MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 1346–1355 doi: 10.1111/1755-0998.12411

A SNP test to identify Africanized honeybees via proportion of 'African' ancestry

NADINE C. CHAPMAN,* BROCK A. HARPUR,† JULIANNE LIM,* THOMAS E. RINDERER,‡ MICHAEL H. ALLSOPP,§ AMRO ZAYED† and BENJAMIN P. OLDROYD*

Behaviour and Genetics of Social Insects Lab, School of Biological Sciences A12, University of Sydney, Sydney, NSW 2006, Australia, †Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3, ‡Honey-bee Breeding Genetics and Physiology Research Laboratory, USDA-ARS, 1157 Ben Hur Road, Baton Rouge, LA 70820, USA, §ARC-Plant Protection Research Institute, Stellenbosch 7599, South Africa

Abstract

The honeybee, *Apis mellifera*, is the world's most important pollinator and is ubiquitous in most agricultural ecosystems. Four major evolutionary lineages and at least 24 subspecies are recognized. Commercial populations are mainly derived from subspecies originating in Europe (75–95%). The Africanized honeybee is a New World hybrid of *A. m. scutellata* from Africa and European subspecies, with the African component making up 50–90% of the genome. Africanized honeybees are considered undesirable for bee-keeping in most countries, due to their extreme defensiveness and poor honey production. The international trade in honeybees is restricted, due in part to bans on the importation of queens (and semen) from countries where Africanized honeybees are extant. Some desirable strains from the United States of America that have been bred for traits such as resistance to the mite *Varroa destructor* are unfortunately excluded from export to countries such as Australia due to the presence of Africanized honeybees in the USA. This study shows that a panel of 95 single nucleotide polymorphisms, chosen to differentiate between the African, Eastern European and Western European lineages, can detect Africanized honeybees with a high degree of confidence via ancestry assignment. Our panel therefore offers a valuable tool to mitigate the risks of spreading Africanized honeybees across the globe and may enable the resumption of queen and bee semen imports from the Americas.

Keywords: Africanized honeybee, ancestry, *Apis mellifera*, breed identification, single nucleotide polymorphisms

Received 23 September 2014; revision received 26 March 2015; accepted 31 March 2015

Introduction

Pollination of crops by honeybees adds billions of dollars to the world economy (Gallai *et al.* 2009), with 84% of crops dependent on or benefitting from insect pollinators (Williams 1994) and one-third of all food production partially or fully dependent on insect pollination (Klein *et al.* 2007). It is therefore essential that threats to honeybees are identified and strategies that prevent losses to the bee-keeping industry are developed. The ability to identify and differentiate between undesirable and desirable honeybee genetic lineages is likely to aid breeding programmes throughout the world.

There are four (or perhaps five) major evolutionary lineages of the Western honeybee (*Apis mellifera*): A (African), M (West European), C (East European) and O (Middle Eastern) (Ruttner 1988; Garnery *et al.* 1992, 1993; Arias & Sheppard 1996; Franck *et al.* 2000; Palmer *et al.* 2000; Whitfield *et al.* 2006; Wallberg *et al.* 2014) and at least 24 named subspecies (Ruttner 1988). Honeybees were introduced into the New World and Australasia by European settlers. In North, South and Central America, Australia and New Zealand, early introductions were mostly of *A. m. mellifera* and *A. m. iberiensis* (*iberica*) from the M lineage (Hopkins 1886; Seeley 1985; Cornuet 1986). Throughout the 20th century, honeybees from the Eastern European (in particular *A. m. ligustica* and *A. m. carnica*) and Middle Eastern (in particular *A. m. caucasica*) lineages were repeatedly introduced to North America and Australasia, as these subspecies are favoured by commercial bee-keepers (Hopkins 1886; Ruttner 1975, 1988; Seeley 1985; Cornuet 1986). Experimental introductions were also made from Africa and the Middle East, leaving a genetic legacy in the honeybees of North America (e.g. Schiff & Sheppard 1993; Sheppard & Smith 2000)

Correspondence: Nadine C. Chapman;
E-mail: nadine.chapman@sydney.edu.au

and likely in Australia (Goodacre 1935; Weatherhead 1986). However, commercial honeybees in Europe, North America and Australia are predominantly (in the order of 75–95%) of Eastern and Western European origin (Whitfield *et al.* 2006; Chapman *et al.* 2008; Oxley & Oldroyd 2009; Harpur *et al.* 2012; Pinto *et al.* 2014; Wallberg *et al.* 2014).

In 1956, *A. m. scutellata* of the African lineage was introduced into Brazil from South Africa and Tanzania. It was hoped that this subspecies, adapted to subtropical and tropical savannah, would perform better in São Paulo state than the existing introduced honeybees that had evolved in Europe (Kerr 1967). Famously, queens and drones escaped from the breeding programme and hybridized with the existing population (Winston 1992b). These hybrid 'Africanized' honeybees have since spread as far south as Argentina and north into the south-western states of the USA (Winston 1992a,b), and their range continues to expand (Schneider *et al.* 2004; Harrison *et al.* 2006; Jarnevich *et al.* 2014). Africanized honeybees are largely of African ancestry in South America (70–90%) and the USA (50–75%) (Whitfield *et al.* 2006; Wallberg *et al.* 2014).

Africanized honeybees abscond more frequently than European subspecies (Winston 1992a), have a propensity for pollen rather than nectar collection (Danka *et al.* 1987; Winston 1992a; Fewell & Bertram 2002) and are regarded as poor honey producers (Rinderer *et al.* 1984, 1985, 1986; Pesante *et al.* 1992; Guzmán-Novoa & Uribe-Rubio 2004) although this is not always the case (Spivak *et al.* 1989; Pereira & Chaud-Netto 2005; Zárate *et al.* 2008; Livanis & Moss 2010). These factors, but particularly their heightened defensiveness (Collins *et al.* 1982; Breed *et al.* 2004), make them unpopular for commercial bee-keeping (Winston 1992a,b; Rinderer *et al.* 1993b; Schneider *et al.* 2004). Commercial bee-keeping relies on high densities of colonies and regular migration of colonies over long distances on public highways to take advantage of ephemeral honey flows and crops that require pollination. Neither migration nor high colony densities are compatible with extreme aggression. For these reasons, a number of countries, notably Canada and Australia, have limitations on honeybee imports due to concerns over Africanized honeybees.

