SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*)

IRENE MUÑOZ,*,† DORA HENRIQUES,*,‡ LAURA JARA,† J. SPENCER JOHNSTON,§
JULIO CHÁVEZ-GALARZA,* PILAR DE LA RÚA† and M. ALICE PINTO*
*Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, †Área de Biología Animal, Dpto. de Zoología y Antropología Física, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain, ‡Centre of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar,4710-057 Braga, Portugal, §Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA

Abstract

The honeybee (*Apis mellifera*) has been threatened by multiple factors including pests and pathogens, pesticides and loss of locally adapted gene complexes due to replacement and introgression. In western Europe, the genetic integrity of the native *A. m. mellifera* (M-lineage) is endangered due to trading and intensive queen breeding with commercial subspecies of eastern European ancestry (C-lineage). Effective conservation actions require reliable molecular tools to identify pure-bred *A. m. mellifera* colonies. Microsatellites have been preferred for identification of *A. m. mellifera* stocks across conservation centres. However, owing to high throughput, easy transferability between laboratories and low genotyping error, SNPs promise to become popular. Here, we compared the resolving power of a widely utilized microsatellite set to detect structure and introgression with that of different sets that combine a variable number of SNPs selected for their information content and genomic proximity to the microsatellite loci. Contrary to every SNP data set, microsatellites did not discriminate between the two lineages in the PCA space. Mean introgression proportions were identical across the two marker types, although at the individual level, microsatellites' performance was relatively poor at the upper range of Q-values, a result reflected by their lower precision. Our results suggest that SNPs are more accurate and powerful than microsatellites for identification of *A. m. mellifera* colonies, especially when they are selected by information content.

Introduction

The western honeybee, *Apis mellifera* L., is currently distributed worldwide. However, prior to the human-assisted global expansion, this species was confined to western Asia, Middle East, Africa and Europe (Ruttner 1988; Chen *et al.* 2016). In such an environmentally heterogeneous range, the honeybee has differentiated into 31 currently recognized subspecies (Ruttner 1988; Sheppard & Meixner 2003; Meixner *et al.* 2011; Chen *et al.* 2016), which have been grouped into four main evolutionary lineages: M (western European), C (eastern European), O (Middle Eastern) and A (African) (Ruttner 1988). This vast diversity has been increasingly

Correspondence: M. Alice Pinto, Fax: +351273325405; E-mail: apinto@ipb.pt

threatened by major factors including habitat loss and fragmentation, pesticides and spread of pests and pathogens (Potts *et al.* 2010; Van Engelsdorp & Meixner 2010). An additional, but less publicized, threat comes from honeybee queen (legal or illegal) trade and intensive queen breeding. Large-scale movements of commercial queen strains, usually of the beekeepers-favoured C-lineage ancestry, represent a risk for local populations, not only because they may bring new parasites and pathogens or more virulent strains of established parasites and pathogens (Mutinelli 2011; Muñoz *et al.* 2014a; McMahon *et al.* 2016), but also because of introgressive hybridization (Jensen *et al.* 2005; Soland-Reckeweg *et al.* 2009; Muñoz *et al.* 2014b; Pinto *et al.* 2014).

There are growing concerns that intensified queen breeding and trade may promote gene flow between

native and commercial strains leading to an irremediable loss of diversity adapted to local conditions (De la Rúa *et al.* 2009; 2013; Büchler *et al.* 2014). This is the case for the M-lineage *A. m. mellifera* (the dark European honeybee), which in a substantial portion of its native range in western Europe is gravely threatened by C-derived (mainly *A. m. ligustica* and *A. m. carnica*) introgression (Jensen *et al.* 2005; Soland-Reckeweg *et al.* 2009; De la Rúa *et al.* 2009; Nedić *et al.* 2014; Pinto *et al.* 2014). In an attempt to reverse this threat, several conservation programmes have been implemented across Europe (De la Rúa *et al.* 2009; Muñoz *et al.* 2015). An efficient management of these programmes requires, however, molecular tools capable of reliably identifying pure-bred *A. m. mellifera* colonies in a cost- and time-effective manner.

While a variety of molecular markers, including RFLPs (Hall 1990), AFLPs (Suazo & Hall 1999), RAPDs (Hunt & Page 1995) and STSs (Arechavaleta-Velasco *et al.* 2003), have been employed in honeybee research, microsatellites (aka short tandem repeats, STRs) have indisputably been the marker of choice over the last 15–20 years (reviewed by Meixner *et al.* 2013). Numerous studies have demonstrated their usefulness in unravelling the signatures of historical and contemporary human-driven events in the native (Franck *et al.* 1998, 2001; Garnery *et al.* 1998; De la Rúa *et al.* 2001, 2003, 2006; Sušnik *et al.* 2004; Bodur & Kence 2007; Dall'Olio *et al.* 2007; Miguel *et al.* 2007; Muñoz *et al.* 2009; Cánovas *et al.* 2011; Coroian *et al.* 2014; Francis *et al.* 2014; Nedić *et al.* 2014; Uzunov *et al.* 2014; Péntek-Zakar *et al.* 2015) and introduced (Clarke *et al.* 2002; Pinto *et al.* 2005; Galindo-Cardona *et al.* 2013; Rangel *et al.* 2016) distributional range of the honeybee. Microsatellites have been particularly useful for identifying introgression of C-derived genes into gene pools of native honeybees and monitoring conservation programmes of *A. m. mellifera* in the Danish island of Læsø (Jensen *et al.* 2005), in the French region of Landes (Strange *et al.* 2008), in the eastern part of Switzerland (Soland-Reckeweg *et al.* 2009) and in the north-eastern part of Poland (Oleksa *et al.* 2011), of *A. m. carnica* in Slovenia (Sušnik *et al.* 2004), of *A. m. iberiensis* in the Canary islands (Muñoz *et al.* 2012) and of *A. m. siciliana* in the Filicudi and Vulcano islands (Muñoz *et al.* 2014c).

Single-nucleotide polymorphisms (SNPs) represent the most recent addition to the molecular toolkit for honeybee genetic analysis and are rapidly becoming popular among honeybee scientists. SNPs have been used to scrutinize the evolutionary history of the honeybee (Whitfield *et al.* 2006; Wallberg *et al.* 2014; Chen *et al.* 2016), to search for footprints of selection (Zayed & Whitfield 2008; Spötter *et al.* 2012; Chávez-Galarza *et al.* 2013; Harpur *et al.* 2014; Wallberg *et al.* 2014; Fuller *et al.* 2015;

Chen *et al.* 2016; Wragg *et al.* 2016), to dissect the evolutionary complexities of the Iberian honeybee hybrid zone (Chávez-Galarza *et al.* 2015) and to examine genomewide recombination patterns (Wallberg *et al.* 2015). At a more practical level, the potential of SNPs for identifying African-derived genes in the European stock of North America (Chapman *et al.* 2015) and C-derived genes in *A. m. mellifera* in western Europe for commercial and conservation purposes has also been investigated (Pinto *et al.* 2014; Muñoz *et al.* 2015). However, whether SNPs can replace microsatellites for identifying genetic stocks needs to be addressed. While as a usually biallelic marker, the per locus information content of a SNP is lower than that of a multiallelic microsatellite, this drawback can be offset by employing large numbers of SNPs, whose identification is greatly facilitated in the genomics era. The average number of random SNPs required to equal the information content of random microsatellites has been estimated to be two to 10 times, depending largely on the question under scrutiny (Kalinowski 2002; Thalamuthu *et al.* 2004; Herráez *et al.* 2005; Liu *et al.* 2005; Schopen *et al.* 2008; Gärke *et al.* 2012). The ratio can, however, be lowered if informative SNPs selected from a larger panel are employed instead of randomly selected SNPs (Glover *et al.* 2010; Gärke *et al.* 2012; Ozerov *et al.* 2013).

