypes belonging to M-lineage, except one single C2 colony from Belgium. The majority of these colonies (S1 out of 64; Table 1) harboured a single haplotype of M4 ancestry, although there were 13 variants (all novel) that differ from each other by six 1-2 bp indels, 14 transitions, of which two in cox2 region were non-synonymous, and three transversions (Fig. S1). Three additional haplotypes were detected in the remaining colonies: M8 (one colony from France), and the novel M4a’ (10 colonies from Scotland; accession number KF274638) and M64 (one colony from France, accession number KF274640). While maternal composition of the protected group was virtually of M-lineage ancestry (63 out of 64), the unprotected group exhibited a high (8 out of 13) proportion of C haplotypes, including the newly described C1h (accession number KF274641; Fig. S1) detected in a colony from France. In spite of this observation, of the three groups, diversity estimates were largest for the protected group (Table 2).

**Table 2.** Number of polymorphic SNP loci (from a total of 1183) and diversity estimates for mtDNA and SNPs per sampled location and group. The protected and unprotected groups refer to colonies sampled from conservation and unprotected areas in the native range of *A. m. mellifera*, respectively. The reference group was sampled in the native range of *A. m. ligustica* (Italy) and *A. m. carnica* (Serbia and Croatia). Polymorphic SNPs were defined by a cut-off criterion of 0.02 for the minor allele, as in Chávez-Galarza et al. (2013). Na represents the mean number of alleles per SNP, Ne the number of effective alleles, uh the unbiased diversity, and Rs the allelic richness. Standard errors are shown within parentheses.

<table>
<thead>
<tr>
<th>Location</th>
<th>Poly-morphic Private</th>
<th>Na MtDNA</th>
<th>SNPs</th>
<th>Ne MtDNA</th>
<th>SNPs</th>
<th>uh MtDNA</th>
<th>SNPs</th>
<th>Rs MtDNA</th>
<th>SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td>733</td>
<td>0</td>
<td>6</td>
<td>1.713(0.013)</td>
<td>4.167</td>
<td>1.325(0.009)</td>
<td>0.844</td>
<td>0.233(0.005)</td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td>279</td>
<td>0</td>
<td>3</td>
<td>1.234(0.012)</td>
<td>3.000</td>
<td>1.187(0.010)</td>
<td>1.000</td>
<td>-</td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td>738</td>
<td>0</td>
<td>1</td>
<td>1.624(0.014)</td>
<td>1.000</td>
<td>1.296(0.009)</td>
<td>0.000</td>
<td>0.211(0.006)</td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
<td>755</td>
<td>1</td>
<td>3</td>
<td>1.638(0.014)</td>
<td>1.991</td>
<td>1.217(0.008)</td>
<td>0.533</td>
<td>0.156(0.005)</td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
<td>575</td>
<td>0</td>
<td>2</td>
<td>1.486(0.015)</td>
<td>1.385</td>
<td>1.292(0.010)</td>
<td>0.333</td>
<td>0.211(0.007)</td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
<td>405</td>
<td>0</td>
<td>1</td>
<td>1.342(0.014)</td>
<td>1.000</td>
<td>1.179(0.009)</td>
<td>0.000</td>
<td>0.122(0.005)</td>
</tr>
<tr>
<td>Norway</td>
<td></td>
<td>332</td>
<td>0</td>
<td>4</td>
<td>1.281(0.013)</td>
<td>3.333</td>
<td>1.145(0.008)</td>
<td>0.778</td>
<td>0.099(0.005)</td>
</tr>
<tr>
<td>All group</td>
<td></td>
<td>1020</td>
<td>1</td>
<td>17</td>
<td>1.938(0.007)</td>
<td>7.262</td>
<td>1.258(0.008)</td>
<td>0.876</td>
<td>0.179(0.004)</td>
</tr>
<tr>
<td>Unprotected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td></td>
<td>827</td>
<td>0</td>
<td>3</td>
<td>1.699(0.013)</td>
<td>2.462</td>
<td>1.406(0.010)</td>
<td>0.679</td>
<td>0.280(0.006)</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td>817</td>
<td>11</td>
<td>4</td>
<td>1.691(0.013)</td>
<td>3.571</td>
<td>1.512(0.011)</td>
<td>0.900</td>
<td>0.361(0.07)</td>
</tr>
<tr>
<td>All group</td>
<td></td>
<td>1073</td>
<td>0</td>
<td>7</td>
<td>1.907(0.008)</td>
<td>5.121</td>
<td>1.555(0.009)</td>
<td>0.805</td>
<td>0.354(0.005)</td>
</tr>
<tr>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td>709</td>
<td>1</td>
<td>2</td>
<td>1.599(0.014)</td>
<td>1.710</td>
<td>1.273(0.009)</td>
<td>0.441</td>
<td>0.182(0.005)</td>
</tr>
<tr>
<td>Serbia</td>
<td></td>
<td>590</td>
<td>0</td>
<td>4</td>
<td>1.499(0.015)</td>
<td>2.077</td>
<td>1.262(0.009)</td>
<td>0.583</td>
<td>0.186(0.006)</td>
</tr>
<tr>
<td>Croatia</td>
<td></td>
<td>681</td>
<td>0</td>
<td>5</td>
<td>1.576(0.014)</td>
<td>3.270</td>
<td>1.267(0.009)</td>
<td>0.764</td>
<td>0.186(0.006)</td>
</tr>
<tr>
<td>All group</td>
<td></td>
<td>969</td>
<td>1</td>
<td>7</td>
<td>1.819(0.011)</td>
<td>4.319</td>
<td>1.297(0.009)</td>
<td>0.790</td>
<td>0.198(0.005)</td>
</tr>
<tr>
<td>All groups</td>
<td></td>
<td>1183</td>
<td>13</td>
<td>25</td>
<td>2.000(0.000)</td>
<td>12.101</td>
<td>1.686(0.008)</td>
<td>0.925</td>
<td>0.388(0.004)</td>
</tr>
</tbody>
</table>