Morphometric measurements can appropriately separate highly Africanized honeybees from non-Africanized honeybees (Daly & Balling 1978; Rinderer *et al.* 1986, 1990, 1993a; Francoy *et al.* 2006, 2008, 2009). Morphometric measures are unable, however, to detect low to medium levels of Africanization, with 43% of hybrid samples being misidentified as European (Guzmán-Novoa *et al.* 1994). Moreover, morphometric measurements cannot be conducted on honeybee semen, which is commonly seen as a low-risk alternative to importing live honeybees that may carry *Varroa* and other parasitic mites. Mitochondrial DNA sequencing is also seen as an unreliable test for Africanization. Mitochondria are maternally inherited without recombination (White *et al.* 2008), and so the offspring of a European queen mated to Africanized drones will be falsely classified as European using mitochondrial DNA sequencing (Sheppard & Smith 2000; Meixner *et al.* 2013), as will each subsequent generation arising from that mating.

Currently, there is no reliable low-cost genetic test for detecting Africanized honeybees available. However, Whitfield *et al.* (2006) were able to clearly distinguish between Africanized honeybees and managed European-derived honeybees using 1136 SNPs (single nucleotide polymorphisms), suggesting that it should be possible to develop a low-cost SNP test suitable for use by industry and biosecurity officials. Such tests have been successfully developed to differentiate breeds of swine (Ramos *et al.* 2011; Wilkinson *et al.* 2012), cattle (Negrini *et al.* 2009; Dimauro *et al.* 2013; Gurgul *et al.* 2013; Mancini *et al.* 2014), sheep (Heaton *et al.* 2014) and salmon (Freamo *et al.* 2011) and to track introgression of domesticated breeds into wild boar (Goedbloed *et al.* 2013) and wild salmon (Karlson *et al.* 2011) populations. The recent explosion of honeybee SNP studies (Whitfield *et al.* 2006; Zayed & Whitfield 2008; Dixon *et al.* 2012; Harpur *et al.* 2012, 2014; Shorter *et al.* 2012; Spotter *et al.* 2012; Chávez-Galarza *et al.* 2013; Pinto *et al.* 2014; Wallberg *et al.* 2014) means that millions of honeybee SNPs are now available (Harpur *et al.* 2014; Wallberg *et al.* 2014) that are potentially effective in separating honeybee lineages (Whitfield *et al.* 2006; Harpur *et al.* 2014; Wallberg *et al.* 2014). These SNPs provide the opportunity to develop a simple SNP-based test for detecting the proportion of African ancestry in any individual honeybees.

Here, we present a panel of just 95 SNPs that were found to be under selection in Africanized populations (Whitfield *et al.* 2006) or with high pairwise population F_{ST} between three ancestral lineages (Harpur *et al.* 2014). Differences in the proportion of African ancestry between these groups thus enable differentiation of African and Africanized honeybees from European-derived honeybees. There are a number of subspecies in Africa, and while we use 'African' ancestry to delineate Africanized honeybees from European honeybees, our current test cannot differentiate between the African subspecies. Given the high rates of 'African' alleles in Africanized populations compared to other populations (50–90% vs 5–25%; Whitfield *et al.* 2006; Harpur *et al.* 2012; Wallberg *et al.* 2014), we use the proportion of African ancestry as a proxy for Africanization. First, we demonstrate that using 95 SNPs, we can effectively differentiate reference samples of subspecies from Africa (*A. m. scutellata*) and Europe (*A. m. ligustica*, *A. m. carnica*, *A. m. mellifera* and

A. m. iberiensis). We then use the SNP panel to estimate the proportion of African ancestry in honeybees from Australian commercial and feral populations, North American commercial and *Varroa*-resistant lines and Africanized honeybees from Brazil and North America. By genotyping both Africanized honeybees and commercial stocks, we determined the effectiveness of our panel at classifying samples as 'Africanized'. Ultimately, we generate a panel that can act as a reliable test of Africanized honeybees from the Americas that has the potential to allow the resumption of international trade in honeybees and honeybee semen into Australia and elsewhere. The panel may also be a cost-effective and efficient means of utilizing SNPs for identifying ancestry of populations, for example on Pacific Islands where the ancestry of introduced populations is not always known.

Methods

Population sampling

Reference samples. Our reference samples came from three of the five major evolutionary lineages. The Eastern European lineage (C) was represented by *A. m. ligustica* ($n = 77$) and *A. m. carnica* ($n = 9$). The Western European lineage (M) was represented by *A. m. iberiensis* ($n = 4$) and *A. m. mellifera* ($n = 9$). The African (A) lineage was represented by *A. m. scutellata* ($n = 128$) (Appendix S1, Supporting information; Rinderer *et al.* 1993a; Oldroyd *et al.* 2011; Harpur *et al.* 2012, 2014). The Harpur *et al.* samples were collected in 2012 and assigned on the basis of present-day distributions and by the experts who collected them in the field and confirmed genetically (Harpur *et al.* 2012). The other European samples were collected between 1989 and 1993 as part of Rinderer *et al.*'s (1993a) work on morphometric identification of honeybee subspecies. The other *A. m. scutellata* samples were collected from South Africa and Botswana between 1984 and 2013 (Appendix S1, Supporting information; Rinderer *et al.* 1993a; Oldroyd *et al.* 2011, 2014).

Test samples. European-derived populations were represented by commercial ($n = 104$) and feral ($n = 102$) populations from Australia (Chapman *et al.* 2008; Hinson *et al.* in press) and commercial populations from Canada ($n = 10$) and the USA ($n = 63$) and by three *Varroa*-resistant strains from the USA collected in 2013 ($n = 58$) (Appendix S1, Supporting information).

