



**HAL**  
open science

## **Identification of runs of homozygosity in Western honey bees ( *Apis mellifera* ) using whole-genome sequencing data**

Annik Imogen Gmel, Matthieu Guichard, Benjamin Dainat, Geoffrey Rhys Williams, Sonia Eynard, Alain Vignal, Bertrand Servin, Markus Beestrong Consortium, Markus Neuditschko

### ► **To cite this version:**

Annik Imogen Gmel, Matthieu Guichard, Benjamin Dainat, Geoffrey Rhys Williams, Sonia Eynard, et al.. Identification of runs of homozygosity in Western honey bees ( *Apis mellifera* ) using whole-genome sequencing data. *Ecology and Evolution*, 2023, 13 (1), pp.e9723. <10.1002/ece3.9723>. <hal-03993788>

**HAL Id: hal-03993788**

**<https://hal.inrae.fr/hal-03993788v1>**

Submitted on 17 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

## RESEARCH ARTICLE

# Identification of runs of homozygosity in Western honey bees (*Apis mellifera*) using whole-genome sequencing data

Annik Imogen Gmel<sup>1</sup> | Matthieu Guichard<sup>1,2</sup> | Benjamin Dainat<sup>2</sup> | Geoffrey Rhys Williams<sup>3</sup> |  
Sonia Eynard<sup>4,5</sup> | Alain Vignal<sup>4,5</sup> | Bertrand Servin<sup>4,5</sup> | the Beestrong Consortium<sup>6</sup> |  
Markus Neuditschko<sup>1</sup> 

<sup>1</sup>Animal GenoPhenomics, Animal Production Systems and Animal Health, Agroscope, Posieux, Switzerland

<sup>2</sup>Swiss Bee Research Centre, Agroscope, Liebefeld, Switzerland

<sup>3</sup>Department of Entomology and Plant Pathology, Auburn University, Auburn, Alabama, USA

<sup>4</sup>GenPhySE, INRAE, INPT, INPENVT, Université de Toulouse, Castanet-Tolosan, France

<sup>5</sup>UMT PrADE, Protection des Abeilles Dans L'Environnement, Avignon, France

<sup>6</sup>Domaine de Vilvert Bat 224, Jouy-en-Josas Cedex, France

## Correspondence

Markus Neuditschko, Animal GenoPhenomics, Animal Production Systems and Animal Health, Agroscope, Posieux, Switzerland.  
Email: [markus.neuditschko@agroscope.admin.ch](mailto:markus.neuditschko@agroscope.admin.ch)

## Funding information

Swiss Federal Office for Agriculture, Grant/Award Number: 627000708

## Abstract

Runs of homozygosity (ROH) are continuous homozygous segments that arise through the transmission of haplotypes that are identical by descent. The length and distribution of ROH segments provide insights into the genetic diversity of populations and can be associated with selection signatures. Here, we analyzed reconstructed whole-genome queen genotypes, from a pool-seq data experiment including 265 Western honeybee colonies from *Apis mellifera mellifera* and *Apis mellifera carnica*. Integrating individual ROH patterns and admixture levels in a dynamic 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 queens. Despite the current applied conservation efforts, we pinpointed 79 admixed queens. Genomic inbreeding ( $F_{ROH}$ ) strongly varied within and between the identified subpopulations. Conserved *A. m. mellifera* from Switzerland had the highest mean  $F_{ROH}$  (3.39%), while queens originating from a conservation area in France, which were also highly admixed, showed significantly lower  $F_{ROH}$  (0.45%). The majority of *A. m. carnica* queens were also highly admixed, except 12 purebred queens with a mean  $F_{ROH}$  of 2.33%. Within the breed-specific ROH islands, we identified 14 coding genes for *A. m. mellifera* and five for *A. m. carnica*, respectively. Local adaptation of *A. m. mellifera* could be suggested by the identification of genes involved in the response to ultraviolet light (*Crh-BP*, *Uvop*) and body size (*Hex70a*, *Hex70b*), while the *A. m. carnica* specific genes *Cpr3* and *Cpr4* are most likely associated with the lighter striping pattern, a morphological phenotype expected in this subspecies. We demonstrated that queen genotypes derived from pooled workers are useful tool to unravel the population dynamics in *A. mellifera* and provide fundamental information to conserve native honey bees.