Given the high number of SNPs available for the honeybee (Whitfield *et al.* 2006; Spötter *et al.* 2012; Pinto *et al.* 2014; Chapman *et al.* 2015), the challenge is to identify the most informative. Several approaches can be implemented to measure the contribution of single SNPs, which can then be ranked by information content (Muñoz *et al.* 2015). In this context, a subset of SNPs, the so-called ancestry informative markers (AIMs), displays large allele frequency differences between populations. Ranking the most informative SNPs allows one to design reduced SNP panels that correctly assign individuals to ancestry origin. Reduced SNP panels have been used by others to delineate the genetic structure of honeybee populations in Canada (Harpur *et al.* 2015) and identify Africanized honeybees in North America (Chapman *et al.* 2015), and by us to estimate introgression in *A. m. mellifera* populations in Europe (Muñoz *et al.* 2015). Building from this previous study, here we compared the resolving power of a widely utilized 11-microsatellite set to detect structure and introgression with that of reduced SNP sets that were selected from a genomewide data set using two criteria: (i) the ranking in terms of information content and (ii) genomic proximity to the 11 microsatellites. Using these criteria, five reduced SNP sets (48, 96 and 144 top-ranked AIMs previously identified by Muñoz *et al.* 2015; and the closest five and 10 SNPs to each of the 11 microsatellites) were built and compared with

the 11-microsatellite set using principal component analysis (PCA) and a model-based clustering algorithm implemented in STRUCTURE (Pritchard *et al.* 2000). The ultimate goal of this study was to appraise the feasibility of a SNP-based alternative to microsatellites that can be used for identifying pure-bred *A. m. mellifera* genetic stock for breeding and for assisting management of conservation centres across Europe.

Methods

Samples

A total of 113 haploid males representing single colonies of three subspecies (*A. m. mellifera*, *A. m. ligustica* and *A. m. carnica*) was collected across Europe in 2010 and 2011. Seventy-seven were collected from the native range of the M-lineage subspecies *A. m. mellifera* in England, France, Belgium, Denmark, the Netherlands, Switzerland, Scotland and Norway, of which 64 originated from protected conservation areas and 13 from unprotected areas. The remainder made up a reference collection of 36 haploid males representing C-lineage diversity sampled in the native range of *A. m. ligustica* (N = 17) in Italy and *A. m. carnica* (N = 19) in Serbia and Croatia. Further details of sampling and DNA extraction procedures are provided in the study by Pinto *et al.* (2014) and Muñoz *et al.* (2015).

Microsatellite genotyping

Eleven widely utilized microsatellites were amplified in two multiplex PCR reactions. Plex-1 consisted of the five loci recommended in the Coloss Beebook (Dietemann *et al.* 2013), which is becoming the standard manual for honeybee research, namely A7, A113, Ap43, Ap55 and B124 (Evans *et al.* 2013). Plex-2 combined A8, A88, Ac11, Ap224, Ap249 and Ap274, which have been used to detect structure and introgression in different honeybee subspecies and populations (Chahbar *et al.* 2012; Coroian *et al.* 2014; Muñoz *et al.* 2014b; Nedić *et al.* 2014; Uzunov *et al.* 2014). Each reaction (10 µL) contained 1.2 mM MgCl$_2$, 0.3 mM of each dNTP, 0.4 µM of each forward and reverse primer, 1.5 U of *Taq* DNA polymerase (Biotools B&M Labs, Madrid, Spain) and 2 µL of extracted DNA. PCR amplification was performed at 95 °C for one 5-min cycle followed by 30 cycles at 95 °C for 30 s, either 54 °C (plex-1) or 50 °C (plex-2) for 30 s, 72 °C for 30 s and a final step at 72 °C for 30 min. PCR products were analysed on an ABI® 3730 automated DNA sequencer (Applied Biosystems, Waltham, USA) and sized with an internal standard. Alleles were subsequently scored using GENEMAPPER® v3.7 software (Applied Biosystems, Foster City, USA).

SNP genotyping

A total of 1536 SNP loci evenly spaced across the 16 honeybee linkage groups were genotyped using Illumina's BeadArray Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay (Illumina, San Diego, USA) following manufacturer's protocols. Upon the filtering process, 353 SNPs were removed because they did not meet the quality criteria for analysis (see Chávez-Galarza *et al.* 2013 for details). Allele frequencies were calculated for the remaining 1183 biallelic SNPs in each population using the program PLINK (Purcell *et al.* 2007).

Comparison analysis for the detection of population structure and introgression

The two types of molecular markers were examined by comparing the 11 microsatellites with the genomewide 1183 SNPs (hereafter named reference SNP set) and with five SNP sets. These included three reduced SNP sets formed by the top-ranked 48, 96 and 144 informative SNPs (hereafter named 48, 96 and 144 AIMs) identified by Muñoz *et al.* (2015) and two reduced SNP sets formed by the 55 and 110 SNPs closest to the 11 microsatellites (hereafter named 55 and 110 closest SNPs). The physical map of these loci shows that the 144 AIMs are distributed across the 16 honeybee linkage groups providing good genome coverage, in stark contrast to the 11 microsatellite loci and corresponding 110 closest SNPs, which only mark seven linkage groups (Fig. 1).

Genetic diversity was assessed for each subspecies from the microsatellite and the six SNP data sets using unbiased estimates of gene diversity (Nei 1987) and allelic richness (Petit *et al.* 1998). The mean number of alleles (Na), number of effective alleles (Ne) and unbiased diversity (uh) were computed using GENALEX 6.501 (Peakall & Smouse 2006, 2012), whereas allelic richness (Rs) was computed after rarefaction using HP-RARE 1.1 (Kalinowski 2005). Differentiation was estimated using the standardized G'_{ST} measure proposed by Hedrick (2005), which allows comparison between markers with different levels of genetic variation. Global and pairwise G'_{ST} values were estimated across the three subspecies and seven data sets with GENALEX 6.501.