Africanized populations were represented by samples from Brazil ($n = 55$) sampled in 1993 (Rinderer *et al.* 1993a; Clarke *et al.* 2001) and an unmanaged Africanized USA population from Texas ($n = 86$) in 2013 (Appendix S1, Supporting information).

Finally, we included test samples from the parasitic *A. m. capensis* clonal lineage ($n = 3$), *A. m. capensis* ($n = 104$) and *scutellata–capensis* hybrids ($n = 17$) (Appendix S1, Supporting information; Oldroyd *et al.* 2011, 2014) as these honeybees are closely related to *A. m. scutellata* and are also highly undesirable in commercial beekeeping due to the ability of workers to parasitize and overrun colonies by producing clones of themselves (Beekman *et al.* 2008). Subspecies within Africa were assigned on the basis of where the sample had been collected in relation to the predefined subspecies and hybrid zones (Ruttner 1988; Hepburn & Crewe 1990; Dietemann *et al.* 2007; Beekman *et al.* 2008; Goudie & Oldroyd 2014).

SNP panel selection and genotyping

Harpur *et al.* (2014) sequenced over 40 whole genomes of honeybees from Africa, Europe and the Middle East. More than 20 000 SNPs identified in this study were found to have high pairwise F_{ST} (Weir & Cockerham 1984) between populations (African, Eastern European and Western European). The effectiveness of selecting genetic markers for population differentiation based on pairwise F_{ST} values is well established (e.g. Reed 1973; Chakraborty *et al.* 1992; Stephens *et al.* 1994; Karlson *et al.* 2011). While other methods are available, they are highly correlated with F_{ST} (Wilkinson *et al.* 2011). We randomly chose 1046 SNPs with high pairwise F_{ST} between populations from this list with the condition that SNPs be greater than 5000 bp apart (Harpur *et al.* 2014). To this list, we added 19 SNPs from a previous study that were hypothesized to be under selection in populations experiencing Africanization and are therefore likely to be important in this population and assist in identification of Africanized honeybees (table S3 in Whitfield *et al.* 2006). From this combined set of 1065 SNPs, we determined which could be multiplexed in an inexpensive SNP genotyping platform using the Sequenom ASSAY DESIGN SUITE (v1.0 Sequenom, CA, USA), which picks SNPs so as to optimize multiplex conditions, for example by avoiding hairpin and dimer formation. With user-assigned priority given to the SNPs from Whitfield *et al.* (2006), the software chose 144 SNPs for amplification in four multiplexes.

High molecular weight DNA was extracted from the tergite (with internal material removed) of one individual per sampled colony with phenol–chloroform–isoamyl alcohol (25:24:1) (Sambrook *et al.* 1989). Individuals were either haploid males or diploid females. Individuals were genotyped at 144 SNPs using the Sequenom MassARRAY MALDI-TOF system (Sequenom, CA, USA). Genotype calling was performed automatically by Sequenom MASSARRAY TYPER software (Sequenom, CA,

USA) using 'moderate parameters' (i.e. allele calls with 'conservative' and 'moderate' descriptions included, 'aggressive' calls excluded) (Sequenom 2006). Reference samples taken from Harpur *et al.* (2014) were not regenotyped.

Population genetics analyses

We used GENEPOP v4.2 (Raymond & Rousset 1995) to determine the level of genetic differentiation (F_{ST}) between populations. An unrooted UPGMA tree was constructed from the F_{ST} matrix using NEIGHBOR in PHYLIP (Felsenstein 1989). Markers were included for analysis if they were successfully typed in more than two-thirds of individuals and individuals were included if they were typed at more than two-thirds of markers. Only markers with minor allele frequencies >5% across all samples were used (Dimauro *et al.* 2013; Goedbloed *et al.* 2013; Mancini *et al.* 2014). This left us with 95 SNPs with minor allele frequencies ranging between 0.08 and 0.496 (Appendix S2, Supporting information). The closest analysed SNPs were 45 945 bp apart and the average was 1 734 863 bp. The honeybee has a very high recombination rate (19 cM per megabase; Beye *et al.* 2006) and the average distance at which linkage decays is 500 bp (Wallberg *et al.* 2014). Therefore, it is highly unlikely that any possible pair of SNPs used in this analysis is linked.

We evaluated the population structure of our samples in STRUCTURE (v2.3.4, Pritchard *et al.* 2000). Haploid male data were coded as missing data at one allele for each marker as per the manual (Pritchard *et al.* 2010). We used a burn-in phase of 50 000 iterations with individuals from the three reference populations (Eastern European (C), Western European (M) and African (A)). We did not identify from which lineage our reference individuals arose. Ancestry of individuals from the test populations was assigned according to an admixture model with uncorrelated allele frequencies in 100 000 Markov chain Monte Carlo iterations based on SNP frequencies in the reference populations. No a priori information was provided regarding population identity or location in our test populations. Allele frequencies were calculated based only on the reference populations (Appendix S2, Supporting information). We performed 5 replicates for each of k = 1–6 populations. We inferred the optimal k using both LnP(D) from STRUCTURE (Pritchard *et al.* 2000) and ΔK (Evanno *et al.* 2005) (Appendix S3, Supporting information).

Accuracy of SNP panel

We determined the functionality of the SNP panel by considering how many individuals from the test samples would be correctly excluded or included at different thresholds of African ancestry (% of SNPs present in our *A. m. scutellata* reference population) in increments of 5% (10–30%) for each population. For each threshold (10%, 15%, 20%, 25%, 30%) of the proportion of African ancestry that was declared acceptable, we counted the number of misclassified individuals. To determine the overall false-negative rate (Africanized individuals declared to be non-Africanized), we counted the number of misclassified Africanized and African (*A. m. capensis*, clone and *scutellata–capensis* hybrid) individuals from the test population and divided it by the total number of African and Africanized individuals in the test population. To determine the overall false-positive rate (European-derived individuals from the test population declared Africanized), we counted the number of misclassified European-derived honeybees and divided it by the total number of European-derived individuals in the test population.

