US
UNIVERSITY OF SUSSEX

Sussex Research

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, Dora Henriques, Julio Chavez-Galarza, Per Kryger, Lionel Garnery, Romee van der Zee, Bjorn Dahle, Gabriele Soland-Reckeweg, Pilar de la Rua, Raffaele Dall' Olio, Norman Carreck, J Spencer Johnson

Publication date
01-05-2014

Licence
This work is made available under the Copyright not evaluated licence and should only be used in accordance with that licence. For more information on the specific terms, consult the repository record for this item.

Citation for this work (American Psychological Association 7th edition)
Pinto, M. A., Henriques, D., Chavez-Galarza, J., Kryger, P., Garnery, L., van der Zee, R., Dahle, B., Soland-Reckeweg, G., de la Rua, P., Dall' Olio, R., Carreck, N., & Johnson, J. S. (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* (Version 1). University of Sussex. https://hdl.handle.net/10779/uos.23403830.v1

Published in
Journal of Apicultural Research

Link to external publisher version
https://doi.org/10.3896/IBRA.1.53.2.08

Copyright and reuse:
This work was downloaded from Sussex Research Open (SRO). This document is made available in line with publisher policy and may differ from the published version. Please cite the published version where possible. Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners unless otherwise stated. For more information on this work, SRO or to report an issue, you can contact the repository administrators at sro@sussex.ac.uk. Discover more of the University's research at https://sussex.figshare.com/

Sussex Research Online

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

Article (Unspecified)

Citation:

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/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

Journal of Apicultural Research 53(2): 269-278 (2014)
DOI 10.3896/IBRA.1.53.2.08
© IBRA 2014

ORIGINAL RESEARCH ARTICLE

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 Henriques[1], Julio Chávez-Galarza[1], Per Kryger[2], Lionel Garnery[3], Romée van der Zee[4], Bjørn Dahle[5,6], Gabriele Soland-Reckeweg[7], Pilar de la Rúa[8], Raffaele Dall' Olio[9], Norman L Carreck[10,11] and J Spencer Johnston[12]

[1]Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, *Campus* de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal.
[2]Århus University, Flakkebjerg, Forsøgsvej 1, 4200 Slagelse, Denmark.
[3]Université de Versailles Saint-Quentin-en-Yvelines, 45 Ave. des Etats-Unis, 78035 Versailles, France.
[4]Netherlands Centre Bee Research, NL-9014 CC Tersoal, Netherlands.
[5]Norwegian Beekeepers Association, NO-2040 Kløfta, Norway.
[6]Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway.
[7]Apigenix, Gaicht 19, 2513 Twann, Switzerland.
[8]Área de Biología Animal, Dpto.de Zoología y Antropología Física, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain.
[9]Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Via di Saliceto 80, 40128 Bologna, Italy.
[10]Laboratory of Apiculture and Social Insects, School of Life Sciences, University of Sussex, Falmer, Brighton, East Sussex, BN1 9QG, UK.
[11]International Bee Research Association, 6, Centre Court, Main Avenue, Treforest, CF37 5YR, UK.
[12]Dept. 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 $tRNA^{leu}$-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 conmensurable de su acervo genético derivado de abejas comerciales del linaje C. Aquí hemos utilizado datos tanto de la secuencia de la región intergénica $tRNA^{leu}$-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 cría 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, $tRNA^{leu}$-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^{leu}-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.