The resolving power of the microsatellite set and the five reduced SNP sets to detect population structure and introgression was compared with the reference SNP set using both PCA and a model-based clustering approach. PCA was performed on a normalized matrix of individuals' genotypes to generate two-dimensional PCA and to visualize the stability of population assignment produced by all sets. PCA was implemented in the PAST
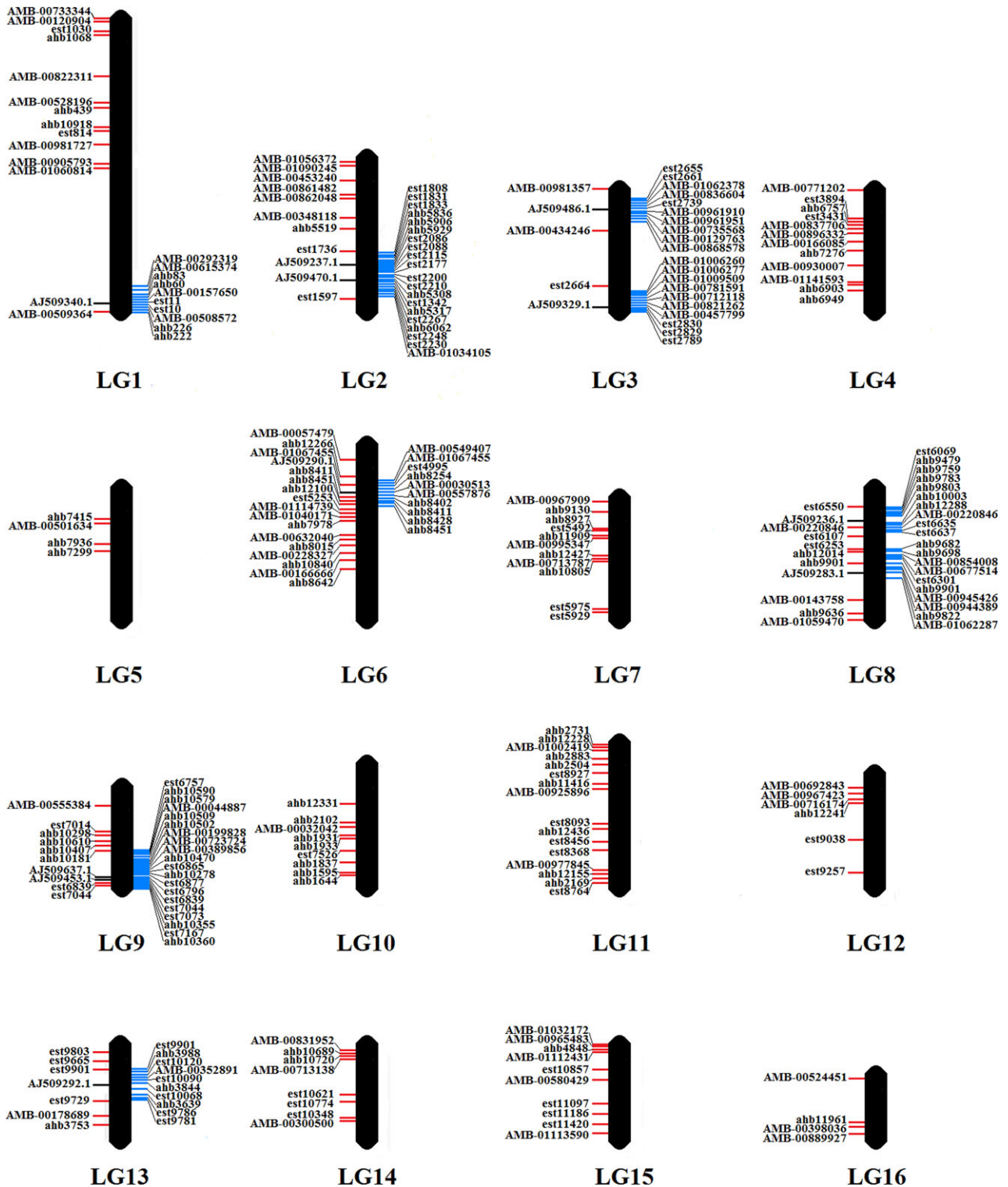
Fig. 1 Physical map of the 16 honeybee linkage groups showing the genomic positions of the 11 microsatellites (coded as AJ509XXX.1) marked in black, the 55 and 110 closest SNPs (five and 10 flanking each of the 11 microsatellite loci) marked in blue and the top-ranked 144 AIMs (includes the 48 and 96 AIMs) marked in red. The map was depicted from the honeybee genome sequence available at http://www.ncbi.nlm.nih.gov/projects/mapview using the Map Viewer tool. [Colour figure can be viewed at wileyonlinelibrary.com]

software (Hammer *et al.* 2001). The model-based Bayesian clustering algorithm implemented in STRUCTURE 2.3.3 (Pritchard *et al.* 2000) was employed to infer membership or introgression proportions (*Q*-value). 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 & 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 & Rosenberg 2007), was used to compute the pairwise 'symmetric similarity coefficient' between pairs of runs and to align the 20 runs for each *K*. Differences in diversity, F_{ST}, and *Q*-values between data sets and subspecies were assessed for statistical significance using the Tukey test implemented in PAST.

The precision of each marker type and set was tested against the reference SNP set by calculating linear regression coefficients (r^2) and the standard deviations of the differences between introgression proportions. Finally, the accuracy of the different sets was estimated via percentage of absolute error of introgression estimates in relation to the reference SNP set.

Results

In this study, the resolving power of microsatellites and SNPs to detect population structure and introgression was compared on 113 honeybee individuals representing three honeybee subspecies (*A. m. mellifera*, *A. m. ligustica*

and *A. m. carnica*) of the two evolutionary lineages (M and C) native to Europe.

Genetic diversity and differentiation

As expected, the SNP loci were biallelic, whereas the 11 microsatellite loci were multiallelic with the number of microsatellite alleles per locus ranging from two (Ap274) to 21 (A7). The mean number of microsatellite alleles per locus varied with subspecies, being 8.7, 3.6 and 4.7 for *A. m. mellifera*, *A. m. carnica* and *A. m. ligustica*, respectively. This wide variation across subspecies may be in part explained by the variable sample size and geographical distribution of samples. A summary of diversity measures (Na, Ne, uh and Rs) inferred from the microsatellite and the six SNP data sets for the three subspecies are shown in Table S1 (Supporting information). Diversity varied across subspecies with the highest values obtained for *A. m. mellifera*. Diversity measures obtained with the 11 microsatellites were significantly higher than those estimated by SNPs (0.0002 ≤ *P*-value ≤0.0066, Tukey test; Table S2, Supporting information).

Global and pairwise G'_{ST} values shown in Table 1 revealed variable levels of differentiation across markers and subspecies. Global G'_{ST} was lower for microsatellites (0.6371), and corresponding 55 (0.6274) and 110 (0.6172) flanking SNPs, than for any reduced SNP panel (0.8889, 0.8966 and 0.9044 for 144, 96 and 48 AIMs, respectively). As expected, pairwise G'_{ST} values showed the lowest differentiation for the two C-lineage subspecies *A. m. ligustica* and *A. m. carnica* (0.1336 for microsatellites and 0.1008 for the reference SNP data set) and the largest between the M-lineage *A. m. mellifera* and the two C-lineage subspecies (0.8098, 0.8082 for microsatellites and 0.7523, 0.7630 for the reference SNPs).

Table 1 Global and pairwise G'_{ST} values estimated by microsatellites and SNPs for the European honeybee subspecies *A. m. mellifera*, *A. m. ligustica* and *A. m. carnica*

Set	Global G'_{ST}	Pairwise G'_{ST}		
		A. m. mellifera vs. *A. m. ligustica*	*A. m. mellifera* vs. *A. m. carnica*	*A. m. ligustica* vs. *A. m. carnica*
11 STRs[*]	0.6371	0.8098	0.8082	0.1336
55 SNPs[†]	0.6274	0.6874	0.7162	0.1408
110 SNPs[‡]	0.6172	0.7008	0.6901	0.1270
48 AIMs[§]	0.9044	0.9651	0.9658	0.0242
96 AIMs[§]	0.8966	0.9623	0.9641	0.0168
144 AIMs[§]	0.8889	0.9581	0.9616	0.0245
1183 SNPs[¶]	0.6682	0.7523	0.7630	0.1008

[*]11 Microsatellite loci.
[†]55 Closest SNPs to the 11 microsatellites.
[‡]110 Closest SNPs to the 11 microsatellites.
[§]Reduced panels of top-ranked 48, 96 and 144 informative SNPs (AIMs—ancestry informative markers) identified by Muñoz et al. (2015).
[¶]Reference SNP data set.