Results

Our data set comprises individuals genotyped at 95 SNPs for three reference populations and 10 test populations, with 829 individuals tested in total. Our three reference populations were well separated; the pairwise F_{ST} between our African (A; *A. m. scutellata*) and Eastern European (C; *A. m. ligustica* and *A. m. carnica*) was 0.855, while that between African and Western European (M; *A. m. mellifera* and *A. m. iberiensis*) was 0.859 (Fig. 1; Table 1). The two European reference populations were similarly well separated (F_{ST} = 0.867; Fig. 1; Table 1). There was little differentiation between the Africanized test populations from USA and Brazil (F_{ST} = 0.058), between the three European-derived North American test populations (Canada commercial, commercial USA and *Varroa*-resistant USA; F_{ST} = 0.020–0.023), the two Australian test populations (commercial and feral; F_{ST} = 0.032) or the South African test populations of *A. m. scutellata, A. m. capensis, scutellata–capensis* hybrids or the clonal parasite (F_{ST} = 0–0.050; Fig. 1; Table 1). The North American (Canada commercial, USA commercial and *Varroa*-resistant) and Australian commercial test populations were very similar (F_{ST} = 0.040–0.055), but the North American populations were quite distant to the Australian feral population (F_{ST} = 0.115–0.139; Table 1).

Analysis using STRUCTURE supported models with k = 3 ancestral populations (average ln[P(D)] = −48 709.9, posterior probability = 1; ΔK = 700.9; Fig. 2; Appendices S3 and S4, Supporting information). Where k = 3, the three reference populations corresponded to the three honeybee evolutionary lineages: African, Western European and Eastern European. Individuals from the Africa reference population had a minimum 94.1% alleles classified as 'African', individuals from the

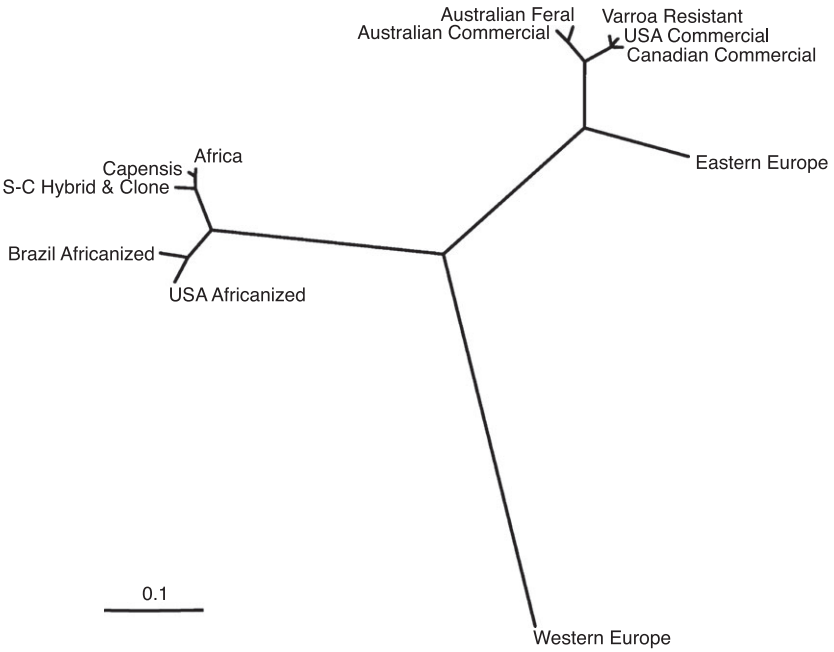

Fig. 1 An unrooted UPGMA tree showing the genetic distance between the 3 reference populations (Africa (A; *A. m. scutellata*), Eastern Europe (C; *A. m. ligustica* and *A. m. carnica*) and Western Europe (M; *A. m. mellifera* and *A. m. iberiensis*)) and the 10 test populations. S-C Hybrid = *scutellata-capensis* hybrid.

Eastern Europe reference population had a minimum 93.4% alleles classified as 'Eastern European', and individuals from the Western Europe reference population had a minimum 99.1% alleles classified as 'Western European'.

Individuals from the Africanized population in Brazil carried African alleles at 80.5% of markers on average, with 17.8% of alleles from Western Europe and 1.7% from Eastern Europe. The Africanized population in the USA carries fewer African alleles on average (62.5%) and more Western European (21.4%) and Eastern European (16.1%) alleles than did the Brazilian samples. There was one outlier Africanized individual from the USA that carried a low proportion of African alleles (19.6%). All the European-derived populations in Canada, Australia and the USA carried a small proportion of alleles present in our African reference population (3.4–4.9% on average) with a greater range among individuals (0.3–32.8%; Fig. 2). On average, European-derived populations from North America have more Eastern European ancestry (*Varroa*-resistant 78.1%; USA commercial 81.1%; Canada 85.4%) than those from Australia (commercial 69.5%; feral 57.5%). Some Australian individuals from both the commercial and feral populations had a high proportion of Western European alleles (up to 92.0%; Fig. 2).

We investigated the number of individuals that would be misclassified if individuals from test samples carrying more than X% African alleles were declared African or Africanized, for values of X between 10 and 30% in 5% increments (Fig. 3; Appendix S5, Supporting information). The threshold that minimizes the number of false positives (European and European-derived individuals being declared Africanized) while preventing any false negatives (African or Africanized individuals being declared non-Africanized) occurs when individuals carrying more than 15% African alleles are rejected. At this threshold none of the 124 African (*A. m. capensis*, *scutellata–capensis hybrids* or clones) or 141 Africanized individuals (from USA and Brazil) were declared non-Africanized and the false-positive rate is 4.7%, with 16 of 337 European-derived test samples declared Africanized (Fig. 3; Appendix S5, Supporting information).

As a check to determine whether our panel represented the most informative design, we calculated the information content (I_n; Rosenberg *et al.* 2003) of the markers we selected. We compared the information content of our panel (mean $I\_n = 0.55 \pm 0.09$ SD) to the average information content of 10 000 possible panels designed from all other SNPs with high pairwise F_{ST} values ($I\_n = 0.496 \pm 0.008$). We found that although there are certainly more informative individual markers available, the set we have chosen is more informative that provides 99.9% of the possible random combinations of high Fst SNPs (permutation test $n = 10\,000$, $P < 0.001$; Appendix S6, Supporting information).

