Runs of homozygosity derived from pool-seq data reveal fine-scale population structures in Western honey bees (Apis mellifera)

Annik Gmel[1], Matthieu Guichard[1], Benjamin Dainat[2], Geoffrey Williams[3], Sonia Eynard[4], Bertrand Servin[4], Alain Vignal[4], the Beestrong Consortium[5], and Markus Neuditschko[1]

[1]Agroscope Location Posieux
[2]Agroscope Liebefeld Location
[3]Auburn University
[4]Universite Toulouse Jean Jaures
[5]Affiliation not available

March 07, 2024

Abstract

Runs of homozygosity (ROH) are continuous homozygous segments that arise through the transmission of haplotypes that are identical by descent (IBD). The length and distribution of ROH segments provide insight into the genetic diversity of populations and are useful to detect selection signatures. Here, we analysed pooled whole-genome sequencing data from 265 Western honey bee colonies from the two subspecies Apis mellifera mellifera and Apis mellifera carnica. Integrating individual ROH patterns and admixture levels in a high-resolution population network visualization allowed us to ascertain major differences between the two subspecies. Within A. m. mellifera, we identified well-defined substructures according to the genetic origin of the colonies and a fair amount of admixed colonies, despite the current applied conservation efforts. In contrast, A. m. carnica colonies were more inbred and could not be differentiated according to the geographical origin. We identified 29 coding genes in overlapping ROH segments within the two subspecies. Genes embedded in A. m. carnica specific homozygosity islands suggested a strong selection for production and behavioural traits, whilst the identified cuticula protein-coding genes (CPR3 and CPR4) were associated with their breed-specific stripe pattern. Local adaption of the two subspecies could be confirmed by the identification of two genes involved in the response to ultraviolet (UV) light. We demonstrated that colony genotypes derived from pooled honey bee workers are reliable to unravel the population dynamics in A. mellifera and provide fundamental information to conserve native honey bees.

Runs of homozygosity derived from pool-seq data reveal fine-scale population structures in Western honey bees *(Apis mellifera)*

Annik Imogen Gmel[1*], Matthieu Guichard[1,2*], Benjamin Dainat[2], Geoffrey Rhys Williams[3], Sonia Eynard[4,5], Alain Vignal[4,5], Bertrand Servin[4,5], the Beestrong Consortium[6]and Markus Neuditschko[1a]

[1] Animal GenoPhenomics, Animal Production Systems and Animal Health, Agroscope, Posieux, Switzerland

[2] Swiss Bee Research Centre, Agroscope, Liebefeld, Switzerland

[3] Department of Entomology and Plant Pathology, Auburn University, Auburn, United States of America

[4] GenPhySE, INRA, INPT, INPENVT, Université de Toulouse, Castanet-Tolosan, France

[5] UMT PrADE, Protection des Abeilles Dans L'Environnement, 84914 Avignon, France

[6] Domaine de Vilvert Bat 224, CS800009 78353 Jouy-en-Josas CEDEX, France

* these authors equally contributed to the manuscript

[a] corresponding author: markus.neuditschko@agroscope.admin.ch

Abstract

Runs of homozygosity (ROH) are continuous homozygous segments that arise through the transmission of haplotypes that are identical by descent (IBD). The length and distribution of ROH segments provide insight into the genetic diversity of populations and are useful to detect selection signatures. Here, we analysed pooled whole-genome sequencing data from 265 Western honey bee colonies from the two subspecies *Apis mellifera mellifera* and *Apis mellifera carnica* . Integrating individual ROH patterns and admixture levels in a high-resolution population network visualization allowed us to ascertain major differences between the two subspecies. Within *A. m. mellifera,*we identified well-defined substructures according to the genetic origin of the colonies and a fair amount of admixed colonies, despite the current applied conservation efforts. In contrast, *A. m. carnica*colonies were more inbred and could not be differentiated according to the geographical origin. We identified 29 coding genes in overlapping ROH segments within the two subspecies. Genes embedded in*A. m. carnica* specific homozygosity islands suggested a strong selection for production and behavioural traits, whilst the identified cuticula protein-coding genes (*CPR3* and *CPR4)* were associated with their breed-specific stripe pattern. Local adaption of the two subspecies could be confirmed by the identification of two genes involved in the response to ultraviolet (UV) light. We demonstrated that colony genotypes derived from pooled honey bee workers are reliable to unravel the population dynamics in *A. mellifera* and provide fundamental information to conserve native honey bees.

Key words

Apis mellifera mellifera , *Apis mellifera carnica* , conservation, genetic diversity, selection signatures, CPR3, CPR4, stripe pattern

Introduction

The Western honey bee (*Apis mellifera* , hereafter honey bee*),* is a key pollinator of agricultural crops (Klein et al., 2007). To date, more than 27 subpopulations have been reported globally, which can be grouped into four distinct lineages, namely M (Western and Northern Europe), C (Eastern Europe), O (Near East and Central Asia) and A (Africa) (Cridland, Tsutsui, & Ramírez, 2017; Friedrich Ruttner, 1988). These lineages are characterized by differences in morphology, physiology and behaviour (Friedrich Ruttner, 1988). Within their native range, honey bees are commonly kept in hives for honey production and pollination purposes. In Europe, several selection programmes have been initiated to increase the productivity of honey bees (Adam, 1983; Büchler, Berg, & Le Conte, 2010; Chauzat et al., 2013; Guichard et al., 2020; H. Ruttner, 1972; Uzunov, Brascamp, & Büchler, 2017), while in Africa the majority of honey bees essentially evolved under natural selection (Dietemann, Pirk, & Crewe, 2009).

In some countries, honey bees are considered as a domesticated livestock species (Geldmann & González-Varo, 2018), due to the strong impact of human-mediated selection. In the beginning of the 19$^{\text{th}}$century, importation of foreign honey bees among European regions began to increase, which profoundly reshaped the genetic structure of this species (Parejo, Wragg, Henriques, Charrière, & Estonba, 2020). Historically, native honey bees of Europe mainly belong to M and C evolutionary lineages. They are locally adapted to different climatic and geographical regions, resulting in several subspecies (Momeni et al., 2021; Friedrich Ruttner, 1988). Nevertheless, beekeepers in Northern Europe continue to replace native honey bees (*A. m. mellifera*) with South-European honey bees (e.g. *A. m. carnica*and *A. m. ligustica*), as these subspecies are known to be more productive, gentle and calm (Bouga et al., 2011; Guichard et al., 2021). This practice has led to multiple admixture events between subspecies and the extinction of locally adapted honey bees (Bieńkowska, Splitt, Węgrzynowicz, & Maciorowski, 2021; F Ruttner, 1995). Furthermore, native honey bees

are threatened by the widespread use of stabilised hybrid strains such as Buckfast (Adam, 1983; Bieńkowska et al., 2021).