**Single nucleotide polymorphisms**

**Genotyping success**

Of the 1536 SNP loci assayed, 353 were excluded from the data set for the following reasons: 124 exhibited poorly separated clusters or low signals, 167 were monomorphic (defined by a cut-off criterion of > 0.98 for the frequency of the most common allele, as in Chávez-Galarza et al., 2013) across all populations, 54 could not be placed in the sequenced honey bee genome, and eight had a double match in the sequenced genome. Accordingly, the final set of SNPs used in all subsequent analyses numbered 1183. Most samples (72 out of 114) exhibited a call rate higher than 99% whereas 38 and three were above 95% and 90%, respectively. One sample from Serbia was excluded from the data set because it had a call rate of 43.2%. The 1183 remaining useful SNPs were distributed across the 16 linkage groups (LG) ranging from 4.6 SNP/Mb (33 SNPs) in LG16 to 6.1 SNP/Mb (81 SNPs) in LG7 with an average of 5.4 SNP/Mb, thereby representing a fine coverage of the honey bee genome. The minor allele frequency (MAF) distribution at the 1183 SNP loci for the three different groups is detailed in Figs. S2, S3, and S4.
Genetic diversity

The number of polymorphic SNP loci and diversity measures per location and group are shown in Table 2. The majority of SNPs were variable across groups, suggesting that most genetic variation is shared among honey bees inhabiting (western and eastern) Europe. The number of private SNPs varied between zero and one in all locations except for the unprotected French sample, which carried 11 private SNPs. Apart from the mean number of alleles, all other diversity measures were highest in the unprotected group ($Ne = 1.555$, $uh = 0.354$, $Rs = 1.339$). In contrast, the protected group exhibited the lowest estimates of effective number of alleles ($Ne = 1.258$), unbiased diversity ($uh = 0.179$), and allelic richness ($Rs = 1.176$), with Norway displaying the lowest values ($Ne = 1.145$, $uh = 0.099$, $Rs = 1.094$) among all sampled locations.