Discussion

Current methods of identifying Africanized individuals, such as morphometric analysis and mitochondrial haplotyping, are arguably ineffective and potentially unreliable (Guzmán-Novoa *et al.* 1994; Meixner *et al.* 2013). In contrast, SNP genotyping provides a means to overcome some of the pitfalls of more traditional methods and can

Table 1 Mean pairwise F_{ST} among the three reference populations (Africa, Eastern Europe and Western Europe) and the 10 test populations

	Eastern Europe	Western Europe	Africa	Scutellata-capensis hybrid	Capensis	Clone	Brazil Africanized	USA Africanized	USA commercial	Varroa resistant	Canada commercial	Australia commercial
Western Europe	0.867											
Africa	0.855	0.859										
Scutellata-capensis hybrid	0.840	0.866	0.042									
Capensis	0.860	0.870	0.014	0.050								
Clone	0.838	0.896	0.028	0.000	0.041							
Brazil Africanized	0.766	0.620	0.153	0.093	0.156	0.012						
USA Africanized	0.606	0.477	0.230	0.135	0.222	0.063	0.058					
USA commercial	0.141	0.659	0.731	0.628	0.723	0.589	0.577	0.409				
Varroa resistant	0.185	0.638	0.724	0.610	0.715	0.572	0.560	0.392	0.023			
Canada commercial	0.181	0.792	0.810	0.741	0.817	0.681	0.615	0.420	0.020	0.023		
Australia commercial	0.205	0.507	0.642	0.530	0.628	0.492	0.485	0.334	0.040	0.046	0.055	
Australia feral	0.396	0.430	0.675	0.519	0.662	0.443	0.448	0.275	0.122	0.115	0.139	0.032

be automated (Meixner *et al.* 2013). Here, we demonstrate that using only 95 SNPs (Appendix S2, Supporting information), we were able to effectively differentiate populations from Africa, Eastern and Western Europe and use this to accurately differentiate individuals with levels of African ancestry (African or Africanized) from non-Africanized populations with high repeatability. Where a threshold of 15% African ancestry is used to reject individuals for importation, all African (124) and Africanized (141) individuals from the test population were correctly identified and rejected, while there was an overall false-positive rate of 4.7% for European-derived New World populations (16 of 337 rejected). This degree of confidence should be sufficient for quarantine purposes, especially given that Africanized populations generally carry 50–90% 'African' alleles (Whitfield *et al.* 2006; Wallberg *et al.* 2014).

The Africanized populations from the Americas conformed to our expectations; individuals from Brazil have little ancestry from the Eastern European lineage, whereas Africanized honeybees from the USA still have some Eastern European ancestry. Over time European alleles are lost due to Africanization, in particular those from the Eastern lineage (Clarke *et al.* 2001, 2002; Schneider *et al.* 2004; Whitfield *et al.* 2006). It is likely that Africanized honeybees from the USA will become even easier to differentiate from the commercial population as the process of Africanization continues.

Honeybees from the commercial populations of Canada, the USA and Australia have between 0.3 and 32.8% African alleles on per individual basis with this test, similar to that found in previous tests completed with more SNPs (Whitfield *et al.* 2006; Wallberg *et al.* 2014). In contrast, Africanized populations from USA (19.2–80.4%) and Brazil (65.3–95.8%) carried significantly more African alleles. There is no evidence that *A. m. scutellata* was ever deliberately introduced into Australia, Canada or the USA. However, there is good evidence of early importations of honeybees from North Africa into both the US and Australia (Seeley 1985; Cornuet 1986; Weatherhead 1986). We suggest that these imports explain the presence of 'African' alleles, as assigned by STRUCTURE from our reference populations, in the feral and commercial populations of these countries. The result is that a small fraction of non-Africanized individuals from Australia, Canada and the USA might not meet our criteria of <15% African ancestry as our reference populations have defined it.

In conclusion, our SNP panel (Appendix S2, Supporting information) and reference data (Appendix S7, Supporting information) make it possible to detect the degree of African ancestry in honeybees and therefore to exclude undesirable Africanized individuals from

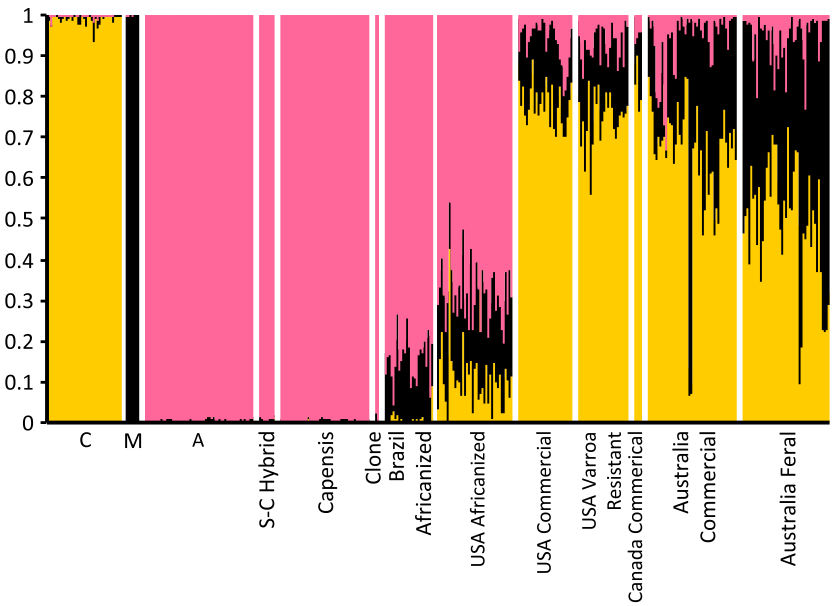

Fig. 2 Results of STRUCTURE analysis for $k = 3$ ancestral lineages: African (A, represented by *A. m. scutellata*) in pink, Western European (M, represented by *A. m. mellifera* and *A. m. iberiensis*) in black and Eastern European (C, represented by *A. m. ligustica* and *A. m. carnica*) in yellow in 13 populations. Each individual is represented by a single vertical bar with the inferred proportion of ancestry shared with each ancestral lineage.