The relocation of subspecies accompanied by admixture is a major risk factor of losing local adaptation and genetic diversity of honey bees (De la Rúa, Jaffé, Dall'Olio, Muñoz, & Serrano, 2009). Therefore, several conservation programmes have been initiated to maintain the genetic diversity of native honey bees. In Switzerland, the first conservation area for *A. m. mellifera* was established in 1977 in canton Glarus (Soland-Reckeweg, Heckel, Neumann, Fluri, & Excoffier, 2009). Nowadays, an additional conservation area exists in canton Obwalden. The two conservatories encompass a total area of 830 km^2 and ˜1050 colonies (Parejo et al., 2016). To limit admixture events with other foreign subspecies (e.g.*A. m. carnica* and Buckfast) these areas are typically located in remote alpine valleys. Besides the maintenance of the conservation areas, the breeding association of *A. m. mellifera* (mellifera.ch) established a selection programme including several mating stations. These stations are also located at geographically isolated areas and consist of 10 up to 20 selected drone-producing colonies. Currently, an ancestry-informative (microsatellites or single nucleotide polymorphisms; SNPs) marker panel is applied to determine the admixture level of conserved and selected *A. m. mellifera* colonies, and highly admixed colonies (>10%) are replaced with purebred*A. m. mellifera* (Parejo, Henriques, Pinto, Soland-Reckeweg, & Neuditschko, 2018). However, the replacement of admixed queens is expected to lead to an increase in inbreeding that could be detrimental to the small conserved *A. m. mellifera* population. Given that the survival of honey bees is strongly dependent on the genetic diversity (Jones, Myerscough, Graham, & Oldroyd, 2004; Kryger, 1990; Mattila, Rios, Walker-Sperling, Roeselers, & Newton, 2012; Mattila & Seeley, 2014; Oldroyd, Rinderer, Harbo, & Buco, 1992), monitoring of inbreeding in small conserved populations, such as *A. m. mellifera* in Switzerland, is crucial.

Inbreeding level indicates the probability that an animal receives the same allele from both parents. Genetic marker information allows to determine that alleles are identical-by-descent (IBD), while pedigree-based estimations require prior knowledge of individual ancestry (Kardos, Luikart, & Allendorf, 2015), which in case of the honey bee is often not available. Runs of homozygosity (ROH), IBD transmitted haplotypes, are one of the tools to estimate inbreeding levels without ancestry information (McQuillan et al., 2008). In fact, the length of ROH segments can be used to ascertain historical changes in population size and structure including admixture (few and short ROH segments), current inbreeding (multiple and long ROH segments) and a recent bottleneck (multiple and short ROH segments); see Ceballos (2018) for a complete review. Furthermore, it is possible to derive the genomic inbreeding coefficient (F_{ROH}) for an animal by dividing the sum of all homozygous segments (S_{ROH}) by the length of the analysed genome (McQuillan et al., 2008). Numerous studies have demonstrated that overlapping ROH segments, so-called homozygosity islands, can be successfully used to identify selection signatures in cattle (Purfield, Berry, McParland, & Bradley, 2012; Zhang, Guldbrandtsen, Bosse, Lund, & Sahana, 2015), sheep (Mastrangelo et al., 2017; Purfield, McParland, Wall, & Berry, 2017; Signer-Hasler, Burren, Ammann, Drogemuller, & Flury, 2019), and horses (Druml et al., 2018; Grilz-Seger, Druml, et al., 2019; Grilz-Seger et al., 2018; Metzger et al., 2015), as well as in cultivated plants such as avocados (Rubinstein et al., 2019), almonds (Pavan et al., 2021) and pears (Kumar et al., 2020).

In this study we investigated the utility of colony genotypes derived from pooled workers to identify ROH segments in honey bees. Furthermore, we integrated individual admixture and F_{ROH} in a high-resolution population structure analysis to enhance the genetic monitoring of conserved *A. m. mellifera.* Finally, we screened the genome for homozygosity islands to detect selection signatures between*A. m. mellifera* and *A. m. carnica* honey bee colonies, related to geographic adaptations and human mediated selection.

Material and Methods

Sampled colonies

We sampled 265 Western honey bee colonies from two different subspecies, namely *A. m. mellifera* (MEL) and *A. m. carnica* (CAR) (Figure 1). Conserved MEL colonies were sampled in Switzerland (CS_CH) and France (CS_FR). The majority of the colonies belonged to MEL from the selection programme (SL_-CH) in Switzerland, which simultaneously represents five different patrilines (P1-P5). The sample size, geographic origin and location of the five different patrilines and conserved MEL colonies are summarised in Table 1 and Supplementary Figure 1. It should be noted that P1 is located in close proximity to the conservation area (CS_CH) and that P4 and P5 have a common maternal origin. The 49 sampled CAR colonies originated from Switzerland (22), Sweden (3), Norway (3) and the United States of America (21). For each colony, approximately 500 workers were collected with a standardized sampling method to include all existing patrilines among workers in the colony.

DNA extraction and pool sequencing

DNA extraction and pool sequencing of the sampled colonies are described in detail by Guichard et al. (2021). Briefly, the approximately 500 workers per colony were shredded in a DNA extraction solution. Pair-end sequencing was performed on an Illumina^{TM} HiSeq 3000 or a NovaSeq^{TM} 6000 platform. To significantly increase computing time, the pool sequence analysis was restricted to an informative marker panel including 7,023,977 genome-wide SNPs. Raw reads from pool sequencing of the 265 colonies were aligned to the honey bee reference genome Amel_HAV3.1, Genebank assembly accession GCA_003254395.2 (Wallberg et al., 2019). After the alignment, the resulting BAM files were converted into pileup files using the samtools mpileup utility (Li et al., 2009). Files produced by mpileup were interpreted by the PoPoolation2 utility mpileup2sync (Kofler, Pandey, & Schlotterer, 2011) for the Sanger Fastq format, with a minimum quality of 20. Finally, sync files were converted to a depth file containing a sequencing depth value for each SNP and count files summarising reference and alternative allele counts for each SNP.

Quality filtering and dosage data conversion

Based on the aforementioned count files, we removed 99,555 SNPs with multiple alternative alleles and 207,904 SNPs with an excessively high and low sequencing depth. After this quality control, we calculated the frequencies of the reference and the alternative alleles for 6,716,518 SNPs and additionally excluded 771,835 homozygous loci. The remaining 5,944,683 SNPs were summarised in PLINK dosage and map files. To detect ROH segments of the colonies, dosage data were converted to hard-called genotypes using the command – *import dosage* with a*hard-call threshold* of 0.4, as implemented in PLINK 2.0 (Chang et al., 2015). Genotypes set to missing during the conversion were imputed with the software BEAGLE 5.2 (Browning, Zhou, & Browning, 2018). For the population structure analyses, hard-called SNP genotypes were further pruned for minor allelic frequency above 5%, which resulted in 1,505,596 genome-wide SNPs.

High-resolution population network

To ascertain the high-resolution population structure of honey bees, we performed a population network visualization. The different components involved in the so-called NetView approach are described in detail by Neuditschko et al. (2012) and Steining et al. (2016). Briefly, we computed genetic distances by subtracting pairwise relationships identical-by-state (IBS) from 1 and applied the algorithm in its default setting (number of k nearest neighbours k -NN = 10). To illustrate the genetic relatedness between neighbouring honey bee colonies, we associated the thickness of edges (connecting lines) with the proportion of the genetic distance, with thicker edges corresponding to lower genetic distances. To identify highly inbred honey bee colonies, we scaled the node size of each colony based on the individual F_{ROH}. The node colour was associated according to the sampled subpopulations and the individual level of admixture at the optimal number of clusters.