Principal component analysis (PCA)

The results of PCA partitioning by the two markers and the reduced SNP sets are shown in Fig. 2. PCA grouping obtained with the microsatellites differed considerably from each of those obtained with the SNPs. The two main PCA components generated from the 11 microsatellites (Fig. 2a) showed a pronounced overlap of the 113 individuals representing the three *A. mellifera* subspecies and did not distinguish the M and C divergent lineages.
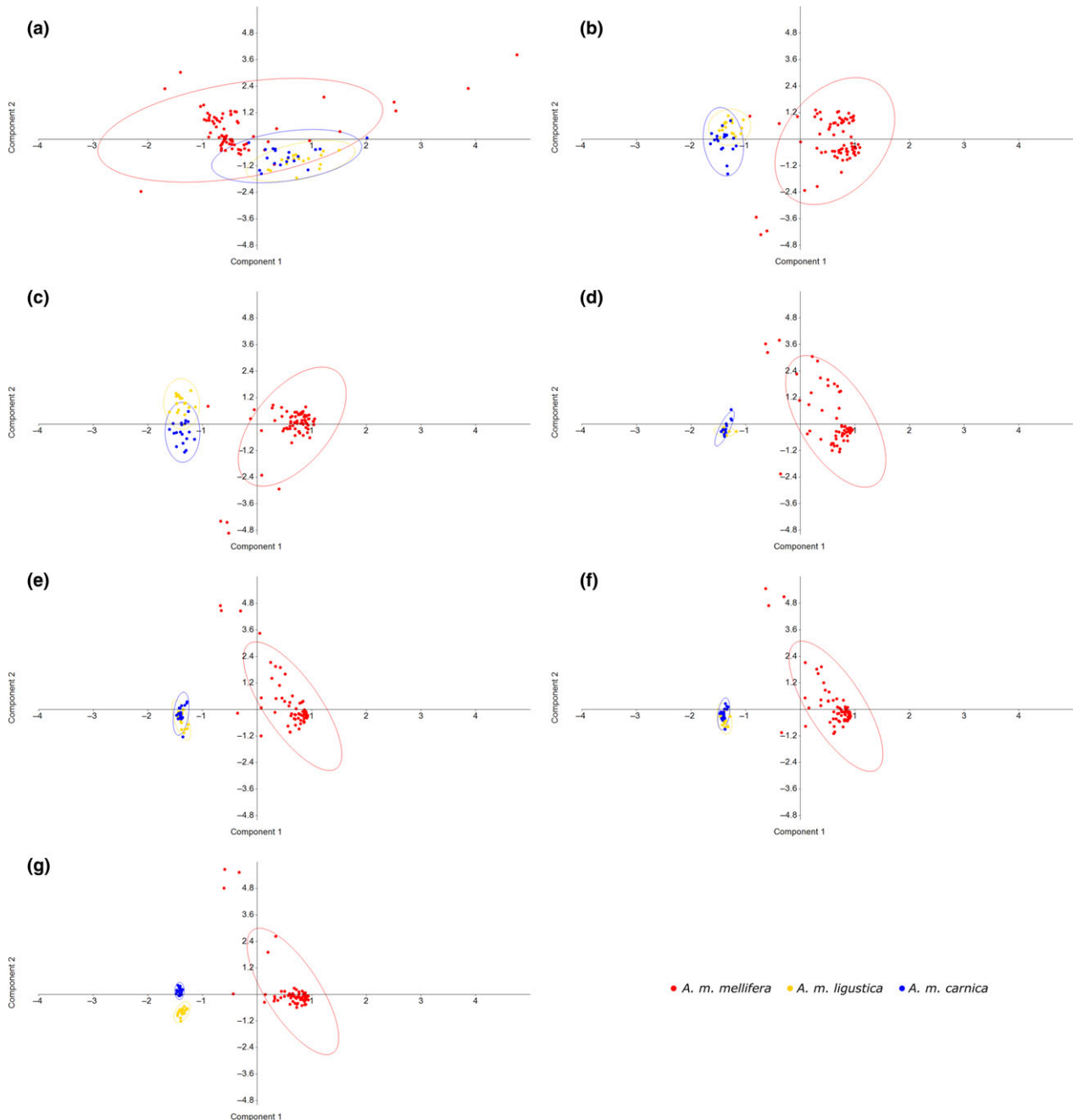

Fig. 2 Principal component analysis (PCA) of (a) 11 microsatellites, (b) 55 closest SNPs, (c) 110 closest SNPs, (d) 48 AIMs, (e) 96 AIMs, (f) 144 AIMs and (g) reference SNP set (d-g taken from Muñoz *et al.* 2015 for comparison purposes). PCAs show the 77 individuals sampled in the native range of the M-lineage honeybee subspecies *A. m. mellifera* in western Europe (marked in red), the 36 individuals sampled in the native range of the C-lineage subspecies *A. m. ligustica* (marked in yellow) and *A. m. carnica* (marked in blue) in eastern Europe. [Colour figure can be viewed at wileyonlinelibrary.com]

Contrasting with microsatellites, all SNP sets were able to distinguish individuals of M-lineage ancestry and C-lineage ancestry, although the degree of overlap varied. A greater overlap was observed for the 55 and 110 closest SNP data sets (Fig. 2b,c) than for the three AIM sets (Fig. 2d–f), which showed a separation pattern more similar to that exhibited by the reference SNP data set (Fig. 2g). The first two PCA components estimated using the 11 microsatellite, the 55 and the 110 closest SNP data sets, and the reference SNP data set explained 53.15, 51.70, 49.95 and 49.80% of the total variances, respectively. Higher values were obtained for the 48, 96 and 144 AIM data sets with 77.50, 76.29 and 75.35% of the total variance explained, respectively.

Clustering analysis with STRUCTURE

Membership proportions (Q) were inferred from microsatellites and SNPs for the 113 honeybee individuals using STRUCTURE (Fig. S1, Supporting information). The ΔK method (Evanno *et al.* 2005) indicated that $K = 2$ was the most likely number of genetic clusters, for both markers and for all data sets tested (Fig. S2, Supporting information). The 36 individuals originating from the C-lineage distributional range in eastern Europe formed one tight cluster with Q-values estimated by the reference SNP and the 11 microsatellite sets all at or above 0.9490 and 0.9217, respectively (Table 2, Table S3, Supporting information). The other cluster contained the 77 individuals sampled from protected and unprotected apiaries across the M-lineage *A. m. mellifera* native range in western Europe. Consistent with a previous report (Pinto *et al.* 2014), these individuals exhibited a wide array of Q-values denoting variable levels of C-lineage introgression (Fig. S1, Table S3, Supporting information).

While estimates of introgression proportions (inferred from Q-values) for the 77 individuals varied across markers and data sets, differences among them were more pronounced for the upper than for the lower range of Q-values (Table S3, Supporting information). For example, the five uppermost Q-values estimated by the reference SNP data set (or the AIMs) ranged from 0.3400 to 0.6902, whereas those estimated by microsatellites ranged from 0.7543 to 0.9602. Q-value differences $>|0.10|$ between these data sets were exhibited by 20 individuals (9 positive and 11 negative values), of which 14 were among the 20 most introgressed (Fig. S3, Supporting information). Nonetheless, when mean introgression proportions (Table 2) were compared across marker types and data sets for the 77 individuals, none of the pairwise comparisons revealed to be significant (P-value ≥ 0.9972, Tukey test; Table S4, Supporting information), despite the higher dispersion of data points observed for microsatellites (SD = 0.2222, Table 2).

Precision and accuracy

Membership proportions in the *A. m. mellifera* cluster estimated by the different data sets were further examined using linear regression (Table 3, Fig. S4, Supporting information). When microsatellites were regressed against the SNP data sets, the highest regression coefficient was obtained for the 55 closest SNPs ($r^2 = 0.6986$, Table 3), which suggests that even though the five flanking SNPs are on average 330425.5 ± 263719.9 bp away from each of the 11 microsatellite loci, and considering the extremely high recombination rate in honeybees (Wallberg *et al.* 2015), they are capturing the same information as microsatellites. On the other hand, when microsatellites and the five reduced SNP data sets were regressed against the reference SNP data set, which with its 1183 loci represent the best genome coverage, the lowest regression coefficient was produced by the microsatellite data set ($r^2 = 0.6202$, Table 3).