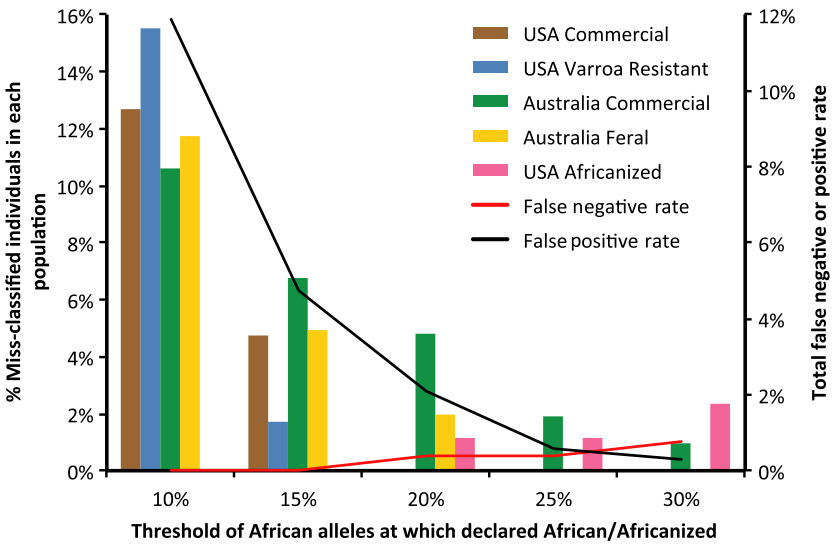

Fig. 3 The percentage of individuals misclassified in each population (bars) and the overall false-positive (European-derived individuals from the test population declared Africanized) and false-negative (Africanized and African individuals from the test population declared non-Africanized) rates (lines) when individuals carrying greater than X% African alleles are declared Africanized for X between 10 and 30% in increments of 5%. Some populations had no misclassified individuals and are thus not represented.

breeding programmes with high reliability. This is a major advance as currently a number of countries have restrictions on honeybee importations due to concerns over Africanization. The current restrictions limit the ability of countries such as Australia to import improved genetic stock, such as lines that are resistant to *Varroa*. The panel will also exclude *A. m. capensis* and all African honeybees from importation and will be useful for identifying African and Africanized honeybees intercepted on shipping and aircraft by biosecurity officials. We hope our panel will provide a new, reliable diagnostic for border control that will enable countries to screen imported bees for Africanization.

Acknowledgements

This project was supported by grants from Rural Industries Research and Development Corporation, Australia PRJ-007774 (BPO), and a Natural Sciences and Engineering Research Council Discovery grant (AZ) and an Ontario Ministry of Research and Innovation's Early Researcher Award (AZ). We thank Robert Cox and Amanda Frake (USDA) for collections of USA Africanized, *Varroa*-resistant and commercial USA samples, many Australian bee-keepers for contributing the Australian samples and Australian Department of Agriculture for samples that were imported to Australia from Canada. We thank the team at Australian Cancer Research Fund at the Garvan Institute for use of their facilities.

References

Arias MC, Sheppard WS (1996) Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution*, **5**, 557–566.

Beekman M, Allsopp MH, Wossler TC, Oldroyd BP (2008) Factors affecting the dynamics of the honeybee (*Apis mellifera*) hybrid zone of South Africa. *Heredity*, **100**, 13–18.

Beye M, Gattermeier I, Hasselmann M *et al.* (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.

Breed MD, Guzmán-Novoa E, Hunt GJ (2004) Defensive behavior of honey bees: organization, genetics and comparisons with other bees. *Annual Review of Entomology*, **49**, 271–298.

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE (1992) Caucasian genes in American blocks: new data. *American Journal of Human Genetics*, **50**, 145–155.

Chapman NC, Lim J, Oldroyd BP (2008) Population genetics of commercial and feral honey bees in Western Australia. *Journal of Economic Entomology*, **101**, 272–277.

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

Clarke KE, Oldroyd BP, Javier J, Quezada-Euán G, Rinderer TE (2001) Origin of honeybees (*Apis mellifera* L.) from the Yucatan peninsula inferred from mitochondrial DNA analysis. *Molecular Ecology*, **10**, 1347–1355.

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

Collins AM, Rinderer TE, Harbo JB, Bolten AB (1982) Colony defense by Africanized and European honey bees. *Science*, **218**, 72–74.

Cornuet J-M (1986) Population genetics. In: *Bee Genetics and Breeding* (ed. Rinderer TE), pp. 235–254. Academic Press, Orlando, Florida.

Daly HV, Balling SS (1978) Identification of Africanized honeybees in the Western Hemisphere by discriminant analysis. *Journal of the Kansas Entomological Society*, **51**, 857–869.

Danka RG, Hellmich RL II, Rinderer TE, Collins AM (1987) Diet-selection ecology of tropically and temperately adapted honey bees. *Animal Behaviour*, **35**, 1858–1863.

Dietemann V, Neumann P, Härtel S, Pirk CWW, Crewe RM (2007) Pheromonal dominance and the selection of a socially parasitic honeybee worker lineage (*Apis mellifera capensis* Esch.). *Journal of Evolutionary Biology*, **20**, 997–1007.

Dimauro C, Cellesi M, Steri R *et al.* (2013) Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. *Animal Genetics*, **44**, 377–382.

Dixon LR, McQuage MR, Lonon EJ *et al.* (2012) Pleiotropy of segregating genetic variants that affect honey bee worker life expectancy. *Experimental Gerontology*, **47**, 631–637.

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.

Felsenstein J (1989) PHYLIP - phylogeny inference package. *Cladistics*, **5**, 165–166.