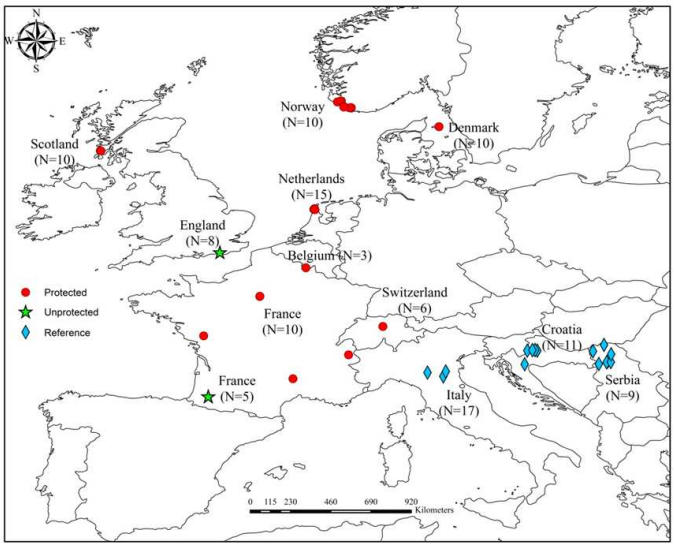

Fig. 1. Location of protected and unprotected groups, sampled in the native range of the M-lineage *A. m. mellifera* (western Europe), and of the reference group sampled in the in the native range of the C-lineage *A. m. ligustica* and *A. m. carnica* (eastern Europe) subspecies. The number of colonies sampled per site was variable. Samples sizes are indicated within parentheses.

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 isoamyl alcohol (25:24:1) protocol (Sambrook *et al.*, 1989) from the thorax of the 114 samples. MtDNA was analysed using the highly polymorphic tRNA^{leu}-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® 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 equidistant without regard to position within or near coding regions. Genotype calling was performed using Illumina's Genome Studio® 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 (uh) were computed using GenAlEx 6.4 (Peakall and Smouse, 2006) whereas allelic richness (Rs) was computed using HP-RARE 1.1 (Kalinowski, 2005), which implements the rarefaction method.

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 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 tRNA^{leu}-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.

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

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.

Location	**Poly-morphic**	**Private**	***Na***		***Ne***		***uh***		***Rs***	
			Mt DNA	**SNPs**	**Mt DNA**	**SNPs**	**Mt DNA**	**SNPs**	**Mt DNA**	**SNPs**
Protected										
France	733	0	6	1.713 (0.013)	4.167	1.325 (0.009)	0.844	0.233 (0.005)	1.800	1.220 (0.005)
Belgium	279	0	3	1.234 (0.012)	3.000	1.187 (0.010)	1.000	-	1.800	1.126 (0.007)
Denmark	738	0	1	1.624 (0.014)	1.000	1.296 (0.009)	0.000	0.211 (0.006)	1.000	1.199 (0.005)
Netherlands	755	1	3	1.638 (0.014)	1.991	1.217 (0.008)	0.533	0.156 (0.005)	1.515	1.151 (0.005)
Switzerland	575	0	2	1.486 (0.015)	1.385	1.292 (0.010)	0.333	0.211 (0.007)	1.303	1.191 (0.006)
Scotland	405	0	1	1.342 (0.014)	1.000	1.179 (0.009)	0.000	0.122 (0.005)	1.000	1.116 (0.005)
Norway	332	0	4	1.281(0.013)	3.333	1.145 (0.008)	0.778	0.099 (0.005)	1.737	1.094 (0.005)
All group	1020	1	17	1.938 (0.07)	7.262	1.258 (0.008)	0.876	0.179 (0.004)	1.869	1.176 (0.004)
Unprotected										
England	827	0	3	1.699 (0.013)	2.462	1.406 (0.010)	0.679	0.280 (0.006)	1.633	1.261 (0.006)
France	817	11	4	1.691 (0.013)	3.571	1.512 (0.011)	0.900	0.361 (0.07)	1.800	1.320 (0.007)
All group	1073	0	7	1.907 (0.08)	5.121	1.555 (0.009)	0.805	0.354 (0.005)	1.836	1.339 (0.005)
Reference										
Italy	709	1	2	1.599 (0.014)	1.710	1.273 (0.009)	0.441	0.182 (0.005)	1.428	1.176 (0.005)
Serbia	590	0	4	1.499 (0.015)	2.077	1.262(0.009)	0.583	0.186 (0.006)	1.549	1.173 (0.006)
Croatia	681	0	5	1.576 (0.014)	3.270	1.267(0.009)	0.764	0.186 (0.006)	1.727	1.177 (0.005)
All group	969	1	7	1.819 (0.011)	4.319	1.297(0.009)	0.790	0.198 (0.005)	1.779	1.196 (0.005)
All groups	1183	13	25	2.000 (0.000)	12.101	1.686 (0.008)	0.925	0.388 (0.004)	1.921	1.386 (0.004)

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 (51 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).

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.