Admixture

Colony admixture levels and genetic distances (F_{ST}) between the subspecies were determined using the program Admixture 1.23 (Alexander, Novembre, & Lange, 2009). We ran Admixture for 100 iterations increasing K from 2 to 10. Convergence between independent runs at the same K was monitored by comparing the resulting log-likelihood scores (LLs) following 100 iterations, and was inferred from stabilized LLs with less than 1 LL unit of variation between runs. Cross validation (CV) error estimation for each K was performed to determine the optimal number of clusters. Admixture results increasing K from 2 to 7 were visualized with the program Distruct 1.1 (Rosenberg, 2004) and integrated in the high-resolution population network, as described above.

Runs of homozygosity

Runs of homozygosity segments were determined with an overlapping window approach implemented in PLINK v.1.9 (Chang et al., 2015) including the aforementioned 5,944,683 genome-wide SNPs. The following settings were applied: a minimum SNP density of one SNP per 80 kb, a maximum gap length of 100 kb, and a minimum length of homozygous segment of 200 kb, while two heterozygotes were permitted in each segment. The total number of ROH (N_{ROH}), the total length of ROH segments (S_{ROH}) and the average length of ROH (L_{ROH}) were summarised for the two subspecies (CAR and MEL) and the respective subpopulations. The genomic-based inbreeding coefficients (F_{ROH}) were calculated by dividing S_{ROH} by the length of the autosomal genome (L_{AUTO}), which was set to 220.76 Mb (Wallberg et al., 2019). Furthermore, we compared F_{ROH} of 74 SL_CH colonies with pedigree-based inbreeding coefficients (F_{PED}) using a linear regression model as implemented in the statistical computing software R (R Core Team, 2013), whereas F_{PED} were calculated following the method described by Brascamp et al. (2014).

Homozygosity islands and gene functions

Homozygosity islands of the two subspecies were determined based on overlapping homozygous regions present in more than 50% of the colonies with the R package detectRUNS (Biscarini, Cozzi, Gaspa, & Marras, 2019). The length and distribution of the homozygosity islands on the chromosomes were visualised using the R package RIdeogram (Hao et al., 2020). We used the NCBI genome data viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/), and the reference genome assembly Amel_HAv3.1 (Wallberg et al., 2019) to identify genes located in homozygosity islands. For each subspecies, we regrouped the characterised genes by the function term based on the annotation chart from the open source Database for Annotation, Visualization, and Integrated Discovery v.6.8 package (https://david.ncifcrf.gov), using the *Apis mellifera* annotation file, with a minimum of two genes per functional group and a Bonferroni-corrected significance threshold of $p<0.05$. Furthermore, we specified the known functions of the identified genes by conducting a literature review.

Results

High-resolution population network

The high-resolution population network clearly divided the honey bees into two distinct population clusters (CAR and MEL), while five MEL colonies (one SL_CH, and four CS_FR) were allocated to CAR (Figure 2A, dashed circle). The hub between CAR and MEL particularly included CS_FR and SL_CH colonies that were not clustering with their respective patrilines. The topology of the network additionally revealed that further substructures exist within MEL, while CAR colonies built a tight population cluster, despite the different geographical sample origin. The most evident substructures within MEL corresponded to CS_CH colonies and two patrilines of the selection programme (P1 and P2). It was interesting to see that four CS_FR colonies were directly connected with two CS_CH colonies, while the remaining CS_FR colonies were

distributed over the network. Furthermore, CS_CH colonies were the nearest neighbours of four SL_CH colonies originating from three different patrilines (P1, P3 and P5), while P1 showed the strongest genetic relationship. Compared to P1 and P2, the remaining three patrilines (P3-P5) did not build a distinct population cluster, P3 colonies were distributed over the network without a discernible pattern, while the majority of P4 and P5 colonies built a common cluster. The association of the node size with F_{ROH}illustrates that with the exception of four colonies, all CAR showed higher F_{ROH} than MEL. Furthermore, it can be noted that CAR located in the neighbourhood of MEL (and vice versa), as well as colonies not clustering with the respective patrilines show lower F_{ROH} in general (Figure 2A).

Admixture

Based on the cross-validation error estimation, an optimal cluster solution at K = 5 was determined increasing K from 2 to 10 (Supplementary Figure 2A). The first level (K = 2) of model-based clustering clearly differentiated CAR from MEL with a F_{ST} of 0.26 (Supplementary Figure 2B). The cluster solution simultaneously highlighted that except for CS_CH and P1, all MEL subpopulations contained admixed colonies, whereas CS_FR showed the highest percentage of highly admixed colonies. At the second (K = 3) and third level (K = 4) of clustering, P2 and CS_FR colonies built a distinct cluster, respectively. At the optimal cluster solution (K = 5), P4 and P5 colonies were allocated in a common population cluster. At the additional two levels of clustering (K = 7 and K = 8), some CAR colonies from the United States of America were allocated in a distinct cluster and the aforementioned common population cluster was further sub-structured, without separating P4 from P5 colonies. Therefore, the hierarchical population clustering (increasing K from 2 to 5) confirmed the findings of the network-based population structure, with the only exception to clearly differentiate P1 colonies from CS_CH. This high agreement between the two applied population structure methods also became visible by integrating the admixture levels at K = 5 in the high-resolution population network (Figure 2B), which simultaneously revealed that colonies not clustering with their respective patrilines and having low F_{ROH} were highly admixed. Based on this observation, we removed five MEL and two CAR outliers from downstream ROH analyses, while 38 MEL colonies with an admixture level greater than 5% (K = 2) were summarized in a distinct population cluster (ADMEL) and excluded from the identification of homozygosity islands.

Runs of homozygosity

The ROH analysis recapitulates the results of the population structure analyses (Table 2). The CAR sample had nearly twice as many ROH segments (N_{ROH} = 23.28 +-4.95) than MEL (N_{ROH} = 13.60 +- 1.66) and ADMEL (N_{ROH} = 12.42 +- 2.62). Concurrently, the total length of ROH was also twice as high in CAR (S_{ROH} = 11.65 Mb +- 0.28) compared to the other two population cluster (4.96 Mb +- 0.91 to 5.50 Mb +- 0.50), while the mean segment length (L_{ROH}) remained similar between the three population clusters (between 0.40 Mb and 0.50 Mb). As expected, the lowest ROH values (N_{ROH} of 12.42 +- 2.62, S_{ROH} of 4.96 Mb +- 0.91, and L_{ROH} of 0.40 Mb +- 0.05) were observed for ADMEL. Similarly, the ADMEL group had the lowest mean F_{ROH} (2.20%), followed by MEL (2.50%) and CAR (5.30%).

Summarizing the ROH results within MEL according to the *a priori*defined subpopulations (patrilines and conservation areas) revealed that CS_FR, including the highest proportion of admixed colonies (62%), was the subpopulation with the lowest ROH values (N_{ROH} = 12.62 +- 2.50, S_{ROH} = 5.00 Mb +- 0.86, L_{ROH} = 0.40 Mb +- 0.04, F_{ROH} = 2.30% +- 0.40) (Supplementary Table 1). The two subpopulations without any admixed colonies, P1 and CS_CH, showed the highest minimal F_{ROH} (2.30% and 2.10%, respectively), while the highest maximal F_{ROH} (3.20%) was identified within P4. Furthermore, P4 was associated with the highest maximal N_{ROH} (18.00) and mean S_{ROH} (5.62 Mb +- 0.59), while P3 had the highest mean N_{ROH} (14.00 +- 1.73) and shortest mean L_{ROH} (0.38 Mb +- 0.03).