Fewell JH, Bertram SM (2002) Evidence for genetic variation in worker task performance by African and European honey bees. *Behavioural Ecology and Sociobiology*, **52**, 318–325.

Franck P, Garnery L, Solignac M, Cornuet JM (2000) Molecular confirmation of a fourth lineage in honeybees from the Near East. *Apidologie*, **31**, 167–180.

Francoy TM, Prado PPR, Gonçalves LS, Costa LD, de Jong D (2006) Morphometric differences in a single wing cell can discriminate *Apis mellifera* racial types. *Apidologie*, **37**, 91–97.

Francoy TM, Wittmann D, Drauschke M *et al.* (2008) Identification of Africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie*, **39**, 488–494.

Francoy TM, Wittmann D, Steinhage V *et al.* (2009) Morphometric and genetic changes in a population of *Apis mellifera* after 34 years of Africanization. *Genetics and Molecular Research*, **8**, 709–717.

Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources*, **11**, 254–267.

Gallai N, Salles J-M, Settele J, Vaissiere BE (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, **68**, 810–821.

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, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Goedbloed DJ, Megens HJ, Van Hooft WF *et al.* (2013) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Molecular Ecology*, **22**, 856–866.

Goodacre WA (1935) The beginner in bee culture. In: *Farmer's Bulletin*, p. 91. Department of Agriculture, Sydney, New South Wales.

Goudie F, Oldroyd BP (2014) Theltyoky in the honey bee. *Apidologie*, **45**, 306–326.

Gurgul A, Rubis D, Zabek T *et al.* (2013) The evaluation of the usefulness of pedigree verification-dedicated SNPs for breed assignment in three Polish cattle populations. *Molecular Biology Reports*, **40**, 6803–6809.

Guzmán-Novoa E, Uribe-Rubio JL (2004) Honey production by European, Africanized and hybrid honey bee (*Apis mellifera*) colonies in Mexico. *American Bee Journal*, **144**, 318–320.

Guzmán-Novoa E, Page RE, Fondrk MK (1994) Morphometric techniques do not detect intermediate and low levels of Africanization in honey bee (Hymenoptera: Apidae) colonies. *Annals of the Entomological Society of America*, **87**, 507–515.

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, USA*, **111**, 2614–2619.

Harrison JF, Fewell JH, Anderson KE, Loper GM (2006) Environmental physiology of the invasion of the Americas by Africanized honeybees. *Integrative and Comparative Biology*, **46**, 1110–1122.

Heaton MP, Leymaster KA, Kalbfleisch TS *et al.* (2014) SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PLoS ONE*, **9**, e94851.

Hepburn HR, Crewe RM (1990) Defining the Cape honeybee: reproductive traits of queenless workers. *South African Journal of Science*, **86**, 524–527.

Hinson EM, Duncan M, Lim J, Oldroyd BP (in press) Density of feral honey bee (*Apis mellifera*) colonies in South Eastern Australia is greater in undisturbed than disturbed habitats. *Apidologie*. doi: 10.1007/s13592-01400334-x

Hopkins I (1886) *Illustrated Australasian Bee Manual and Complete Guide to Modern Bee Culture in the Southern Hemisphere*, 3rd edn. Issac Hopkins, Auckland, New Zealand.

Jarnevich CS, Esaias WE, Ma PLA *et al.* (2014) Regional distribution models with lack of proximate predictors: Africanized honeybees expanding north. *Diversity and Distributions*, **20**, 193–201.

Karlson S, Moen T, Lien S, Glovers KA, Hindar K (2011) Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular Ecology Resources*, **11**, 247–253.

Kerr WE (1967) The history of the introduction of Africanized honey bees to Brazil. *South African Bee Journal*, **39**, 3–5.

Klein AM, Vaissiere BE, Cane JH *et al*. (2007) Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B Biological Sciences*, **274**, 303–313.

Livanis G, Moss CB (2010) The effect of Africanized honey bees on honey production in the United States: an informational approach. *Ecological Economics*, **69**, 895–904.

Mancini G, Gargani M, Chillemi G *et al*. (2014) Signatures of selection in five Italian cattle breeds detected by a 54K SNP panel. *Molecular Biology Reports*, **41**, 957–965.

Meixner MD, Pinto MA, Bouga M *et al*. (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *Journal of Apicultural Research*, **52**, 27.

Negrini R, Nicoloso L, Crepaldi P *et al*. (2009) Assessing SNP markers for assigning individuals to cattle populations. *Animal Genetics*, **40**, 18–26.

Oldroyd BP, Allsopp MH, Lim J, Beekman M (2011) A thelytokous lineage of socially parasitic honey bees has retained heterozygosity despite at least 10 years of inbreeding. *Evolution*, **65**, 860–868.

Oldroyd BP, Allsopp MH, Roth KM *et al*. (2014) A parent-of-origin effect on honeybee worker ovary size. *Proceedings of the Royal Society B*, **281**, 1–7.

Oxley PR, Oldroyd BP (2009) Mitochondrial sequencing reveals five separate origins of 'Black' *Apis mellifera* (Hymenoptera: Apidae) in Eastern Australian commercial colonies. *Journal of Economic Entomology*, **102**, 480–484.

Palmer MR, Smith DR, Kaftanoglu O (2000) Turkish honeybees: genetic variation and evidence for a fourth lineage of *Apis mellifera* mtDNA. *Journal of Heredity*, **91**, 42–46.

Pereira AM, Chaud-Netto J (2005) Africanized honeybees: biological characteristics, urban nesting behavior and accidents caused in Brazilian cities (Hymenoptera: Apidae). *Sociobiology*, **46**, 535–550.

Pesante DG, Rinderer TE, Collins AM, Boykin DL, Buco SM (1992) Honey production in Venezuela: effects of feeding sugar syrup on colony weight gains by Africanized and European colonies. *Apidologie*, **23**, 545–552.

Pinto MA, Henriques D, Chávez-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.