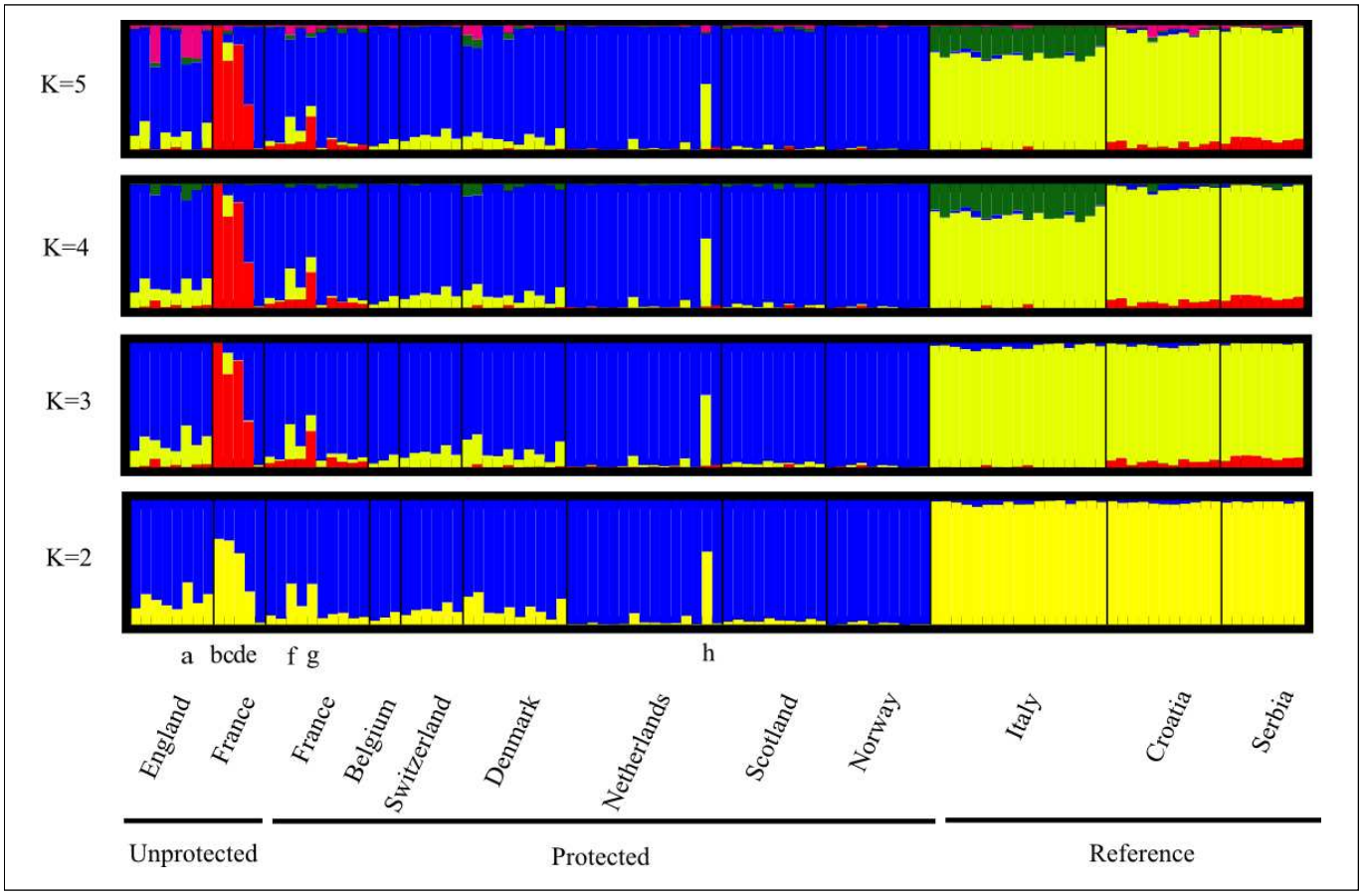

Fig. 2. Estimated population structure and admixture levels obtained with STRUCTURE based on 1183 SNP loci. Each individual is represented by a bar, which is partitioned into K collared segments that represent the individual's estimated membership proportions (Q) in K clusters. Black lines separate individuals of different locations clustered into the three studied groups (ordered from left to right): unprotected, protected and reference groups. The lower case letters (a-h) mark the same colonies in Figs. 2 and 4.