The F_{PED} of the 74 SL_CH colonies including all patrilines except P4 ranged from 0.00 to 11.26%, with a mean F_{PED} of 3.13% +- 0.26, whereas F_{ROH}ranged from 1.40 to 3.10%, with a mean F_{ROH} of 2.40% +-

0.28. The association between the two inbreeding coefficient measures was low ($R^2 = 0.02$). Including the individual admixture levels, as a covariate, in the linear regression model improved the concordance between F_{PED} and F_{ROH} ($R^2 = 0.10$). Absolute differences between F_{PED} and F_{ROH} were higher in the ADMEL (15 colonies > 5% CAR admixture, mean absolute difference = 2.84%) compared to the remaining colonies (59 colonies < 5% CAR admixture, mean absolute difference = 1.82%).

Homozygosity islands

We identified 24 CAR-specific (private) homozygosity islands (mean length= 0.53 +- 0.69 Mb) distributed over nine chromosomes, while MEL had 17 homozygosity islands (mean length= 0.35 +- 0.19 Mb) over seven chromosomes (Figure 3, Supplementary Table 2). Chromosomes 14 and 16 did not bear any homozygosity islands. The largest homozygosity island was identified for CAR on chromosome 11 at 3,737,355 bp to 7,286,231 bp, which covers 3.55 Mb, or approximately 20% of the entire chromosome. The shortest homozygosity island was private for MEL, spanning just 465 bp on chromosome 7. The majority of homozygosity islands were located at the starting end of the chromosomes. Four chromosomes (6, 9, 12 and 15) comprised homozygosity islands for both subspecies, but only one segment on chromosome 12 overlapped between CAR and MEL, spanning from 3,124,678 bp to 3,335,849 bp, which did not contain any annotated genes.

Only 11 out of 24 homozygosity islands for CAR and 5 out of 17 for MEL contained annotated genes (Table 3), and one CAR homozygosity island did not even contain any uncharacterised loci (Supplementary Table 2). There were substantially more uncharacterised loci than annotated genes within the homozygosity islands (i.e. 788 uncharacterised loci and 24 characterised genes in CAR).

The 24 annotated genes embedded within CAR-specific homozygosity islands clustered into six functional groups with high significance levels (Bonferroni-adjusted p-value < 0.05, Supplementary Table 3):*Neuroactive ligand-receptor interaction* , *Insect cuticle protein* , *Structural constituent of cuticle* , *Ion transport, cell junction* and *Synapse* . Half of the characterised genes (12 out of 24) could not be regrouped by function, namely *Ndufs1, PHRF1, Rep, Snf, Chmp1, Crh-BP, Gpdh, Grp, Rga, RpL35, Tpx-4 and Uqcr11* . The five genes embedded in MEL-specific homozygosity islands (*GstS1, Pban, WRNexo, Uvop* and *Mad*) did not share functional terms.

Discussion

We demonstrated that colony genotypes derived from pooled honey bee workers can successfully be applied to ascertain high-resolution population structures, including the computation of $F_{ROH}$and the detection of breed-specific selection signatures. To date, mostly drone genomes were used to assess the genetic diversity of honey bees, as their haploid nature facilitates cost-efficient whole-genome sequencing (Parejo et al., 2016). However, based on the haploid data structure, which exhibits systematic homozygosity, it is likely to overestimate genetic relationships and subsequently inbreeding, compared to other livestock species (Wragg et al., 2016). Another disadvantage of honey bee drones is that they only explain part of the genetic diversity, as multiple patrilines are involved in the formation of honey bee colonies (Estoup, Solignac, & Cornuet, 1994; Neumann, Moritz, & van Praagh, 1999; Tarpy, Nielsen, & Nielsen, 2004).

The applied population structure analyses clearly differentiated MEL from CAR, whilst the $F_{ST} = 0.26$ between the two subspecies was lower compared to our previous findings using drone genomes ($F_{ST} = 0.36$) (Parejo et al., 2016). This reduced genetic difference might be the result of highly admixed MEL colonies, which simultaneously highlights the challenges to conserve native honey bees due to random mating. Therefore, the high-resolution population network illustrates that a successful honey bee conservation programme requires an appropriate management tool including a legal framework, a suitable geographical isolated location, and ancestry informative marker testing, like the conservation strategy of MEL in the Canton Glarus, the only area without admixed colonies (CS_CH and P1). However, in our view the strong gene flow between

the two subpopulations can have a negative impact on the in-situ conservation as foreign genetic variants might be introduced to the CS_CH gene pool by the mating station.

Compared to CS_CH, the origin of CS_FR colonies has only been sporadically assessed in the past based on wing vein measurements, which simultaneously explains the highly observed diversity of the colonies, whereas four colonies showed a high genetic relatedness with CS_CH. The population structure of CS_FR and the random mating events observed in some SL_CH patrilines indicate that current applied conservation strategies including the geographical locations are not suitable for in-situ conservation. Ex-situ conservation by means of artificial insemination (Cobey, Tarpy, & Woyke, 2013), could be a more efficient alternative to maintain the gene pool of native honey bees.

In spite of the small population size, MEL showed significantly lower F_{ROH} than CAR. This result corresponds to the current applied selection strategies of both subspecies. Within MEL, SL_CH are carefully selected to contribute to the local genetic diversity, whereas CAR can be considered as a highly specialized transboundary honey bee breed. This breed characteristic of CAR was also evident in the high-resolution population network, which successfully ascertained the numerous substructures within MEL, but failed to clearly separate CAR colonies according to their geographical origin. The ROH results, according to the observed population structure (Table 2), revealed that the population admixture of MEL was another reason for the low F_{ROH}. It also simultaneously indicated a direct relationship between admixture and ROH length in honey bees. The effect of population admixture on homozygosity patterns has already been documented in other populations, such as cattle (Purfield et al., 2012) and goats (Bertolini et al., 2018). The population admixture also had an effect on the relationship between F_{ROH} and F_{PED}, whereas the improved value ($R^2 = 0.10$) was significantly lower than commonly observed values in livestock, such as goats ($R^2 = 0.27 – 0.65$) (Burren et al., 2016). Compared to other livestock populations, the paternal origin of honey bees is not known, as honey bee queens naturally mate in flight with 10 – 20 drones (polyandrous mating system) (Estoup et al., 1994; Neumann et al., 1999; Tarpy et al., 2004). Hence, the paternal origin must be estimated by restricting paternal origins to the drone-producing colonies located at the mating station. However, the proportion of foreign drones contributing to the mating remains unknown. Therefore, F_{PED} of colonies from insufficiently isolated mating stations (with higher admixture proportions) are overestimated, while a low pedigree completeness results in lower F_{PED}compared to F_{ROH}. To improve the pedigree quality of honeybees, we suggest confirming the parental origin with a marker-based parentage analysis.