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

Pritchard JK, Wen X, Falush D (2010) Documentation for *structure* software: Version 2.3, p. 7, http://pritchardlab.stanford.edu/structure_software/release_versions/v2.3.4/structure_doc.pdf.

Ramos AM, Megens HJ, Crooijmans RPMA, Schook LB, Groenen MAM (2011) Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. *Animal Genetics*, **42**, 613–620.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumeniscism. *Journal of Heredity*, **86**, 248–249.

Reed TE (1973) Number of gene loci required for accurate estimation of ancestral populations proportion in individual human hybrids. *Nature*, **244**, 575–576.

Rinderer TE, Bolten AB, Collins AM, Harbo JB (1984) Nectar-foraging characteristics of Africanized and European honeybees in the Neotropics. *Journal of Apicultural Research*, **23**, 70–79.

Rinderer TE, Collins AM, Tucker KW (1985) Honey production and underlying nectar harvesting activities of Africanized and European honeybees. *Journal of Apicultural Research*, **23**, 161–167.

Rinderer TE, Sylvester HA, Brown MA *et al*. (1986) Field and simplified techniques for identifying Africanized and European honey bees. *Apidologie*, **17**, 33–48.

Rinderer TE, Daly HV, Sylvester HA *et al*. (1990) Morphometric differences among Africanized and European honey bees and their FI hybrids (Hymenoptera, Apidae). *Annals of the Entomological Society of America*, **83**, 346–351.

Rinderer TE, Buco SM, Rubink WL *et al*. (1993a) Morphometric identification of Africanized and European honey bees using large reference populations. *Apidologie*, **24**, 569–585.

Rinderer TE, Oldroyd BP, Sheppard WS (1993b) Africanized bees in the United States. *Scientific American*, **269**, 84–90.

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

Ruttner F (1975) Races of bees. In: *The Hive and the Honey Bee* (ed. Sons Da), pp. 19–35. Dadant and Sons, Hamilton, Illinois.

Ruttner F (1988) *Biogeography and Taxonomy of Honeybees*. Springer-Verlag, Berlin, Germany.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, New York, New York.

Schiff NM, Sheppard WS (1993) Mitochondrial DNA evidence for the 19th century introduction of African honey bees into the United States. *Experientia*, **49**, 350–352.

Schneider SS, DeGrandi-Hoffman G, Smith DR (2004) The African honey bee: factors contributing to a successful biological invasion. *Annual Review of Entomology*, **49**, 351–376.

Seeley TD (1985) *Honeybee Ecology: A Study of Adaptation in Social Life*. Princeton University Press, Princeton, New Jersey.

Sequenom (2006) MassARRAY Typer 3.4 Software User's Guide for iPLEX and hME. p. 117. San Diego, California.

Sheppard WS, Smith DR (2000) Identification of African-derived bees in the Americas: a survey of methods. *Annals of the Entomological Society of America*, **93**, 159–176.

Shorter JR, Arechavaleta-Velasco M, Robles-Rios C, Hunt GJ (2012) A genetic analysis of the stinging and guarding behaviors of the honey bee. *Behavior Genetics*, **42**, 663–674.

Spivak M, Batra S, Segreda F, Castro AL, Ramirez W (1989) Honey production by Africanized and European honey bees in Costa Rica. *Apidologie*, **20**, 207–220.

Spotter A, Gupta S, Nurnberg G, Peinsch 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.

Stephens JC, Briscoe D, O'Brien S (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *American Journal of Human Genetics*, **55**, 809–824.

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.

Weatherhead T (1986) *Boxes to Bar Hives: Beekeeping History of Queensland*. International Colour Productions, Stanthorpe, Queensland.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

White DJ, Wolff JN, Pierson M, Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, **17**, 4925–4942.

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.

Wilkinson S, Wiener P, Archibald AL *et al*. (2011) Evaluation of approaches for identifying population informative markers from high density SNP chips. *BMC Genetics*, **12**, 45.

Wilkinson S, Archibald AL, Haley CS *et al*. (2012) Development of a genetic tool for product regulation in the diverse British pig breed marker. *BMC Genomics*, **13**, 580.

Williams IH (1994) The dependence of crop production within the European Union on pollination by honey bees. *Agricultural Zoology Reviews*, **6**, 229–257.

Winston ML (1992a) The biology and management of Africanized honey bees. *Annual Review of Entomology*, **37**, 173–193.

Winston ML (1992b) *Killer Bees: The Africanized Honey Bee in the Americas*. Harvard University Press, Cambridge, Massachusetts.

Zárate O, de Araujo-Freitas C, Medina LA, Velásquex A, Quezada-Euán G (2008) Phenotypic correlations of field and laboratory tests with

honey production in Africanized honey bees (*Apis mellifera*). *Apidologie*, **39**, 523–530.

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, USA*, **105**, 3421–3426.

B.P.O., A.Z. and N.C.C. designed the experiment. N.C.C. and J.L. extracted DNA. N.C.C. performed genotyping and data analysis. N.C.C., B.P.O., B.A.H. and A.Z. wrote the manuscript. A.Z. and B.A.H. provided genotype data for some reference population samples, and B.A.H. performed informativeness analysis. B.P.O., N.C.C., M.H.A. and T.E.R. provided samples.

Data Accessibility

Sampling locations (Appendix S1, Supporting information), marker information (Appendix S2, Supporting information) and genotypes of reference samples (Appendix S7, Supporting information) and instructions for ancestry assignment (Appendix S8, Supporting information) are available as supplementary material. Raw STRUCTURE, GENEPOP and PHYLIP files are available at http://hdl.handle.net/2123/12853.

Supporting Information

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

Appendix S1 List of samples and the populations and locations sampled from.

Appendix S2 SNP marker information for the 95 SNP panel.

Appendix S3 K inference.

Appendix S4 Triangle plot showing proportion of ancestry from the three ancestral lineages for each sample.

Appendix S5 Table of the number of individuals misclassified in each population.

Appendix S6 Informativeness.

Appendix S7 Reference data in format for use in STRUCTURE.

Appendix S8 Instructions in the use of STRUCTURE for implementing the ancestry test for Africanization.