Annik Imogen Gmel and Matthieu Guichard equally contributed to the manuscript.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## KEYWORDS

*Apis mellifera carnica*, *Apis mellifera mellifera*, conservation, genetic diversity, pooled sequences, selection signatures

## TAXONOMY CLASSIFICATION

Conservation genetics

## 1 | 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 subspecies 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 et al., 2017; Ruttner, 1988). These lineages are characterized by genetic differences leading to variable morphology, physiology, and behavior (Ruttner, 1988). Honey bees are commonly kept in hives for honey production and pollination purposes. Varying selection pressures have been applied by humans to honey bees within their native range: in Europe, several selection programs have been initiated to increase their productivity (Adam, 1983; Büchler et al., 2010; Chauzat et al., 2013; Guichard et al., 2020; Uzunov et al., 2017), while in Africa the majority of honey bees evolved without large-scale selection (Dietemann et al., 2009).

In the beginning of the 19th century, importation of foreign honey bees among European regions began to increase, which profoundly reshaped the genetic structure of this species (Parejo et al., 2020). Historically, native honey bees of Europe mainly belong to A, M, and C evolutionary lineages. They are locally adapted to different climatic and geographical regions, resulting in several subspecies (Momeni et al., 2021; Ruttner, 1988). Nevertheless, beekeepers in Northern Europe continue to replace native honey bees (e.g., *A. m. mellifera*) with honey bees of South-European origin (e.g., *A. m. carnica* and *A. m. ligustica*), as these subspecies are considered to be more productive, gentle and calm (Bouga et al., 2011; Guichard et al., 2021). In North America, most honey bees are hybrids of these two historically imported strains, and key selection traits in United States (US) breeding programs are productivity and resistance traits to certain pathogens (Saelao et al., 2020). In Northern Europe, the favored use of South-European honey bees has led to multiple admixture events between subspecies and the extinction of native honey bees (Bieńkowska et al., 2021; Ruttner, 1995). Furthermore, these bees are also threatened by the widespread use of stabilized hybrid strains such as Buckfast (Adam, 1983; Bieńkowska et al., 2021).

The relocation of subspecies accompanied by admixture is a major risk factor for the loss of local adaptation and genetic diversity of honey bees (De la Rúa et al., 2009). Therefore, in Europe, several conservation programs have been initiated to maintain the genetic diversity of native honey bees, by establishing conservation areas on islands (e.g., Denmark, Scotland, and the Canary Islands) or on the mainland (e.g., France, Norway, Slovenia, and Austria), and excluding hybrids and invasive breeds mainly by their morphotype or behavior (De la Rúa et al., 2009). In Switzerland, the first conservation area of

*A. m. mellifera* was established in 1977 in canton Glarus under a legal framework (Soland-Reckeweg et al., 2009). Nowadays, an additional conservation area exists in canton Obwalden. The two conservatories encompass a total area of 830km<sup>2</sup> and ~1050 colonies (Parejo et al., 2016). To limit admixture events with other non-native 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 program including several mating stations. These stations are also located in geographically isolated areas and consist of 10 to 20 selected drone-producing colonies. Currently, an ancestry-informative marker panel (microsatellites or single nucleotide polymorphisms; SNPs) is applied to determine the hybridization of conserved and selected *A. m. mellifera* queens, and queens with an admixture level greater than 10% are replaced with purebred *A. m. mellifera* (Parejo et al., 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 their genetic diversity (Jones et al., 2004; Kryger, 1990; Mattila et al., 2012; Mattila & Seeley, 2014; Oldroyd et al., 1992), monitoring of inbreeding in small conserved populations, such as *A. m. mellifera* in Switzerland, is crucial.