Patterns of genetic variation

Patterns of genetic variation and admixture, inferred with STRUCTURE for a number of K clusters ranging from two to five, are shown in Fig. 2. For $K = 2$, which is the most likely number following the $\Delta K$ method (Evanno et al., 2005), colonies of the reference group formed one cluster (marked in yellow) with an average membership coefficient of $0.98 \pm 0.002$ (SE). Several colonies of the unprotected group revealed admixed ancestry with a membership proportion in the yellow cluster as high as 0.69 in France and 0.34 in England (Figs 2 and 3). In contrast, colonies of the protected group from Norway and Scotland showed the largest membership coefficients in the blue cluster, with average values of $0.99 \pm 0.003$ and $0.97 \pm 0.003$ (Fig. 3), respectively, reflecting virtually no introgression from the reference group. Colonies from the Netherlands displayed low introgression levels, as well, excepting for a single colony with probability of assignment to the yellow cluster of 0.59. The other representatives of the protected group displayed larger introgression levels with average membership proportions to the blue cluster varying from 0.86 $\pm$ 0.02, in Denmark, to 0.88 $\pm$ 0.01, in Switzerland (Fig. 3). As the $K$ cluster number increased ($K = 3$), the major change in genome partitioning was observed for three colonies (marked by letters b, c, and d in Figs. 2 and 3) of the unprotected French population, with a high probability of assignment ($0.87 \pm 0.07$) to the cluster marked in red (Fig. 2). Structure analysis using additional reference subspecies revealed that those three colonies shared a common ancestry with African subspecies (data not shown). Principal components analysis (PCA; Fig. 4) supports the patterns revealed by STRUCTURE. The first component (PC1) separated two
Fig. 3. Distribution of posterior mean estimates of membership proportion (Q), in the blue cluster of Fig. 2, for each individual of the protected and unprotected groups obtained with STRUCTURE, based on 1183 SNP loci, for K = 2. Average membership proportion ± SE, for each location, are indicated at the right side of the chart.

Fig. 4. Principal component analysis (PCA) based on the 1183 SNP loci. PC1 separates colonies sampled in eastern Europe (native range of A. m. carnica and A. m. ligustica) from colonies sampled in western Europe (native range of A. m. mellifera) whereas PC2 separates colonies of the unprotected group from France. The lower case letters (a-h) mark the same colonies in Figs. 2 and 4. PC1 and PC2 explain 42.2% and 6.1% of the variance, respectively.
major clusters formed by colonies of the reference eastern European group and colonies of the western European group whereas PC2 separated colonies of the unprotected group from France that exhibited the highest levels of introgression (Figs. 2 and 3). PCA indicates the non C-lineage origin of introgressed genes into those colonies. Both STRUCTURE and PCA revealed a close relationship among populations of Norway, Scotland, and the Netherlands.

Discussion

This study suggests that efforts to preserve A. m. mellifera have proved successful in most conservation programmes, although there is clear evidence that colonies of some protected populations still carry an important component of C-lineage ancestry. While mtDNA revealed a single C-lineage derived colony, SNPs showed varying levels of admixture across locations. Colonies from Norway and Scotland form the most homogenous and the “purest” populations with average membership proportions in the A. m. mellifera cluster higher than 0.99 and 0.97, respectively. Colonies from the Netherlands were also homogenous and showed high proportions of individual genotype memberships in the A. m. mellifera cluster, excepting for a single highly introgressed colony, which might represent a recent introduction event into the closed breeding protected population.

Colonies from France, Belgium, Switzerland and Denmark exhibited higher admixture proportions, although their genomes are mostly derived from A. m. mellifera (Q > 0.86). Earlier studies using microsatellites reported lower introgression proportions in the same populations of Switzerland (Soland-Reckeweg et al., 2009), Scotland and Denmark (Jensen et al., 2005). While this discrepancy could be explained by a sampling effect or a temporal change, it is also possible that the genotype-wide scan is capturing hidden introgression undetected by the microsatellite loci. While we cannot compare surveys that used different molecular markers, this study revealed that those populations still hold an important C-derived component suggesting that management strategies implemented in the conservation programmes have not been successful in purging all foreign alleles. This finding calls for adjustments in conservation strategies, which might involve a better control of matings, if introgression is still ongoing. Alternatively, a more thorough selection within each conservation populations may be necessary, while carefully observing the risk of further reducing native genetic diversity.