Within CAR-specific homozygosity islands, we identified several genes that are directly associated with the current applied selection traits, including increased productivity, as well as reduced defensive behaviour and swarming drive (Bouga et al., 2011; Guichard et al., 2021). Based on highly selected *A. m. ligustica* lineages, it has already been demonstrated that *RpL35* controls royal jelly production and larval growth (Ararso et al., 2018). Furthermore, the differential expression of *Ndufs1* (Guo et al., 2019) and *Gpdh* (Seehuus, Taylor, Petersen, & Aamodt, 2013) may also increase foraging behaviour, and consequently, productivity. The reduced aggressiveness in CAR (Guichard, Neuditschko, Fried, Soland, & Dainat, 2019; Friedrich Ruttner, 1988) may be associated to the private homozygosity island containing the gene 5-ht7, coding for serotonin receptor 7, as higher serotonin levels increase the likelihood of bee stings, the ultimate defence mechanism of the colony (Nouvian et al., 2018).

The longest private homozygosity island of CAR included several cuticular protein-coding genes (*CPR3* and *CPR4* in particular). These proteins may be involved in the CAR specific morphotype of broader hairy stripes (Figure 1), as they effect the thickness and colouring of the exoskeleton (Costa et al., 2016; Soares et al., 2013). Another gene, *Chmp1,* is known to influence the veining pattern in Drosophila (Valentine, Hogan, & Collier, 2014), which might explain the morphological differences in vein patterns between the two subspecies (Bouga et al., 2011; Friedrich Ruttner, 1988).

The genes *chr-BP* and *Uvop* , embedded in private islands for CAR and MEL respectively, are both involved in the resistance to ultraviolet (UV) exposure, but reveal different adaptive mechanisms due to the ancestral geographical origin of the two subspecies. The gene crh-BP was shown to be upregulated in honey bees in response to UV exposure and heat stress (Even, Devaud, & Barron, 2012). CAR-specific homozygosity in

that gene could therefore indicate local adaptation to more constant sun exposure and higher temperatures of Southern Europe. The homozygous state of the *Uvop* gene in MEL is associated with retinal development and the circadian rhythm (Lichtenstein, Grubel, & Spaethe, 2018), which may enable MEL to deal with seasonally more variable sun exposure. For example, within fish certain Malawi Cichlids also show differential expression of opsin genes depending on their photic environment (Parry et al., 2005). Diurnal mammal species also produce different quantities of ultraviolet-sensitive pigments depending on their ecological niche (Emerling, Huynh, Nguyen, Meredith, & Springer, 2015). Furthermore, it has recently been demonstrated that genes involved in the response to UV exposure are associated with the local adaption of horse breeds (Grilz-Seger, Neuditschko, et al., 2019).

Several characterised genes in a homozygous state for CAR and MEL shared functions associated with stress response (*ATP5G2* , *Nmdar1, Tpx-4, GstS1* (Alburaki, Karim, Lamour, Adamczyk, & Stewart, 2019; Watts, Williams, Nithianantharajah, & Claudianos, 2018; Yan, Jia, Gao, Guo, & Xu, 2013)), DNA integrity (*Uzip* , *WRNexo* (Ding et al., 2011; Rossi, Ghosh, & Bohr, 2010)) and immunity (*CPR2, SP34, Pban, CTL5* (Badaoui et al., 2017; Lin et al., 2020; McDonnell et al., 2013; Zou, Lopez, Kanost, Evans, & Jiang, 2006)). However, at the current stage of research, it is not clear whether the homozygosity state of these genes has a positive or negative effect on the aforementioned functions. Therefore, fine-tuned gene expression studies are required to assess the selection direction within the two subspecies.

In summary, we have described a number of novel aspects to investigate the genetic diversity of honey bees that are of potential interest. Firstly, the application of colony genotypes derived from pooled honey bee workers to ascertain fine-scale population structures. Secondly, the identification of ROH segments from pool-seq data to compute genomic inbreeding of honey bee colonies. Finally, the assessment of breed-specific selection signatures by means of ROH islands. Therefore, we believe that ROH derived from pool-seq data will be of invaluable benefit to investigate complex population structures in honey bees and other insects, whereas the hard-call genotype threshold and ROH setting should be further investigated.

Acknowledgements

Financial support for this study was provided by Bundesamt fur Landwirtschaft BLW (Swiss Federal Office for Agriculture FOAG) Grant No. 627000708, by Labogena and FranceAgriMer (Programme d'Investissements d'Avenir), the Alabama Agricultural Experiment Station, and the USDA NIFA Multi-state Hatch Project NC1173. We thank Lucie Genestout (Labogena) for coordinating the Beestrong project, Christina Sann and Yannick Poquet (Labogena) for organising the sampling in Savoie, and Alice Revel (Agroscope) for her help in organising the sampling in Switzerland. We thank Rachid Mahla (Labogena) for DNA extraction and Olivier Bouchez (GeT-PlaGe) for sequencing. Sequencing was performed in collaboration with the GeT platform, Toulouse (France), a partner of the National Infrastructure France Genomique, thanks to support by the Commissariat aux Grands Investissements (ANR-10-INBS-0009). Sequence alignments and allele frequencies estimations were performed on the computer cluster of the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (www.bioinfo.genotoul.fr). We are grateful to all of the experts for colony evaluation and to all of the beekeepers for supporting the sample collection. This study was performed in collaboration with the Beestrong Consortium. Beestrong Consortium: Sonia Eynard, Alain Vignal, Bertrand Servin (INRAE, GenPhyse, 31326 Castanet-Tolosan, France), Lucie Genestout, Rachid Mahla, Christina Sann, Yannick Poquet, Yves Agez (LABOGENA DNA, Laboratoire d'Analyses Genetiques pour les Especes Animales, Domaine de Vilvert, 78353 Jouy-en-Josas, France), Francois Guillaume (Evolution XY, rue Eric Tabarly, 35538 Noyal-Sur-Vilaine, France), Benjamin Basso, Yves Le Conte, Fanny Mondet (INRAE, Abeilles et Environnement, 84914 Avignon, France & UMT PrADE, 84914 Avignon, France), Benjamin Basso, Axel Decourtye (ITSAP, 84914 Avignon, France), Florence Phocas, Jean-Pierre Bidanel (Universite Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France), Olivier Bouchez (GeT-PlaGe, Genotoul, INRAE Auzeville, 31320, Castanet-Tolosan, France).

References

Adam, B. (1983). *In search of the best strains of bees* : Northern Bee Books.

Alburaki, M., Karim, S., Lamour, K., Adamczyk, J., & Stewart, S. D. (2019). RNA-seq reveals disruption of gene regulation when honey bees are caged and deprived of hive conditions. *Journal of Experimental Biology, 222* (18), jeb207761.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research, 19* (9), 1655-1664.

Ararso, Z., Ma, C., Qi, Y., Feng, M., Han, B., Hu, H., . . . Li, J. (2018). Proteome comparisons between hemolymph of two honeybee strains (Apis mellifera ligustica) reveal divergent molecular basis in driving hemolymph function and high royal jelly secretion. *Journal of proteome research, 17* (1), 402-419.

Badaoui, B., Fougeroux, A., Petit, F., Anselmo, A., Gorni, C., Cucurachi, M., . . . Formato, G. (2017). RNA-sequence analysis of gene expression from honeybees (Apis mellifera) infected with Nosema ceranae.*PloS one, 12* (3), e0173438.