Estimates of inbreeding indicate the probability that an animal receives alleles that are identical by descent from each parent. This can be estimated using genetic markers, while pedigree-based estimations require prior knowledge of individual ancestry (Kardos et al., 2015), which in case of the honey bee is often not available. Runs of homozygosity (ROH), caused by inheritance of parental haplotypes that are identical by descent, are one of the common methods to estimate inbreeding levels without ancestry information (McQuillan et al., 2008). 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 et al., 2018) for a complete review. Furthermore, it is possible to derive the genomic inbreeding coefficient ( $F_{ROH}$ ) of an animal by dividing the sum of all homozygous segments ( $S_{ROH}$ ) by the length of the analyzed genome (McQuillan et al., 2008). Numerous studies have demonstrated that overlapping ROH segments across individuals, so-called ROH islands can be found in breed-specific selection signatures in cattle (Purfield et al., 2012; Zhang et al., 2015), sheep (Mastrangelo et al., 2017; Purfield et al., 2017; Signer-Hasler et al., 2019), and horses (Druml et al., 2018; Grilz-Seger et al., 2018; Grilz-Seger, Druml, et al., 2019; 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).

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). Due to the hemizygosity of drones, ROH cannot be estimated based on such data and it becomes 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 paternal origins are involved in the formation of honey bee colonies (Estoup et al., 1994; Neumann et al., 1999; Tarpay et al., 2004). However, genotyping of honey bee queens for the evaluation of admixture and genomic inbreeding without harming them remains difficult (Bubnič et al., 2020; Madella et al., 2020). Therefore, a novel method for deriving queen genotypes based on pooled sequences of diploid worker bees was recently presented (Eynard et al., 2022), which could enable more genomic studies requiring diploid data in honey bees and other haplo-diploid eusocial insects.

In this study, we investigated the utility of queen genotypes derived from pooled worker sequences to ascertain population substructures and to identify ROH segments in honey bees. Furthermore, we integrated estimates of individual admixture and  $F_{ROH}$  in a dynamic population network visualization to enhance the genetic monitoring of conserved *A. m. mellifera*. Finally, we screened the genomes for ROH islands to detect genes associated with geographic adaptations and human-mediated selection within *A. m. mellifera* and *A. m. carnica*.

## 2 | MATERIAL AND METHODS

### 2.1 | Sampled colonies

We sampled 265 honey bee colonies from two different subspecies, namely *A. m. mellifera* (MEL) and *A. m. carnica* (CAR). Conserved MEL colonies were sampled in Switzerland (CS\_CH) and France (CS\_FR). The majority of the MEL colonies came from the selection program in Switzerland (SL\_CH), which represents five different paternal origins (P1–P5), that is, drone-producing colonies headed by sister queens. The sample size, geographic origin, and location of the five different paternal origins and conserved MEL colonies are summarized in Table 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 (CAR\_CH,  $n = 22$ ), Sweden (CAR\_SWE,  $n = 3$ ), Norway (CAR\_NOR,  $n = 3$ ), and the US (CAR\_US,  $n = 21$ ), while the majority of these colonies descended from open mating. For each colony, approximately 500 workers were sampled inside the hive on brood combs. Following this sample strategy, it was estimated to include all existing paternal origins among workers in the colony.

TABLE 1 Number of sampled colonies, geographic origin (CH = Switzerland, FR = France), legal framework and protection radius of paternal origins and conserved *Apis mellifera mellifera*