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 ΔK method (Evanno *et al.*, 2005), colonies of the reference group formed one cluster (marked in yellow) with an average membership coefficient of 0.98 ± 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 ± 0.003 and 0.97 ± 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 ± 0.02, in Denmark, to 0.88 ± 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 ± 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

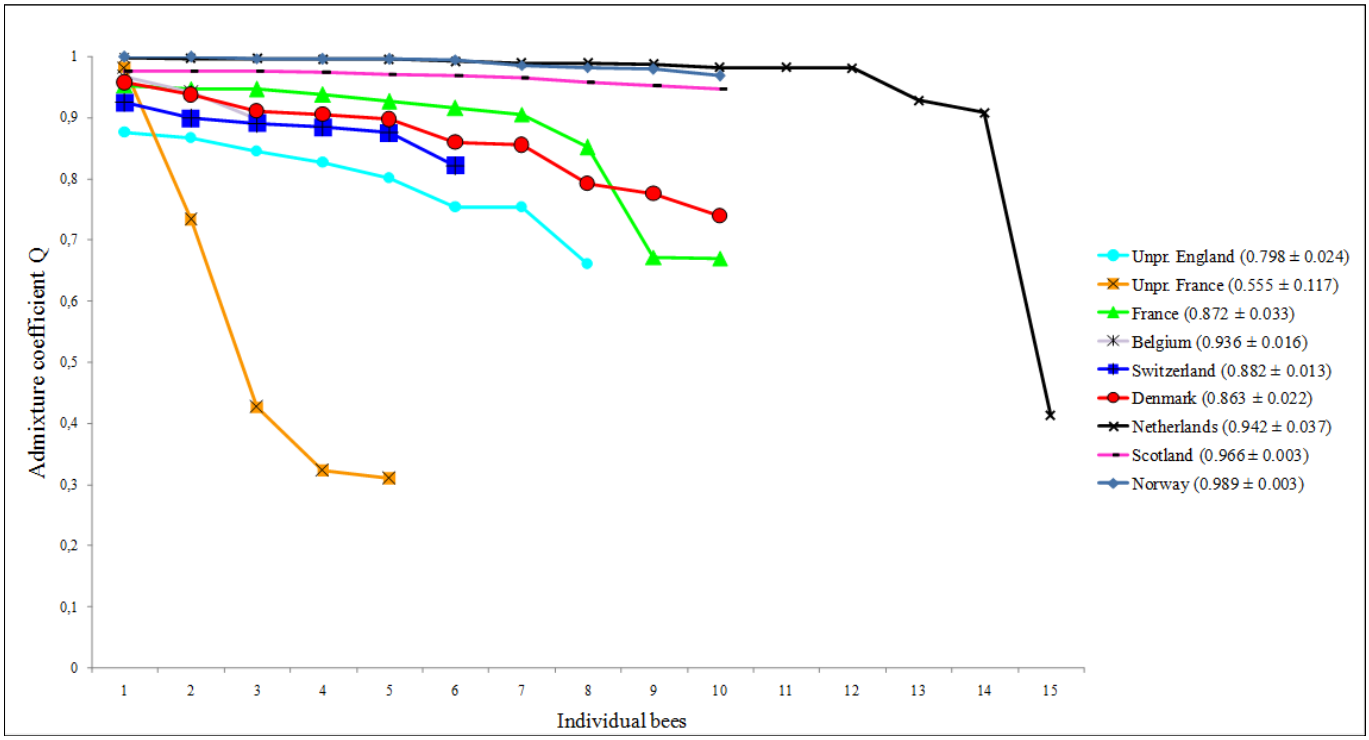

Fig. 3. Distribution of posterior mean estimates of membership proportion (Q), in the blue cluster of Fig. 2, for each individual of the protected and unprotected groups obtained with STRUCTURE, based on 1183 SNP loci, for K = 2. Average membership proportion ± SE, for each location, are indicated at the right side of the chart.