Precision and accuracy in estimating C-derived introgression into *A. m. mellifera* varied between marker types and data sets (Fig. 3). The standard deviations of

Table 2 Statistics of Q-values inferred from STRUCTURE for the individuals sampled in the M-lineage *A. m. mellifera* native range in western Europe ($N = 77$) and in the C-lineage *A. m. ligustica* and *A. m. carnica* native range in eastern Europe ($N = 36$)

Set	M-lineage group			C-lineage group		
	Max.	Min.	Mean $\pm$ SD	Max.	Min.	Mean $\pm$ SD
11 STRs	0.9602	0.0090	0.1177 ± 0.2222	0.9910	0.9217	0.9762 ± 0.0205
55 SNPs	0.8185	0.0050	0.1239 ± 0.1921	0.9940	0.8885	0.9727 ± 0.0272
110 SNPs	0.7790	0.0030	0.1230 ± 0.1825	0.9950	0.8842	0.9768 ± 0.0291
48 AIMs	0.6765	0.0040	0.1129 ± 0.1573	0.9960	0.9315	0.9822 ± 0.0170
96 AIMs	0.6706	0.0020	0.1075 ± 0.1531	0.9980	0.9487	0.9763 ± 0.0161
144 AIMs	0.6592	0.0020	0.1077 ± 0.1462	0.9980	0.9452	0.9746 ± 0.0142
1183 SNPs	0.6902	0.0010	0.1131 ± 0.1496	0.9990	0.9490	0.9829 ± 0.0123

Set	Parameter	11 STRs	55 SNPs	110 SNPs	48 AIMs	96 AIMs	144 AIMs
11 STRs	Slope a	–	0.7225	0.6703	0.5355	0.5190	0.4961
	Intercept b	–	0.0388	0.0441	0.0498	0.0464	0.0492
	r^2	–	0.6986	0.6661	0.5723	0.5678	0.5684
1183 SNPs	Slope a	0.5301	0.7392	0.7910	0.9259	0.9623	1.0103
	Intercept b	0.0507	0.0215	0.0158	0.0086	0.0097	0.0043
	r^2	0.6202	0.9014	0.9317	0.9484	0.9695	0.9758

Table 3 Parameters and coefficients obtained by the linear regression analysis of C-lineage introgression proportions in *A. m. mellifera*

Q-value differences were higher for microsatellites than for the five reduced SNP panels (Fig. 3a), which indicates that precision of the microsatellites was the lowest. When comparing among SNP data sets, the 55 and 110 closest SNPs provided less precise estimates than the AIMs, despite the identical number of loci included in the two groups of the reduced panels. Accuracy was high for the six sets, but the mean accuracy of microsatellites (91.84%) was lower than that provided by SNPs, which ranged from 95.17%, for the 55 closest SNPs, to 98.23%, for the 144 AIMs. It is noteworthy that at the individual level, microsatellite accuracy for the upper range of *Q*-values was highly variable and below 80% for 11 individuals, suggesting that mean values should be interpreted with caution (Fig. 3b). In summary, the SNP sets provided more accurate introgression estimates

than the microsatellite set, especially when they were selected by their information content (Muñoz *et al.* 2015).

Discussion

Reliable molecular tools for detecting C-lineage introgression and genetic identification of pure-bred *A. m. mellifera* colonies are crucial to effectively manage conservation centres not only for restoring and preserving genetic identity and diversity but also for increasing adaptively important traits of this endangered honeybee subspecies. While the PCR-RFLP of the intergenic tRNA[leu]-cox2 mitochondrial DNA region has proved to be a powerful and cost-effective tool for monitoring *A. m. mellifera* conservation centres in France (Bertrand *et al.* 2015), its maternal inheritance precludes identification of
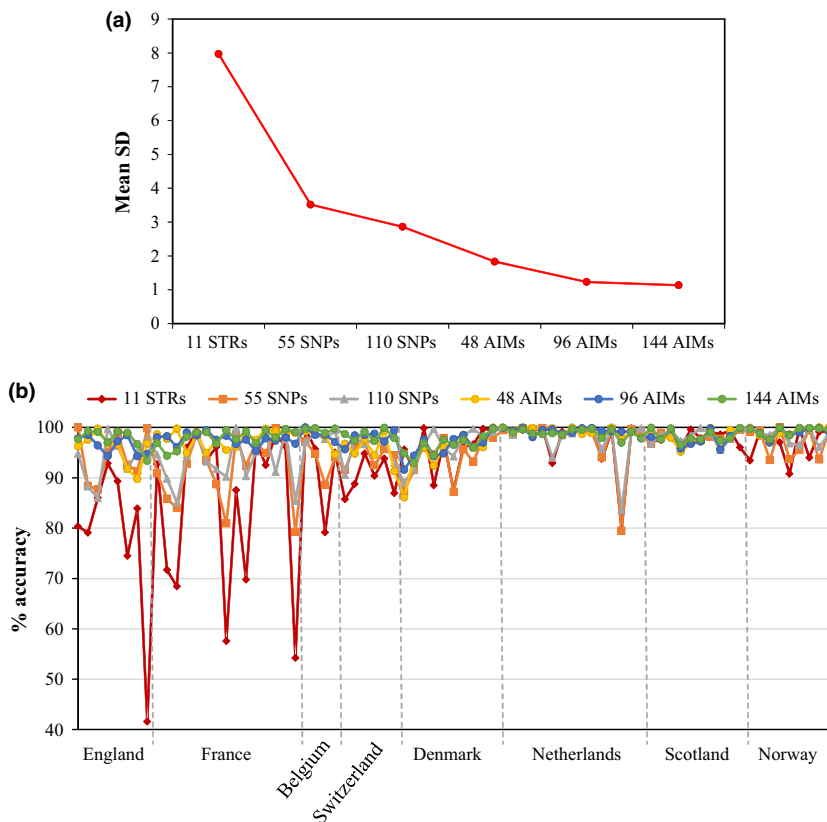

Fig. 3 (a) Precision estimates obtained from standard deviations (SD) of the differences between introgression inferred from the six sets in relation to the reference SNP set. The data set included the microsatellites (11 STRs) and the reduced SNP panels containing the 5 and 10 flanking SNPs of each microsatellite locus (55 and 110 closest SNPs) and the top-ranked AIMs (48, 96 and 144 AIMs). (b) Accuracy of the six data sets for each of the 77 *A. m. mellifera* individuals. Individuals are ordered as in Fig. S1 (Supporting information), which shows their introgression proportions. [Colour figure can be viewed at wileyonlinelibrary.com]

male-directed C-lineage introgression. Of the 77 *A. m. mellifera* colonies examined in this study, 76 carried haplotypes of M-lineage ancestry and one was a maternal descendant of a C-lineage colony (Pinto *et al.* 2014). However, these colonies exhibited variable levels of nuclear C-lineage introgression suggesting that a full identification of pure-bred *A. m. mellifera* requires biparentally inherited markers.