Paternal origins and conserved MEL	Number of sampled colonies	Geographic location	Canton (Ct.) and description of surroundings	Altitude (m)	Legal framework	Protection radius (distance to nearest independent apiary)
P1	17	Krauchtal, CH	Ct. Glarus, semi-isolated valley	1400	Cantonal law: beekeepers can only breed <i>A. m. mellifera</i>	1.8 km
P2	47	Gental, CH	Ct. Bern, isolated valley	1300	Cantonal law: protection of the mating station	3 km
P3	9	Säntis, CH	Ct. Appenzell Ausserrhoden, open valley	1100	Relies on the agreement between beekeepers	2.2 km
P4	34	Schilistal, CH	Ct. St-Gallen, semi-isolated valley	1100	Relies on the agreement between beekeepers	3.5 km
P5	39	Rothbach, CH	Ct. Luzern, isolated valley	1300	Relies on the agreement between beekeepers	4 km
CS_CH	45	Glarus, CH	Ct. Glarus, colonies widespread across the valley (from isolated to open)	400–1000	Cantonal law: beekeepers can only breed <i>A. m. mellifera</i>	min. 4 km (distance to Canton limit)
CS_FR	25	Savoie, FR	Colonies widespread across the valley (from isolated to open)	500–1700	Relies on the agreement between beekeepers	No apiary registration

## 2.2 | DNA extraction and pool sequencing

DNA extraction and pool sequencing of the sampled colonies are described in detail by Guichard et al. (2021). Briefly, approximately 500 workers per colony were shredded in a DNA extraction solution. Pair-end sequencing was performed on an Illumina™ HiSeq 3000 or a NovaSeq™ 6000 platform. To significantly decrease computing time, the pool sequence analysis was restricted to an informative marker panel including 7,023,977 genome-wide SNPs, as previously described by Wragg et al. (2022). Raw reads from pool sequencing of the 265 colonies were aligned to the honey bee reference genome Amel\_HAV3.1, Genbank 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 et al., 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 summarizing reference and alternative allele counts for each SNP.

## 2.3 | Reconstruction of queen genotypes and quality control

We used the method described in Eynard et al. (2022) to reconstruct honeybee queen genotypes. In brief, this method follows a two-step procedure using two statistical models. First, the genetic composition, in terms of proportion of the three main European honey bee subspecies (*A. m. mellifera*, *A. m. ligustica*, and *A. m. caucasica*) was estimated for each colony. For this purpose, reference allele frequencies for each of these subspecies were estimated from the data available from Wragg et al. (2022) and used as prior in the statistical model. Second, based on to estimated genetic composition (e.g., pure *A. m. mellifera*, pure *A. m. carnica* and hybrids) colonies were divided into different groups and queen genotype reconstruction was performed across colonies within such a group. On average our pool-seq data showed the same sequencing depth (~30X), which was used to simulate the aforementioned statistical models. Therefore, we expect the same genotype errors and accuracies, previously reported by Eynard et al. (2022).

After the reconstruction of genome-wide queen genotypes, we removed 99,555 SNPs with multiple alternative alleles and 207,904 SNPs with an excessively high and low sequencing depth. Furthermore, we excluded 771,835 homozygous SNPs to account for the very large non recombining, low polymorphic regions within the honey bee genome (Wragg et al., 2022). Finally, missing genotypes of the remaining 5,944,683 SNPs were imputed with BEAGLE 5.2 (Browning et al., 2018) to detect ROH segments along the genome, while for the population structure analyses queen genotypes were further edited for minor allelic frequency (MAF > 5%), which resulted in 1,609,447 genome-wide SNPs.

## 2.4 | Dynamic population network

To ascertain the high-resolution population structure of honey bees, we performed a dynamic population network visualization. The different components involved in the so-called NetView approach are described in detail by Neuditschko et al. (2012) and Steinig et al. (2016). Briefly, we computed genetic distances by subtracting pairwise relationships identical-by-state (IBS), as provided by PLINK v.1.9 (Chang et al., 2015), from 1 and applied the algorithm in its default setting (number of  $k$  nearest neighbors  $k$ -NN = 10). To illustrate the genetic relatedness between neighboring honey bee queens, we associated the thickness of edges (connecting lines) with the magnitude of the genetic distance, with thicker edges corresponding to lower genetic distances. To identify highly inbred honey bee queens, we scaled the node size of each queen based on the individual  $F_{ROH}$ . The node color denotes the sampled subpopulations and the individual level of admixture at  $K = 2$  and  $K = 7$  (the optimal number of clusters).