Levels of mtDNA diversity were higher in the protected group than in the reference group. This finding is consistent with previous studies that have reported higher maternal diversity in M than in C-lineage populations (Garnery et al., 1998; Franck et al., 2000; Jensen et al., 2005). This disparity has been attributed to the shorter intergenic sequence characteristic of C-lineage (it lacks the P element and it possesses a single Q element) offering less targets for site mutations and less possibilities of duplication/deletion of the Q element (Cornuet et al., 1991). Sequence analysis, performed in this study for each colony, allowed identification of variation that would have gone undetected by the popular PCR-RFLP method known as the DraI test (Garnery et al., 1993). In this study, the PCR-RFLP M4 pattern was the most frequent in both protected and unprotected groups, congruent with previous surveys of A. m. mellifera (Garnery et al., 1998; Jensen et al., 2005; Oleksa et al., 2011; Rortais et al., 2011). However, our sequence data distinguished 13 variants of the M4 pattern indicating that colonies from France, Belgium, the Netherlands, Switzerland, and Norway do not descend from a single maternal ancestral, as the DraI test (Garnery et al., 1993) would have suggested.

SNP diversity showed a different pattern: protected and reference groups exhibited similar diversity levels, although lower than those exhibited by the unprotected group. Given the admixed nature of the latter group, this was an anticipated result as admixture is a major mechanism for increasing genetic diversity in managed honey bee populations (Harpur et al., 2012). Genetic diversity is important at both population and colony level, and its decrease has been linked to recent honey bee declines in Europe and North America (vanEngelsdorp and Meinken, 2010). At the population level, genetic diversity is required for populations to evolve to cope with increasingly challenging environmental conditions (e.g. novel parasites, novel diseases, and pesticides). At the colony level, genetic diversity is essential to colony health (Tarpy, 2003; Seeley and Tarpy, 2007) and fitness (Page, 1980; Mattila and Seeley, 2007; Oldroyd and Fewell, 2007). Admixture may lead to increased genetic diversity, yet it may also compromise local adaptations by disrupting co-evolved gene complexes fine-tuned by natural selection over evolutionary time (De la Rúa et al., 2013). Accordingly, native honey bee subspecies represent reservoirs of unique combinations of genes and adaptations to local conditions that must be preserved and passed on to future generations of beekeepers.

Acknowledgments

We are deeply indebted to numerous people that contributed to this study, namely: Gilles Fert and Andrew Abrahams for providing honey bee samples; Danny Weaver for permission to use the OPA; Colette Abbey for DNA extractions and running GoldenGate assays; Clare Gill for assistance with the GenomeStudio software and for providing the facilities at Texas A&M University for SNP genotyping; José Rufino for providing computer power for running the simulations. Three reviewers made insightful comments on an earlier draft of the manuscript. GoldenGate SNP data were produced with support from the Whole Systems Genomics Initiative at Texas A&M University in cooperation with Texas A&M AgriLife Research. Julio Chávez-Galarza was supported by Fundação para a Ciência e Tecnologia through the PhD scholarship SFRH/BD/68682/2010. Financial support for this research was provided.
Diversity of protected A. m. mellifera

by Fundação para a Ciência e Tecnologia and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008.

References


SHEN, R; FAN, J B; CAMPBELL, D; CHANG, W H; CHEN, J; DOUCET, D; YEAKLEY, J; BIKIKOVA, M; GARCIA, E W; MCBRIDE, C; STEEMERS, F; GARCIA, F; KERMANI, B G; GUNDERSON, K; OLIPHANT, A (2005) High-throughput SNP genotyping on universal bead arrays. Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 573: 70-82. http://dx.doi.org/10.1016/j.mrfmmm.2004.07.022


Online Supplementary Material: http://www.ibra.org.uk/downloads/20140515_1/download