Microsatellites are still the most popular molecular tool for monitoring nuclear C-lineage introgression across *A. m. mellifera* conservation centres in Europe (L. Garnery, P. Kryger and G. Soland, pers. comm.). However, the benefits of using SNPs over microsatellites have been increasingly reported for many organisms (Väli *et al.* 2008, 2010; Glover *et al.* 2010; Hauser *et al.* 2011; Rašić *et al.* 2014), and honeybees are no exception. In this study, comparisons between different marker types and data sets showed that reduced sets of top-ranked informative SNPs (Muñoz *et al.* 2015) provide higher power in resolving the two highly divergent western (M) and eastern (C) European lineages, and are more accurate at estimating introgression proportions than microsatellites or their flanking SNPs (55 and 110 closest SNP sets). Our findings add to an increasing list of studies showing that SNPs outperform microsatellites in a variety of applications (Karlsson *et al.* 2007; Hauser *et al.* 2011; Gärke *et al.* 2012; Moore *et al.* 2014; Rašić *et al.* 2014; Puckett & Eggert 2016). This suggests that although biallelic SNPs provide lower information per locus than multiallelic microsatellites, the drawback can be offset by using a disproportionally larger number of randomly selected SNPs than microsatellites (Herráez *et al.* 2005; Liu *et al.* 2005; Narum *et al.* 2008; Hauser *et al.* 2011; Ciani *et al.* 2013; Rašić *et al.* 2014) or, instead, by using a reduced number of SNPs selected by information content (Rosenberg *et al.* 2003; Liu *et al.* 2005; Tokarska *et al.* 2009; Gärke *et al.* 2012; Hess *et al.* 2011; Oserov *et al.* 2013; this study).

Diversity estimates were lower for SNPs than for microsatellites (Table S1, Supporting information), which is expected given the disparity in the number of alleles per locus between the two markers (biallelic *versus* multiallelic). Global differentiation values obtained with the standardized G'_{ST} were identical across the two marker types (Table 1), although they were considerably higher for the reduced SNP data sets. This is an expected result given that the SNP loci of the reduced panels were selected by their highest resolution power for discriminating subspecies of the divergent M and C evolutionary lineages (Muñoz *et al.* 2015). The degree of differentiation between M and C lineages was high for both marker types and in the range reported by other honeybee studies (Whitfield *et al.* 2006; Harpur *et al.* 2012).

The PCA showed that subspecies partitioning differs substantially between the two markers (Fig. 2).

Specifically, microsatellites exhibited lower clustering in the PCA space than SNPs, an unexpected result given their widely claimed higher power in detecting population clustering (Herráez *et al.* 2005; Liu *et al.* 2005; Livingstone *et al.* 2011; Ciani *et al.* 2013; DeFaveri *et al.* 2013; Ross *et al.* 2014). While the number of loci used here is in the range of other studies (Narum *et al.* 2008; Schopen *et al.* 2008; Glover *et al.* 2010; Hauser *et al.* 2011; Hess *et al.* 2011; Livingstone *et al.* 2011), it is possible that the power of microsatellites was limited by the low genome coverage (only seven of the 16 honeybee chromosomes). However, the 55 and 110 closest SNP sets, which sample the same chromosomes, provided a lineage separation in the PCA space identical to that of the 144 AIMs, which are spread across the 16 chromosomes (Fig. 1). An alternative explanation for the poor lineage separation provided by microsatellites is homoplasy. Due to allele size constraints and high mutation rates, homoplasy is expected to occur relatively often in microsatellites, a problem that is aggravated with increasing divergence times (Kimura & Crow 1964; Estoup *et al.* 2002). Although geographically close, the two European lineages (M and C) are the most divergent among the four honeybee lineages (Garnery *et al.* 1992; Wallberg *et al.* 2014). Accordingly, it is possible that convergence of allele size has obscured lineage differentiation. This hypothesis is supported by a simulation study showing that for moderate to high levels of divergence, SNPs have generally greater power than microsatellites in detecting structure (Haasl & Payseur 2011).

While all reduced SNP data sets were able to separate the two lineages, *A. m. carnica* and *A. m. ligustica* could only be unambiguously distinguished by the 1183 SNP data set (Fig. 2g). This is not a surprising result given that the AIM panels were selected by their high discriminatory power in separating variation between and not within lineages. Furthermore, in accordance with the simulations of Haasl & Payseur (2011), as divergence time decreases, an exponential increase in the number of SNP loci is required for population separation. On the other hand, these authors found that in the presence of low levels of divergence, microsatellites may outperform SNPs, which was not the case here. It is possible that the 11 microsatellite loci were not sufficient to distinguish *A. m. carnica* from *A. m. ligustica*, as suggested by recent surveys using 25 microsatellite loci that resolved a number of C-lineage subspecies (Francis *et al.* 2014; Uzunov *et al.* 2014).

The results of STRUCTURE provided further insights into the relative performance of the two marker types. Although mean introgression proportions (inferred from Q-values) into *A. m. mellifera* estimated by microsatellites and SNPs were identical, at the individual level there were some major discrepancies, mostly in the upper

range of Q-values. The honeybee individuals 8 and 18 are two examples of highly skewed positive and negative microsatellite introgression estimates, respectively (Fig. S3, Supporting information). This finding has important implications for conservation programmes because decision-making works at the individual colony level; the decision of whether these two colonies would be maintained or removed from the protected population would be determined by the marker used for colony identification.

Precision and accuracy were lowest for microsatellites and highest for the SNP panel containing the largest number of top-ranked AIM loci. When comparing among SNP sets, both precision and accuracy were lowest for the two SNP sets flanking microsatellites, which indicate that SNPs carefully selected by their discriminatory power perform better than equivalent numbers of unselected SNPs. While mean accuracy was high across markers and data sets, at the individual level there was a trend showing lower accuracy at the upper range of Q-values, especially for microsatellites, suggesting that mean values can be misleading and are of little help for monitoring conservation programmes.

In summary, our results showed the superiority of SNPs in distinguishing the two European evolutionary lineages and estimating C-lineage introgression, especially when they are selected by their information content. These findings, together with high throughput, ease of analysis, transferability between laboratories, low genotyping error and low per locus genotyping cost (Vignal *et al.* 2002; Brumfield *et al.* 2003; Morin *et al.* 2004), make SNP markers more compliant to the test of tracking introgression, promising to supersede microsatellites in *A. m. mellifera* conservation programmes across Europe.

Acknowledgements

References

Arechavaleta-Velasco ME, Hunt GJ, Emore C (2003) Quantitative trait loci that influence the expression of guarding and stinging behaviors of individual honey bees. *Behavior Genetics*, **33**, 357–364.

Bertrand B, Alburaki M, Legout H, Moulin S, Mougel F, Garnery L (2015) MtDNA COI-COII marker and drone congregation area: an efficient method to establish and monitor honeybee (*Apis mellifera* L.) conservation centres. *Molecular Ecology Resources*, **15**, 673–683.

Bodur MK, Kence A (2007) Genetic structure of honeybee, *Apis mellifera* L. (Hymenoptera: Apidae) populations of Turkey inferred from microsatellite analysis. *Journal of Apicultural Research*, **46**, 50–56.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Chahbar N, Muñoz I, Dall'Olio R, De la Rúa P, Serrano J, Doumandji S (2012) Population structure of North African honey bees is influenced by both biological and anthropogenic factors. *Journal of Insect Conservation*, **17**, 385–392.

Chapman NC, Harpur BA, Lim J *et al.* (2015) A SNP test to identify Africanized honeybees via proportion of "African" ancestry. *Molecular Ecology Resources*, **15**, 1346–1355.

Chávez-Galarza J, Henriques D, Johnston JS *et al.* (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 5890–5907.

Chávez-Galarza J, Henriques D, Johnston JS *et al.* (2015) Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. *Molecular Ecology*, **24**, 2973–2992.

Chen C, Liu Z, Pan Q *et al.* (2016) Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. *Molecular Biology and Evolution*, **33**, 1337–1348.