## 2.5 | Admixture

Queen admixture levels and genetic distances ( $F_{ST}$ ) between the subspecies were determined using the program Admixture 1.23 (Alexander et al., 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 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 dynamic population network, as described above.

## 2.6 | Runs of homozygosity

Continuous homozygous 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 40 kb, a maximum gap length of 100 kb, and a minimum length of homozygous segment of 200 kb. The total number of ROH ( $N_{ROH}$ ), the total length of ROH segments ( $S_{ROH}$ ), and the average length of ROH ( $L_{ROH}$ ) were summarized 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). Differences between subspecies and subpopulations were investigated using  $t$ -tests (for the two subspecies) and ANOVA with post hoc Tukey's honestly significant difference (HSD) tests at a significance level of  $\alpha < 0.05$  as implemented in the R package multcompView (Graves et al., 2015). We also correlated

$F_{ROH}$  with the admixture proportions at  $K = 2$  for each subspecies as implemented in the statistical computing software R (R Core Team, 2020). Furthermore, we compared  $F_{ROH}$  of 74 SL\_CH queens with pedigree-based inbreeding coefficients ( $F_{PED}$ ). Pedigree-based inbreeding coefficients of the selected queens were calculated following the method described by Brascamp and Bijma (2014) based on the pedigree information of 1082 *A. m. mellifera* queens (Guichard et al., 2020) born between 1991 and 2017. The identity of the queen, of her mother and the grand-mother (queen of the drone producing colonies) were largely known and used to establish a pedigree file, from which an inverse relationship matrix between all entries was calculated to determine  $F_{PED}$  (Guichard et al., 2020).

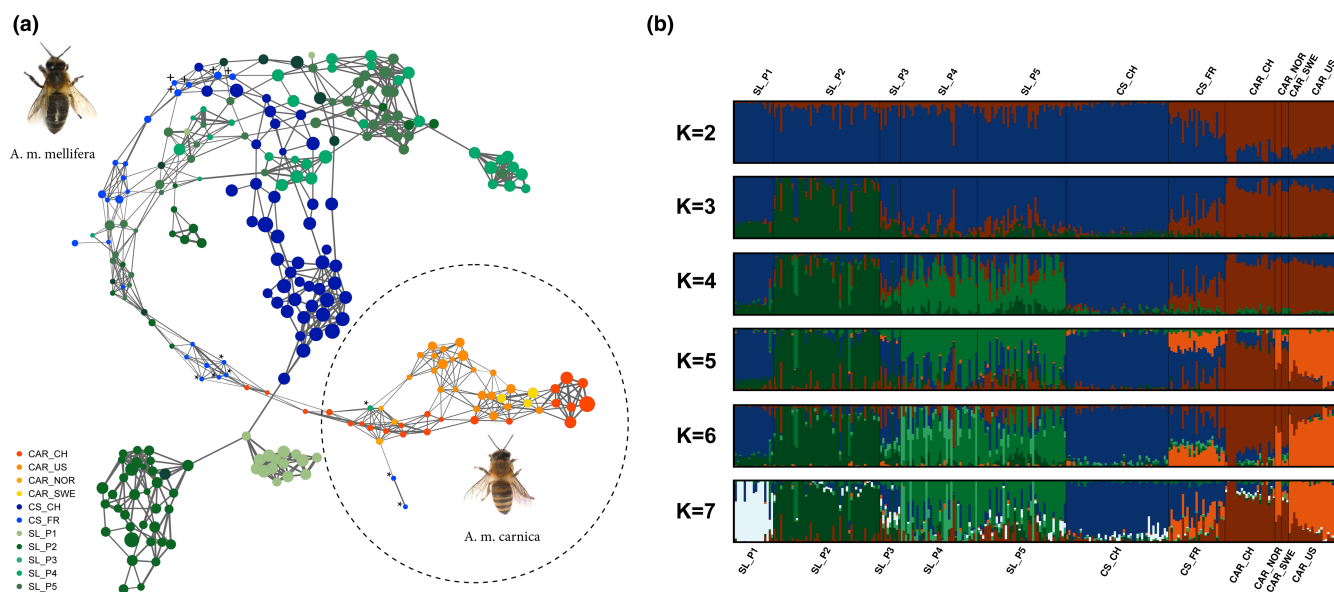
## 2.7 | Homozygosity islands and gene functions