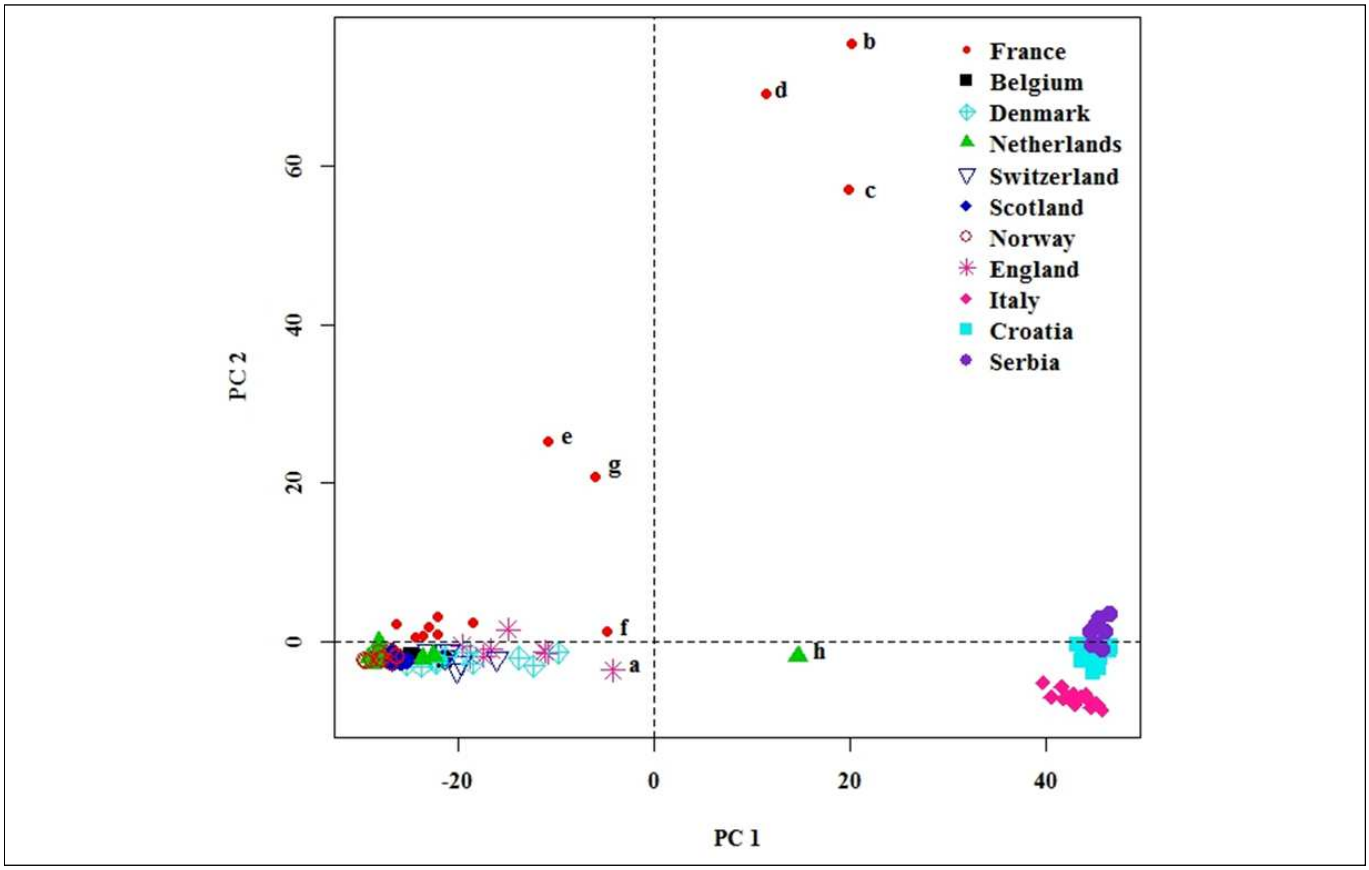

Fig. 4. Principal component analysis (PCA) based on the 1183 SNP loci. PC1 separates colonies sampled in eastern Europe (native range of *A. m. carnica* and *A. m. ligustica*) from colonies sampled in western Europe (native range of *A. m. mellifera*) whereas PC2 separates colonies of the unprotected 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 genome-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 *Dra*I 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 *Dra*I 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 Meixner, 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

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

References

CHÁVEZ-GALARZA, J; HENRIQUES, D; JOHNSTON, J S; AZEVEDO, J C; PATTON, J C; MUÑOZ, I; DE LA RÚA, P; PINTO, M A (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms (SNPs). *Molecular Ecology* 22: 5890-5907. http://dx.doi.org/10.1111/mec.12537

CORNUET, J M; GARNERY, L; SOLIGNAC, M (1991) Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. *Genetics* 128: 393-403.

CROZIER, R H; CROZIER, Y C (1993) The mitochondrial genome of the Honey bee *Apis mellifera:* complete sequence and genome organization. *Genetics* 133: 97-117.