Ciani E, Cecchi F, Castellana E *et al.* (2013) Poorer resolution of low-density SNP vs. STR markers in reconstructing genetic relationships among seven Italian sheep breeds. *Large Animal Review*, **19**, 236–241.

Clarke KE, Rinderer TE, Franck P, Quezada-Euán JG, Oldroyd BP (2002) The Africanization of honey bees (*Apis mellifera* L.) of the Yucatan: a study of a massive hybridization event across time. *Evolution*, **56**, 1462–1474.

Coroian CO, Muñoz I, Schluns EA *et al.* (2014) Climate rather than geography separates two European honeybee subspecies. *Molecular Ecology*, **23**, 2353–2361.

Dall'Olio R, Marino A, Lodesani M, Moritz RFA (2007) Genetic characterization of Italian honey bees, *Apis mellifera ligustica*, based on microsatellite DNA polymorphisms. *Apidologie*, **38**, 207–217.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2001) Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. *Molecular Ecology*, **19**, 1733–1742.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2003) Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. *Genetics, Selection, Evolution*, **35**, 339–350.

De la Rúa P, Galián J, Pedersen BV, Serrano J (2006) Molecular characterization and population structure of *Apis mellifera* from Madeira and the Azores. *Apidologie*, **37**, 699–708.

De la Rúa P, Jaffé R, Dall'Olio R, Muñoz I, Serrano J (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie*, **40**, 263–284.

DeFaveri J, Viitaniemi H, Leder E, Merilä J (2013) Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources*, **13**, 377–392.

Dietemann V, Ellis JD, Neumann P (eds.) (2013) *The COLOSS BEE-BOOK, Vol. I. Standard methods for Apis mellifera research*. International Bee Research Association, Cardiff, UK. 636 pp. ISBN 978-0-86098-274-6.

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. *Molecular Ecology*, **11**, 1591–1604.

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.

Evans JD, Schwarz RS, Chen YP et al. (2013) *Standard Methods for Molecular Research in Apis mellifera*. In The COLOSS BEEBOOK, Vol. I. Dietemann V, Ellis JD, and Neumann P (eds). Standard methods for *Apis mellifera* research. *Journal of Apicultural Research* **52**.

Francis RM, Kryger P, Meixner M et al. (2014) The genetic origin of honey bee colonies used in the COLOSS genotype-environment interactions experiment: a comparison of methods. *Journal of Apicultural Research*, **53**, 188–204.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Franck P, Garnery L, Loiseau A et al. (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Fuller ZL, Niño EL, Patch HM et al. (2015) Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools. *BMC Genomics*, **16**, 518.

Galindo-Cardona A, Acevedo-González JP, Rivera-Marchand B, Giray T (2013) Genetic structure of the gente Africanized honey bee population (gAHB) in Puerto Rico. *BMC Genetics*, **14**, 65.

Gärke C, Ytournel F, Bed'hom B et al. (2012) Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Animal Genetics*, **43**, 419–428.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**, 145–154.

Garnery L, Franck P, Baudry E, Vautrin D, Cornuet JM, Solignac M (1998) Genetic diversity of the West European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, 49–74.

Glover KA, Hansen MM, Lien S, Als TD, Høyheim B, Skaala O (2010) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics*, **11**, 2.

Haasl RJ, Payseur BA (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, **106**, 158–171.

Hall HG (1990) Parental analysis of introgressive hybridization between African and European honeybees using nuclear DNA RFLPs. *Genetics*, **125**, 611–621.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron*, **4**, art. 4.

Harpur BA, Minaei S, Kent CF, Zayed A (2012) Management increases genetic diversity of honey bees via admixture. *Molecular Ecology*, **21**, 4414–4421.

Harpur BA, Kent CF, Molodtsova D et al. (2014) Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences of the USA*, **111**, 2614–2619.

Harpur BA, Chapman NC, Krimus L et al. (2015) Assessing patterns of admixture and ancestry in Canadian honey bees. *Insectes Sociaux*, **62**, 479–489.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, **11**, 150–161.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Herráez DL, Schafer H, Mosner J, Fries HR, Wink M (2005) Comparison of microsatellite and single nucleotide polymorphism markers for the genetic analysis of a Galloway cattle population. *Journal of Biosciences*, **60**, 637–643.

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, **11**, 137–149.

Hunt GJ, Page RE (1995) Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics*, **139**, 1371–1382.

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.

Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, **14**, 93–106.

Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity*, **88**, 62–65.

Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology Notes*, **5**, 187–189.

Karlsson EK, Baranowska I, Wade CM et al. (2007) Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nature Genetics*, **39**, 1321–1328.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.

Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, **6**, S26.

Livingstone DS, Motamayor JC, Schnell RJ et al. (2011) Development of single nucleotide polymorphism markers in *Theobroma cacao* and comparison to simple sequence repeat markers for genotyping of Cameroon clones. *Molecular Breeding*, **27**, 93–106.

McMahon DP, Natsopoulou ME, Doublet V et al. (2016) Elevated virulence of an emerging viral genotype as a driver of honeybee loss. *Proceedings of the Royal Society B*, **283**, 2016081.

Meixner MD, Leta MA, 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.

Meixner MD, Pinto MA, Bouga M, Kryger P, Ivanova E, Fuchs S (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *Journal of Apicultural Research*, **52**, 1–28.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the Western Europe. *Apidologie*, **38**, 141–155.

Moore JS, Bourret V, Dionne M et al. (2014) Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology*, **23**, 5680–5697.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution, and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Muñoz I, Dall'Olio R, Lodesani M, De la Rúa P (2009) Population genetic structure of coastal Croatian honeybees (*Apis mellifera carnica*). *Apidologie*, **40**, 617–626.

Muñoz I, Madrid-Jiménez MJ, De la Rúa P (2012) Temporal genetic analysis of an introgressed island honey bee population (Tenerife, Canary Islands, Spain). *Journal of Apicultural Research*, **51**, 144–146.

Muñoz I, Cepero A, Pinto MA, Martín-Hernández R, Higes M, De la Rúa P (2014a) Presence of *Nosema ceranae* associated with honeybee queen introductions. *Infection, Genetics and Evolution*, **23**, 161–168.

Muñoz I, Pinto MA, De la Rúa P (2014b) Effects of queen importation on the genetic diversity of island honey bee populations (*Apis mellifera* Linneaus 1758). *Journal of Apicultural Research*, **53**, 296–302.

Muñoz I, Dall'Olio R, Lodesani M, De la Rúa P (2014c) Estimating introgression in *Apis mellifera siciliana* populations: are the conservation islands really effective? *Insect Conservation and Diversity*, **7**, 563–571.

Muñoz I, Henriques D, Johnston JS, Chavez-Galarza J, Kryger P, Pinto MA (2015) Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*). *PLoS ONE*, **10**, e0124365.

Mutinelli F (2011) The spread of pathogens through trade in honey bees and their products (including queen bees and semen): overview and recent developments. *Revue scientifique et technique (International Office of Epizootics)*, **30**, 257–271.

Narum SR, Banks M, Beacham TD *et al.* (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, **17**, 3464–3477.

Nedić N, Francis RM, Stanisavljević L *et al.* (2014) Detecting population admixture in the honey bees of Serbia. *Journal of Apicultural Research*, **53**, 303–313.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, USA.

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**, 116–129.

Ozerov M, Vasemägi A, Wennevik V *et al.* (2013) Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS ONE*, **8**, e82434.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics*, **28**, 2537–2539.