Homozygosity islands of the three different groups (CAR, SL\_CH, and CS\_CH) were determined based on overlapping homozygous regions present in more than 50% of the queens with <10% admixture applying the R package detectRUNS (Biscarini et al., 2019). Considering the small sample size, we used all CAR with admixture proportions <10% for the identification of breed-specific ROH islands. Finally, 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 ROH islands and specified the known functions of the identified genes by conducting a literature review.

## 3 | RESULTS

### 3.1 | Dynamic population network

The dynamic population network separated CAR (Figure 1a, dashed circle) from MEL, while seven MEL queens (one SL\_CH, and six CS\_FR) were allocated in the immediate neighborhood of CAR (Figure 1a, indicated by “\*”). The hub between CAR and MEL included CS\_FR and SL\_CH queens that did not cluster with their respective genetic origin. The topology of the network additionally revealed that further substructures exist within MEL queens. The most evident substructures within MEL corresponded to CS\_CH queens and two selected strains (P1 and P2). It was interesting to see that five CS\_FR queens (Figure 1a, top left, indicated by “+”) were directly connected with four CS\_CH queens, while the remaining CS\_FR queens were frequently distributed over the network. Furthermore, CS\_CH queens were the nearest neighbors of five SL\_CH queens originating from three different strains (P1, P4, and P5), while P1 showed the strongest genetic relationship with this cluster. Compared to P1 and P2, the three remaining strains (P3–P5) did not build a distinct population cluster. P3 queens were distributed over the network without a discernible pattern and the majority of P4 and P5 queens were highly related to each other, while especially P5 queens built two small sub-clusters each including a P4 queen. Such a small sub-cluster was also evident in CAR including seven highly related queens. The association of the node size with  $F_{ROH}$  illustrates that the majority of CS\_CH and three CAR queens, included in the



**FIGURE 1** Dynamic population network and model-based clustering of honey bees (*Apis mellifera*). (a) Dynamic population network, where each queen is illustrated by a node, with individual node size proportional to  $F_{ROH}$ , while the node color represents the sample origin. The thickness of edges varies in the proportion of the genetic distance to visualize individual relationships between the colonies. The topology of the network clearly differentiated *Apis mellifera carnica* (CAR, dashed circle) from *Apis mellifera mellifera* (MEL) and described well-defined substructure within MEL according to the genetic origin. MEL queens allocated in the immediate neighborhood of CAR are indicated by “\*”, while CS\_FR queens directly connecting with CS\_CH queens are highlighted by “+”. (b) Model-based clustering assignment of honey bees using 2–7 clusters (K). Queens are presented by a single vertical column divided into K colors. Each color represents one cluster and the length of the colored segment corresponds to the individual membership proportion in that cluster.

forementioned sub-cluster, showed the highest  $F_{ROH}$ . Furthermore, it can be noted that CAR located in the neighborhood of MEL (and vice versa), as well as queens not clustering with their strains show in general lower  $F_{ROH}$  (Figure 1a).