Bertolini, F., Cardoso, T. F., Marras, G., Nicolazzi, E. L., Rothschild, M. F., & Amills, M. (2018). Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution, 50* (1), 1-12.

Bieńkowska, M., Splitt, A., Węgrzynowicz, P., & Maciorowski, R. (2021). The Buzz Changes within Time: Native Apis mellifera mellifera Honeybee Subspecies Less and Less Popular among Polish Beekeepers Since 1980.*Agriculture, 11* (7), 652.

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2019). detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes.

Bouga, M., Alaux, C., Bienkowska, M., Büchler, R., Carreck, N. L., Cauia, E., . . . De la Rúa, P. (2011). A review of methods for discrimination of honey bee populations as applied to European beekeeping. *Journal of Apicultural Research, 50* (1), 51-84.

Brascamp, E. W., & Bijma, P. (2014). Methods to estimate breeding values in honey bees. *Genetics Selection Evolution, 46* (1), 1-15.

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics, 103* (3), 338-348.

Büchler, R., Berg, S., & Le Conte, Y. (2010). Breeding for resistance to Varroa destructor in Europe. *Apidologie, 41* (3), 393-408.

Burren, A., Neuditschko, M., Signer-Hasler, H., Frischknecht, M., Reber, I., Menzi, F., . . . Flury, C. (2016). Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Animal genetics, 47* (6), 727-739.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics, 19* (4), 220.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4* (1), s13742-13015-10047-13748.

Chauzat, M.-P., Cauquil, L., Roy, L., Franco, S., Hendrikx, P., & Ribiere-Chabert, M. (2013). Demographics of the European apicultural industry. *PloS one, 8* (11), e79018.

Cobey, S. W., Tarpy, D. R., & Woyke, J. (2013). Standard methods for instrumental insemination of Apis mellifera queens. *Journal of Apicultural Research, 52* (4), 1-18.

Costa, C. P., Elias-Neto, M., Falcon, T., Dallacqua, R. P., Martins, J. R., & Bitondi, M. M. G. (2016). RNAi-mediated functional analysis of bursicon genes related to adult cuticle formation and tanning in the honeybee, Apis mellifera. *PloS one, 11* (12), e0167421.

Cridland, J. M., Tsutsui, N. D., & Ramirez, S. R. (2017). The complex demographic history and evolutionary origin of the western honey bee, Apis mellifera. *Genome biology and evolution, 9* (2), 457-472.

De la Rua, P., Jaffe, R., Dall'Olio, R., Munoz, I., & Serrano, J. (2009). Biodiversity, conservation and current threats to European honeybees. *Apidologie, 40* (3), 263-284.

Dietemann, V., Pirk, C. W. W., & Crewe, R. (2009). Is there a need for conservation of honeybees in Africa? *Apidologie, 40* (3), 285-295.

Ding, Z. Y., Wang, Y. H., Luo, Z. K., Lee, H. F., Hwang, J., Chien, C. T., & Huang, M. L. (2011). Glial cell adhesive molecule unzipped mediates axon guidance in Drosophila. *Developmental Dynamics, 240* (1), 122-134.

Druml, T., Neuditschko, M., Grilz-Seger, G., Horna, M., Ricard, A., Mesarič, M., . . . Brem, G. (2018). Population networks associated with runs of homozygosity reveal new insights into the breeding history of the Haflinger horse. *Journal of Heredity, 109* (4), 384-392.

Emerling, C. A., Huynh, H. T., Nguyen, M. A., Meredith, R. W., & Springer, M. S. (2015). Spectral shifts of mammalian ultraviolet-sensitive pigments (short wavelength-sensitive opsin 1) are associated with eye length and photic niche evolution. *Proceedings of the royal society B: biological sciences, 282* (1819), 20151817.

Estoup, A., Solignac, M., & Cornuet, J.-M. (1994). Precise assessment of the number of patrilines and of genetic relatedness in honeybee colonies. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 258* (1351), 1-7.

Even, N., Devaud, J.-M., & Barron, A. B. (2012). General stress responses in the honey bee. *Insects, 3* (4), 1271-1298.

Geldmann, J., & González-Varo, J. P. (2018). Conserving honey bees does not help wildlife. *Science, 359* (6374), 392-393.

Grilz-Seger, G., Druml, T., Neuditschko, M., Dobretsberger, M., Horna, M., & Brem, G. (2019). High-resolution population structure and runs of homozygosity reveal the genetic architecture of complex traits in the Lipizzan horse. *BMC genomics, 20* (1), 174.

Grilz-Seger, G., Mesarič, M., Cotman, M., Neuditschko, M., Druml, T., & Brem, G. (2018). Runs of homozygosity and population history of three horse breeds with small population size. *Journal of Equine Veterinary Science, 71* , 27-34.

Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesarič, M., . . . Brem, G. (2019). Genome-wide homozygosity patterns and evidence for selection in a set of European and near eastern horse breeds. *Genes, 10* (7), 491.

Guichard, M., Dainat, B., Eynard, S., Vignal, A., Servin, B., Consortium, B., & Neuditschko, M. (2021). Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences. *Animal genetics* .

Guichard, M., Neuditschko, M., Fried, P., Soland, G., & Dainat, B. (2019). A future resistance breeding strategy against Varroa destructor in a small population of the dark honey bee. *Journal of Apicultural Research, 58* (5), 814-823.

Guichard, M., Neuditschko, M., Soland, G., Fried, P., Grandjean, M., Gerster, S., . . . Brascamp, E. W. (2020). Estimates of genetic parameters for production, behaviour, and health traits in two Swiss honey bee populations. *Apidologie* .

Guo, Y., Fu, B., Qin, G., Song, H., Wu, W., Shao, Y., . . . Yu, L. (2019). Proteome analysis reveals a strong correlation between olfaction and pollen foraging preference in honeybees. *International journal of biological macromolecules, 121* , 1264-1275.

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., & Chen, J. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science, 6* , e251.

Jones, J. C., Myerscough, M. R., Graham, S., & Oldroyd, B. P. (2004). Honey bee nest thermoregulation: diversity promotes stability.*Science, 305* (5682), 402-404.

Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity, 115* (1), 63-72.

Klein, A.-M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the royal society B: biological sciences, 274* (1608), 303-313.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics, 27* (24), 3435-3436.

Kryger, P. (1990). 4. Die Bedeutung der genotypischen Varianz für das hygienische Verhalten der Honigbiene. *Apidologie, 21* (4), 332-333.

Kumar, S., Deng, C. H., Hunt, M., Kirk, C., Wiedow, C., Rowan, D., . . . Brewer, L. (2020). Homozygosity mapping reveals population history and trait architecture in self-incompatible pear (Pyrus spp.).*Frontiers in Plant Science, 11* .

Li, H., Bob, H., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map (SAM) Format and.

Lichtenstein, L., Grübel, K., & Spaethe, J. (2018). Opsin expression patterns coincide with photoreceptor development during pupal development in the honey bee, Apis mellifera. *BMC developmental biology, 18* (1), 1-11.

Lin, J., Yu, X.-Q., Wang, Q., Tao, X., Li, J., Zhang, S., . . . You, M. (2020). Immune responses to Bacillus thuringiensis in the midgut of the diamondback moth, Plutella xylostella. *Developmental & Comparative Immunology, 107* , 103661.

Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., & Portolano, B. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution, 49* (1), 1-10.

Mattila, H. R., Rios, D., Walker-Sperling, V. E., Roeselers, G., & Newton, I. L. (2012). Characterization of the active microbiotas associated with honey bees reveals healthier and broader communities when colonies are genetically diverse. *PloS one, 7* (3), e32962.

Mattila, H. R., & Seeley, T. D. (2014). Extreme polyandry improves a honey bee colony's ability to track dynamic foraging opportunities via greater activity of inspecting bees. *Apidologie, 45* (3), 347-363.

McDonnell, C. M., Alaux, C., Parrinello, H., Desvignes, J.-P., Crauser, D., Durbesson, E., . . . Le Conte, Y. (2013). Ecto-and endoparasite induce similar chemical and brain neurogenomic responses in the honey bee (Apis mellifera). *BMC ecology, 13* (1), 25.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., . . . Tenesa, A. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics, 83* (3), 359-372.

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., . . . Distl, O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses.*BMC*

genomics, 16 (1), 764.

Momeni, J., Parejo, M., Nielsen, R. O., Langa, J., Montes, I., Papoutsis, L., . . . Charrière, J.-D. (2021). Authoritative subspecies diagnosis tool for European honey bees based on ancestry informative SNPs. *BMC genomics, 22* (1), 1-12.

Neuditschko, M., Khatkar, M. S., & Raadsma, H. W. (2012). NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PloS one, 7* (10), e48375.

Neumann, P., Moritz, R. F., & van Praagh, J. (1999). Queen mating frequency in different types of honey bee mating apiaries. *Journal of Apicultural Research, 38* (1-2), 11-18.

Nouvian, M., Mandal, S., Jamme, C., Claudianos, C., d'Ettorre, P., Reinhard, J., . . . Giurfa, M. (2018). Cooperative defence operates by social modulation of biogenic amine levels in the honey bee brain. *Proceedings of the royal society B: biological sciences, 285* (1871), 20172653.

Oldroyd, B. P., Rinderer, T. E., Harbo, J. R., & Buco, S. M. (1992). Effects of intracolonial genetic diversity on honey bee (Hymenoptera: Apidae) colony performance. *Annals of the Entomological Society of America, 85* (3), 335-343.

Parejo, M., Henriques, D., Pinto, M. A., Soland-Reckeweg, G., & Neuditschko, M. (2018). Empirical comparison of microsatellite and SNP markers to estimate introgression in Apis mellifera mellifera. *Journal of Apicultural Research, 57* (4), 504-506.

Parejo, M., Wragg, D., Gauthier, L., Vignal, A., Neumann, P., & Neuditschko, M. (2016). Using whole-genome sequence information to foster conservation efforts for the European Dark Honey Bee, Apis mellifera mellifera. *Frontiers in Ecology and Evolution, 4* , 140.

Parejo, M., Wragg, D., Henriques, D., Charrière, J.-D., & Estonba, A. (2020). Digging into the genomic past of Swiss honey bees by whole-genome sequencing museum specimens. *Genome biology and evolution, 12* (12), 2535-2551.

Parry, J. W., Carleton, K. L., Spady, T., Carboo, A., Hunt, D. M., & Bowmaker, J. K. (2005). Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Current Biology, 15* (19), 1734-1739.

Pavan, S., Delvento, C., Mazzeo, R., Ricciardi, F., Losciale, P., Gaeta, L., . . . Ricciardi, L. (2021). Almond diversity and homozygosity define structure, kinship, inbreeding, and linkage disequilibrium in cultivated germplasm, and reveal genomic associations with nut and seed weight. *Horticulture research, 8* (1), 1-12.

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC genetics, 13* (1), 70.

Purfield, D. C., McParland, S., Wall, E., & Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PloS one, 12* (5), e0176780.

R Core Team. (2013). R: A language and environment for statistical computing (Version 3.6.1).

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular ecology notes, 4* (1), 137-138.

Rossi, M. L., Ghosh, A. K., & Bohr, V. A. (2010). Roles of Werner syndrome protein in protection of genome integrity. *DNA repair, 9* (3), 331-344.

Rubinstein, M., Eshed, R., Rozen, A., Zviran, T., Kuhn, D. N., Irihimovitch, V., . . . Ophir, R. (2019). Genetic diversity of avocado (Persea americana Mill.) germplasm using pooled sequencing. *BMC genomics, 20* (1), 1-10.

Ruttner, F. (1988). *Biogeography and Taxonomy of Honeybees* : Springer.

Ruttner, F. (1995). Von der "dunklen" zur "grauen" deutschen Landbiene.*Imkerfreund, 95* (8), 16-18.

Ruttner, H. (1972). *Technische Empfehlungen zur Methodik der Leistungsprüfung von Bienenvölkern.* Paper presented at the Proceedings of Paarungskontrolle und Selektion bei der Honigbiene. Internationales Symposium. Lunz am See, Austria.

Seehuus, S.-C., Taylor, S., Petersen, K., & Aamodt, R. M. (2013). Somatic maintenance resources in the honeybee worker fat body are distributed to withstand the most life-threatening challenges at each life stage. *PloS one, 8* (8), e69870.

Signer-Hasler, H., Burren, A., Ammann, P., Drogemuller, C., & Flury, C. (2019). Runs of homozygosity and signatures of selection: A comparison among eight local Swiss sheep breeds. *Animal genetics, 50* (5), 512-525.

Soares, M. P. M., Barchuk, A. R., Simoes, A. C. Q., dos Santos Cristino, A., de Paula Freitas, F. C., Canhos, L. L., & Bitondi, M. M. G. (2013). Genes involved in thoracic exoskeleton formation during the pupal-to-adult molt in a social insect model, Apis mellifera. *BMC genomics, 14* (1), 1-17.

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 honeybee, Apis mellifera. *Journal of Insect Conservation, 13* (3), 317-328.

Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W., & Zenger, K. R. (2016). netview p: a network visualization tool to unravel complex population structure using genome-wide SNP s. *Molecular Ecology Resources, 16* (1), 216-227.

Tarpy, D., Nielsen, R., & Nielsen, D. (2004). A scientific note on the revised estimates of effective paternity frequency in Apis.*Insectes Sociaux, 51* (2), 203-204.

Uzunov, A., Brascamp, E., & Buchler, R. (2017). The basic concept of honey bee breeding programs. *Bee World, 94* (3), 84-87.

Valentine, M., Hogan, J., & Collier, S. (2014). The Drosophila Chmp1 protein determines wing cell fate through regulation of epidermal growth factor receptor signaling. *Developmental Dynamics, 243* (8), 977-987.

Wallberg, A., Bunikis, I., Pettersson, O. V., Mosbech, M.-B., Childers, A. K., Evans, J. D., . . . Webster, M. T. (2019). A hybrid de novo genome assembly of the honeybee, Apis mellifera, with chromosome-length scaffolds. *BMC genomics, 20* (1), 275.

Watts, M. E., Williams, S. M., Nithianantharajah, J., & Claudianos, C. (2018). Hypoxia-induced MicroRNA-210 targets neurodegenerative pathways.*Non-coding RNA, 4* (2), 10.