Péntek-Zakar E, Oleksa A, Borowik T, Kusza S (2015) Population structure of honey bees in the Carpathian Basin (Hungary) confirms introgression from surrounding subspecies. *Ecology and Evolution*, **5**, 5456–5467.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Pinto MA, Rubink WL, Patton JC, Coulson RN, Johnston JS (2005) Africanization in the United States: replacement of Feral European Honeybees (*Apis mellifera* L.) by an African Hybrid Swarm. *Genetics*, **170**, 1653–1665.

Pinto MA, Henriques D, Chavez-Galarza J *et al.* (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**, 269–278.

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution*, **25**, 345–353.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puckett EE, Eggert LS (2016) Comparison of SNP and microsatellite genotyping panels for spatial assignment of individuals to natal range: a case study using the American black bear (*Ursus americanus*). *Biological Conservation*, **193**, 86–93.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *The American Journal of Human Genetics*, **81**, 559–575.

Rangel J, Giresi M, Pinto MA *et al.* (2016) Africanization of a feral honey bee (Apis mellifera) population in South Texas: does a decade make a difference? *Ecology and Evolution*, **6**, 2158–2169.

Rašić G, Filipovic I, Weeks AR, Hoffmann AA (2014) Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics*, **15**, 275.

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *The American Journal of Human Genetics*, **73**, 1402–1422.

Ross CT, Weise JA, Bonnar S *et al.* (2014) An empirical comparison of short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) for relatedness estimation in Chinese rhesus macaques (*Macaca mulatta*). *American Journal of Primatology*, **76**, 313–324.

Ruttner F (1988) *Biogeography and Taxonomy of Honey Bees*. Springer, Berlin.

Schopen GCB, Bovenhuis H, Visker MHPW, Van Arendonk JAM (2008) Comparison of information content for microsatellites and SNPs in poultry and cattle. *Animal Genetics*, **39**, 451–453.

Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, **34**, 367–375.

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.

Spötter A, Gupta P, Nürnberg G, Reinsch N, Bienefeld K (2012) Development of a 44K SNP assay focussing on the analysis of a varroa-specific defence behaviour in honey bees (*Apis mellifera carnica*). *Molecular Ecology Resources*, **12**, 323–332.

Strange JP, Garnery L, Sheppard WS (2008) Morphological and molecular characterization of the Landes honey bee (*Apis mellifera* L.) ecotype for genetic conservation. *Journal of Insect Conservation*, **12**, 527–537.

Suazo A, Hall HG (1999) Modification of the AFLP protocol applied to honey bee (*Apis mellifera* L.) DNA. *BioTechniques*, **26**, 704–705.

Sušnik S, Kozmus P, Poklukar J, Meglic V (2004) Molecular characterisation of indigenous *Apis mellifera carnica* in Slovenia. *Apidologie*, **35**, 623–636.

Thalamuthu A, Mukhopadhyay I, Ray A, Weeks DE (2004) A comparison between microsatellite and single-nucleotide polymorphism markers with respect to two measures of information content. *BMC Genetics*, **6**, S27.

Tokarska M, Marshall T, Kowalczyk R *et al.* (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity*, **103**, 326–332.

Uzunov A, Meixner MD, Kiprijanovska H *et al.* (2014) Genetic structure of *Apis mellifera macedonica* in the Balkan Peninsula based on microsatellite DNA polymorphism. *Journal of Apicultural Research*, **53**, 288–295.

Väli U, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808–3817.

Väli Ü, Saag P, Dombrovski V *et al.* (2010) Microsatellites and single nucleotide polymorphisms in avian hybrid identification: a comparative case study. *Journal of Avian Biology*, **41**, 34–49.

Van Engelsdorp D, Meixner MD (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**, 80–95.

Vignal A, Milan D, Sancristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, **34**, 275–305.

Wallberg A, Han F, Wellhagen G *et al.* (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.

Wallberg A, Glémin S, Webster MT (2015) Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. *PLoS Genetics*, **11**, e1005189.

Whitfield CW, Behura SK, Berlocher SH *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.

Wragg D, Marti-Marimon M, Basso B *et al.* (2016) Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly. *Scientific Reports*, **6**, 27168.

Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of USA*, **105**, 3421–3426.

Data accessibility

Microsatellite and SNP genotypes are deposited in Dryad: http://dx.doi.org/10.5061/dryad.5vp20.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Diversity measures estimated by the 11 microsatellites and the six different SNP data sets for each of three subspecies *A. m. mellifera*, *A. m. ligustica* and *A. m. carnica*. Na - mean number of alleles, Ne - effective number of alleles, uh - unbiased diversity, and Rs - allelic richness

Table S2 *P*-values obtained with the Tukey test for the comparisons of the diversity measures between marker types and data sets of the three subspecies *A. m. mellifera*, *A. m. ligustica* and *A. m. carnica*

Table S3 Membership proportions (*Q*-values) in the C-lineage cluster, inferred from STRUCTURE, for the 77 *A. m. mellifera* individuals sampled in western Europe (M-lineage) and for the 17 *A. m. ligustica* and 19 *A. m. carnica* individuals sampled in eastern Europe (C-lineage). Sampling locations of *A. m. mellifera* marked in bold represent unprotected apiaries; the remaining sampling locations represent apiaries under conservation management. The 113 individuals are ordered as in Fig. S1

(Supporting information)

Table S4 *P*-values obtained with the Tukey test for pairwise comparisons of C-lineage introgression proportions into *A. m. mellifera*

Fig. S1 Clusters identified by the Bayesian-based software STRUCTURE (Pritchard *et al.* 2000) for 77 individuals sampled in the native range of the M-lineage honeybee subspecies *A. m. mellifera* in western Europe and the 36 individuals sampled in the native range of the C-lineage subspecies *A. m. ligustica* (N = 17) and *A. m. carnica* (N = 19) in eastern Europe. Individual membership proportions (*Y*-axis) were inferred from **(a)** 11 microsatellites, **(b)** 55 closest SNPs, **(c)** 110 closest SNPs, **(d)** 48 AIMs, **(e)** 96 AIMs, **(f)** 144 AIMs, and **(g)** reference SNP data set. Each individual is represented by a bar, which is partitioned into two coloured segments that represent the individual's estimated membership proportions in K = 2 optimal number of clusters, as determined by the ΔK method (Evanno *et al.* 2005)

Fig. S2 Determination of the optimal number of genetic clusters (K) using the ΔK method (Evanno *et al.* 2005) for **(a)** 11 microsatellites, **(b)** 55 closest SNPs, **(c)** 110 closest SNPs, **(d)** 48 AIMs, **(e)** 96 AIMs, **(f)** 144 AIMs, and **(g)** reference SNP sets

Fig. S3 Differences of *Q*-values inferred from the six data sets in relation to those of the reference SNP set for each of the 77 individuals sampled in the native range of the M-lineage honeybee subspecies *A. m. mellifera* in western Europe (see *Q*-values in Table S3, Supporting information). The six data sets included the 11 microsatellites (11 STRs) and the reduced SNP panels containing the 5 and 10 flanking SNPs of each microsatellite locus (55 and 110 closest SNPs) and the top-ranked informative SNPs (48, 96, and 144 AIMs). Individuals (ID 1 to 77) are ordered as in Fig. S1 (Supporting information), which shows introgression proportions and geographical origin of the 77 individuals

Fig. S4 C-lineage introgression into *A. m. mellifera* inferred with STRUCTURE from the 11 microsatellites plotted as linear regressions against the **(a)** 55 closest SNPs, **(b)** 110 closest SNPs, **(c)** 48 AIMs, **(d)** 96 AIMs, **(e)** 144 AIMs, and **(f)** reference SNP sets; and from the reference SNPs against the **(g)** 55 closest SNPs, **(h)** 110 closest SNPs, **(i)** 48 AIMs, **(j)** 96 AIMs, **(k)** and 144 AIMs