DE LA RÚA, P; JAFFÉ, R; DALL'OLIO, R; MUÑOZ, I; SERRANO, J (2009) Biodiversity, conservation and current threats to European honey bees. *Apidologie* 40: 263-284. http://dx.doi.org/10.1051/apido/2009027

DE LA RÚA, P; JAFFÉ, R; MUÑOZ, I; SERRANO, J; MORITZ, R F A; KRAUS, F B (2013) Conserving genetic diversity in the honey bee. Comments on Harpur *et al.* (2012). *Molecular Ecology* 22: 3208-3210. http://dx.doi.org/10.1111/MEC.12333

DIETEMANN, V; PIRK, C W W; CREWE, R (2009) Is there a need for conservation of honey bees in Africa? *Apidologie* 40: 285-295. http://dx.doi.org/10.1051/apido/2009013

EARL, D A; VONHOLDT, B M (2012) STRUCTURE HARVESTER: a web-site and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361. http://dx.doi.org/10.1007/s12686-011-9548-7

ENGEL, M S (1999) The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; *Apis*). *Journal of Hymenoptera Research* 8: 165-196.

EVANNO, G; REGNAUT, S; GOUDET, J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x

FRANCK, P; GARNERY, L; CELEBRANO, G; SOLIGNAC, M; CORNUET, J M (2000) Hybrid origins of honey bees from Italy (*Apis mellifera ligustica*) and Sicily (*A. m. sicula*). *Molecular Ecology* 9: 907-921. http://dx.doi.org/10.1046/j.1365-294x.2000.00945.x

GARNERY, L; FRANCK, P; BAUDRY, E; VAUTRIN, D; CORNUET, J M; SOLIGNAC, M (1998) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution* 30: S31-S47.

GARNERY, L; SOLIGNAC, M; CELEBRANO, G; CORNUET, J M (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia* 49: 1016-1021.

HARPUR, B A; MINAEI, S; KENT, C F; ZAYED, A (2012) Management increases genetic diversity of honey bees via admixture. *Molecular Ecology* 21: 4414-4421. http://dx.doi.org/10.1111/j.1365-294X.2012.05614.x

HARPUR, B A; MINAEI, S; KENT, C F; ZAYED, A (2013) Admixture increases diversity in managed honey bees: reply to De la Rúa *et al.*, 2013. *Molecular Ecology* 22: 3211-3215. http://dx.doi.org/10.1111/mec.12332

HEPBURN, H R; RADLOFF, S E (1998) *Honey bees of Africa*. Springer; Berlin, Germany. 370 pp.

JAKOBSSON, M; ROSENBERG, N A (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806. http://dx.doi.org/10.1093/bioinformatics/btm233

JENSEN, A B; PALMER, K A; BOOMSMA, J J; PEDERSEN, B V (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honey bee, *Apis mellifera mellifera*, in north west Europe. *Molecular Ecology* 14: 93-106. http://dx.doi.org/10.1111/j.1365-294X.2004.02399.x

JOMBART, T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. http://dx.doi.org/10.1093/bioinformatics/btn129

KALINOWSKI, S T (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187-189. http://dx.doi.org/10.1111/j.1471-8286.2004.00845.x

MATTILA, H R; SEELEY, T D (2007) Genetic diversity in honey bee colonies enhances productivity and fitness. *Science* 317: 362-364. http://dx.doi.org/10.1126/science.1143046

MEIXNER, M D; COSTA, C; KRYGER, P; HATJINA, F; BOUGA, M; IVANOVA, E; BÜCHLER, R (2010) Conserving diversity and vitality for honey bee breeding. *Journal of Apicultural Research* 49(1): 85-92. http://dx.doi.org/10.3896/IBRA.1.49.1.12

MEIXNER, M D; LETA, M A; KOENIGER, N; FUCHS, S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera* - *Apis mellifera simensis* n. ssp. *Apidologie* 42: 425-437. http://dx.doi.org/10.1007/s13592-011-0007-y

MEIXNER, M D; PINTO, M A; BOUGA, M; KRYGER, P; IVANOVA, E; FUCHS, S (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera.* In *V Dietemann; J D Ellis; P Neumann (Eds) The COLOSS* BEEBOOK*, Volume I: standard methods for* Apis mellifera *research*. *Journal of Apicultural Research* 52(4): http://dx.doi.org/10.3896/IBRA.1.52.4.05