Wragg, D., Marti-Marimon, M., Basso, B., Bidanel, J.-P., Labarthe, E., Bouchez, O., . . . Vignal, A. (2016). Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly. *Scientific reports, 6* (1), 1-13.

Yan, H., Jia, H., Gao, H., Guo, X., & Xu, B. (2013). Identification, genomic organization, and oxidative stress response of a sigma class glutathione S-transferase gene (AccGSTS1) in the honey bee, Apis cerana cerana. *Cell Stress and Chaperones, 18* (4), 415-426.

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S., & Sahana, G. (2015). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC genomics, 16* (1), 542.

Zou, Z., Lopez, D. L., Kanost, M. R., Evans, J. D., & Jiang, H. (2006). Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect molecular biology, 15* (5), 603-614.

Data accessibility

The data that support the findings of this study remain the property of Agroscope (Swiss samples) and the Beestrong Consortium (French samples). However, data are available from the authors upon reasonable request.

Author contributions

M.N., M.G. and B.D. conceived and planned the study. A.I.G., M.G. and M.N. performed the analyses. M.G., B.D., G.W., S.E., A.V. and B.S. contributed to sample preparation and pooled sequence analysis. A.I.G., M.G., B.D. and M.N. contributed to the interpretation of the results. A.I.G. and M.N. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Tables and Figures

Table 1: Number of sampled colonies, geographic origin, legal framework and protection radius of patrilines and conserved Apis mellifera mellifera

Patrilines and conserved MEL	Number of sampled colonies	Geographic location	Canton (Ct.) and description of surro
P1	17	Krauchtal	Ct. Glarus, semi-isolated valley
P2	47	Gental	Ct. Bern, isolated valley
P3	9	Säntis	Ct. Appenzell Ausserrhoden, open va
P4	33	Schilstal	Ct. St-Gallen, semi-isolated valley
P5	39	Rothbach	Ct. Luzern, isolated valley
CS_CH	45	Glarus	Ct. Glarus, colonies widespread acros
CS_FR	25	Savoie (FR)	Colonies widespread across the valley

Table 2: Mean values, standard deviation (SD), and minimum and maximum values for total number of ROH (N_{ROH}), the total length of ROH segments (S_{ROH}), the average length of ROH (L_{ROH}), and genomic inbreeding coefficients (F_{ROH}) for three population clusters: Apis mellifera carnica (CAR), Apis mellifera mellifera (MEL) and MEL colonies with a CAR admixture proportion >5% (ADMEL)

Sample	**n**	**Mean**	**SD**	**Min**	**Max**
CAR	47				
N_{ROH}		23.28	4.95	9.00	35.00
S_{ROH} (Mb)		11.65	2.78	4.88	17.66
L_{ROH} (Mb)		0.50	0.07	0.37	0.67
F_{ROH} (%)		5.30	1.30	2.20	8.00
MEL	173				
N_{ROH}		13.60	1.66	8.00	18.00
S_{ROH} (Mb)		5.50	0.50	3.73	7.06
L_{ROH} (Mb)		0.41	0.03	0.34	0.53
F_{ROH} (%)		2.50	0.20	1.70	3.20
ADMEL	38				
N_{ROH}		12.42	2.62	7.00	17.00
S_{ROH} (Mb)		4.96	0.91	2.73	6.83
L_{ROH} (Mb)		0.40	0.05	0.30	0.50
F_{ROH} (%)		2.20	0.40	1.20	3.10

Table 3: Homozygosity islands for Apis mellifera carnica (CAR) and Apis mellifera mellifera (MEL), with characterised genes in the region.

Pop	Chr.	Begin (bp)	End (bp)	Length (Mb)	Characterised genes
CAR					
	1	20,503,934	20,691,670	0.19	*Uqcr11*
	3	45,010	795,764	0.75	*Ndufs1*
	3	4,292,409	4,539,710	0.25	*Nmdar1, uzip*
	5	109,374	769,034	0.66	*PHRF1, Chmp1*
	6	457,554	697,716	0.24	*5-ht7*
	8	486,627	1,847,468	1.36	*RpL35, CTL5, crh-BP, ATP5G2*
	9	831,231	1,400,908	0.57	*Grp, Rep, LCCH3, Amel_8916*
	9	2,670,767	2,901,055	0.23	*Gpdh, Tpx-4*
	11	3,737,355	7,286,231	3.55	*CPR1, CPR2, CPR3, CPR4, Rga*
	13	1,041,176	1,747,547	0.71	*SP34*
	15	2,649,222	2,877,623	0.23	*Snf*
MEL					
	4	1,352,064	2,061,845	0.006	*GstS1*
	6	13,132,241	13,364,154	0.004	*Pban*
	9	1,493,548	2,109,752	0.005	*WRNexo,*
	10	17,777	464,733	0.004	*Uvop*
	15	701,467	985,135	0.004	*Mad*

Uqcr11: ubiquinol-cytochrome c reductase complex III subunit XI,*Ndufs1:* NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase), *Nmdar1* : Glutamate [NMDA] receptor subunit 1, *uzip* : Unzipped, *PHRF1* : PHD and RING finger domain-containing protein 1, *Chmp1:* chromatine modifying protein 1, *5-ht7* : serotonin receptor 7, *RpL35:* ribosomal protein L35, *CTL5:* C-type lectin 5, *crh-BP:*corticotropin-releasing hormone binding protein, *ATP5G2:* ATP synthase H+ transporting mitochondrial F_0 complex, subunit C2 (subunit 9), *Grp* : glycine-rich cuticle protein,*Rep:* Rab escort protein, LCCH3: ligand-gated chloride channel homolog 3, Amel_8916: cys-loop ligand-gated ion channel subunit 8916,*Gpdh:* glycerol-3-phosphate dehydrogenase, *Tpx-4* : thioredoxin peroxidase 4, *CPR* : cuticular protein, *Rga:*regulator of gene activity protein, *SP34* : venom serine protease 34, *Snf* : U1 small nuclear ribonucleoprotein A, *GstS1:*glutathione S-transferase S1, *Pban* : pheromone biosynthesis-activating neuropeptide, *WRNexo* : WRN exonuclease,*Uvop* : ultraviolet-sensitive opsin, *Mad* : MAX dimerization protein

Figure 1: **Photographs of Apis mellifera mellifera and Apis mellifera carnica worker bees sampled in Switzerland.**

Figure 2: **High-resolution population network of honey bees (Apis mellifera)** . Each colony is illustrated by a node, with individual node size proportional to F_{ROH}, whilst the node color represents the sample origin (A) and the individual levels of admixture (B) at the optimal cluster solution (K = 5). The thickness of edges varies in the proportion of the genetic distance to visualize individual relationship between the colonies. The topology of the network clearly differentiated Apis mellifera carnica (CAR) from Apis mellifera mellifera (MEL) (dashed circles) and described well-defined substructure within MEL according to the genetic origin (patrilines and conserved MEL).

Figure 3: **Identified homozygosity islands across the honey bee (Apis mellifera) genome** (16 chromosomes). Breed-specific homozygosity islands of Apis mellifera carnica (CAR) are illustrated in dark grey; those of Apis mellifera mellifera (MEL) are in brown and the common segment in light grey.

A. m. mellifera

A. m. carnica