MUÑOZ, I; MADRID-JIMÉNEZ, M J; DE LA RÚA, P (2011) Temporal genetic analysis of an introgressed island honey bee population (Tenerife, Canary Islands, Spain). *Journal of Apicultural Research* 51(1): 144-146. http://dx.doi.org/10.3896/IBRA.1.51.1.20

MUÑOZ, I; PINTO, M A; DE LA RÚA, P (2013) Temporal changes in mitochondrial diversity highlights contrasting population events in Macaronesian honey bees. *Apidologie* 44: 295-305. http://dx.doi.org/10.1007/s13592-012-0179-0

NEI, M (1987) *Molecular evolutionary genetics.* Columbia University Press; New York, USA. 512 pp.

OLDROYD, B P; FEWELL, J H (2007) Genetic diversity promotes homeostasis in insect colonies. *Trends in Ecology & Evolution* 22: 408-413. http://dx.doi.org/10.1016/j.tree.2007.06.001

OLEKSA, A; CHYBICKI, I; TOFILSKI, A; BURCZYK, J (2011) Nuclear and mitochondrial patterns of introgression into native dark bees (*Apis mellifera mellifera*) in Poland. *Journal of Apicultural Research* 50(2): 116-129. http://dx.doi.org/10.3896/IBRA.1.50.2.03

PAGE, R E (1980) The evolution of multiple mating-behaviour by honey bee queens *Apis mellifera* L. *Genetics* 96: 263-273.

PEAKALL, R; SMOUSE, P E (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x

PETIT, R J; EL MOUSADIK, A; PONS, O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844-855.

PRITCHARD, J K; STEPHENS, M; DONNELLY, P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

RORTAIS, A; ARNOLD, G; ALBURAKI, M; LEGOUT, H; GARNERY, L (2011) Review of the *Dra*I COI-COII test for the conservation of the black honey bee (*Apis mellifera mellifera*). *Conservation Genetics Resources* 3: 383-391. http://dx.doi.org/10.1007/s12686-010-9351-x

ROSENBERG, N A (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137-138. http://dx.doi.org/10.1046/j.1471-8286.2003.00566.x

RUTTNER, F (1988) *Biogeography and taxonomy of honey bees.* Springer; Berlin, Germany. 284 pp.

RUTTNER, F; MILNER, E; DEWS, J E (1990) *The Dark European honey bee,* Apis mellifera mellifera *Linnaeus 1758*. British Isles Bee Breeders Association; Derby, UK.

SAMBROOK, J; FRITSCH, E F; MANIATIS, T (1989) *Molecular cloning (2nd Ed.).* Cold Spring Harbor Laboratory Press; New York, USA. 1626 pp.

SEELEY, T D; TARPY, D R (2007) Queen promiscuity lowers disease within honey bee colonies. *Proceedings of the Royal Society B-Biological Sciences* 274: 67-72. http://dx.doi.org/10.1098/rspb.2006.3702

SHEN, R; FAN, J B; CAMPBELL, D; CHANG, W H; CHEN, J; DOUCET, D; YEAKLEY, J; BIBIKOVA, 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

SHEPPARD, W S; MEIXNER, M D (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie* 34: 367-375. http://dx.doi.org/10.1051/apido:2003037

SOLAND-RECKEWEG, G; HECKEL, G; NEUMANN, P; FLURI, P; EXCOFFIER, L (2009) Gene flow in admixed populations and implications for the conservation of the Western honey bee, *Apis mellifera*. *Journal of Insect Conservation* 13: 317-328. http://dx.doi.org/10.1007/s10841-008-9175-0

TAMURA, K; PETERSON, D; PETERSON, N; STECHER, G; NEI, M; KUMAR, S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739. http://dx.doi.org/10.1093/molbev/msr121

TARPY, D R (2003) Genetic diversity within honey bee colonies prevents severe infections and promotes colony growth. *Proceedings of the Royal Society B-Biological Sciences* 270: 99-103. http://dx.doi.org/10.1098/rspb.2002.2199

VANENGELSDORP, D; MEIXNER, M D (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* 103: S80-S95. http://dx.doi.org/10.1016/j.jip.2009.06.011

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