### 3.2 | Admixture

Based on the cross-validation error estimation increasing  $K$  from 2 to 10, an optimal cluster solution at  $K = 7$  was determined (Figure S1). The first level ( $K = 2$ ) of model-based clustering clearly differentiated CAR from MEL with a  $F_{ST}$  of 0.45 (Figure 1b). This cluster solution simultaneously highlighted that except for CS\_CH and P1, all MEL subpopulations contained highly admixed queens, while CS\_FR showed the highest percentage of admixed queens. At the second ( $K = 3$ ) and third level ( $K = 4$ ) MEL was further differentiated by allocating P2, P4, and P5 queens into two distinct clusters. At the fourth ( $K = 5$ ) and fifth level ( $K = 6$ ), the CAR\_US queens built a distinct cluster and the common population cluster of P4 and P5 queens was further sub-structured, without separating P4 from P5 queens. Finally, at the optimal cluster solution ( $K = 7$ ), P1 queens were differentiated from the CS\_CH cluster. Therefore, the hierarchical population clustering (increasing  $K$  from 2 to 7) confirmed the findings of the dynamic population network. This high agreement between the two applied population structure methods also became visible by integrating the admixture levels at  $K = 2$  (Figure 2a) and  $K = 7$  (Figure 2b) in the dynamic population network, which simultaneously revealed that queens not clustering with their respective geographical origin and having low  $F_{ROH}$  were highly admixed (Figure 2a). This observation was also reflected by an overall high negative correlation ( $r = -.75$ ) between  $F_{ROH}$  and the respective admixture level of all queens (CAR and MEL) at  $K = 2$ . Furthermore, it can be noticed that the differentiation of the common population cluster of P4 and P5 queens at  $K = 6$  was associated with one of the aforementioned sub-clusters (Figure 2b, center), while the hierarchical clustering failed to also detect the other one within the common cluster (Figure 2b, top right).

### 3.3 | Runs of homozygosity

The number of ROH segments ( $N_{ROH}$ ), the total length of ROH ( $S_{ROH}$ ), and the  $F_{ROH}$  were significantly different between MEL and CAR, while the mean segment length was equal ( $L_{ROH} = 0.34 \pm 0.10$  Mb, Table 2). The number of ROH segments and total ROH length was nearly twice as high in MEL ( $N_{ROH} = 13.43 \pm 7.21$ ,  $S_{ROH} = 4.84 \pm 2.70$  Mb) as in CAR ( $N_{ROH} = 7.88 \pm 5.64$ ,  $S_{ROH} = 2.85 \pm 2.27$  Mb). The presence of admixture in both subspecies was reflected by the presence of queens with no ROH segments, resulting in high standard deviation (SD) values. Removing MEL queens with a CAR admixture proportion  $>10\%$  at  $K = 2$  ( $n = 79$ ) increased the mean number of segments ( $N_{ROH} = 17.41 \pm 4.61$ ) and total length of segments ( $S_{ROH} = 6.30 \pm 1.78$ ) in the new subset (MEL $_{<10\%}$ ). Only 12 CAR queens remained after

removing all samples with an admixture proportion  $>10\%$  (CAR $_{<10\%}$ ), 10 from Switzerland (CAR\_CH), and two from the US (CAR\_US). The number, total length, mean length of segments, and mean genomic inbreeding coefficient all increased ( $N_{ROH} = 13.67 \pm 5.16$ ,  $S_{ROH} = 5.14 \pm 2.53$ ,  $L_{ROH} = 0.36 \pm 0.06$ ,  $F_{ROH} = 2.33 \pm 1.15$ ), but the SD increased in all parameters except for the  $L_{ROH}$ .

Summarizing the  $F_{ROH}$  results of all MEL queens according to the a priori defined subpopulations underscored some results from the dynamic population network (Figure 3). Pairwise Tukey's HSD comparisons between the subpopulations revealed that the CS\_FR showed significantly (adjusted  $p$ -value  $>.05$ ) lower  $F_{ROH}$  ( $0.45 \pm 0.53$ ), while CS\_CH had a significantly higher mean  $F_{ROH}$  ( $3.39 \pm 0.77$ ) compared to the other subpopulations. P1 ( $F_{ROH} = 2.63 \pm 0.82$ ) was also significantly different from P5 ( $F_{ROH} = 1.84 \pm 0.93$ ), but not from P2 ( $F_{ROH} = 1.98 \pm 0.87$ ), P3 ( $F_{ROH} = 1.84 \pm 0.93$ ), and P4 ( $F_{ROH} = 2.39 \pm 1.18$ ). P2, P3, and P4 were neither significantly different from each other nor from P5. There were no significant differences in ROH values between the CAR subpopulations due to the small sample size.

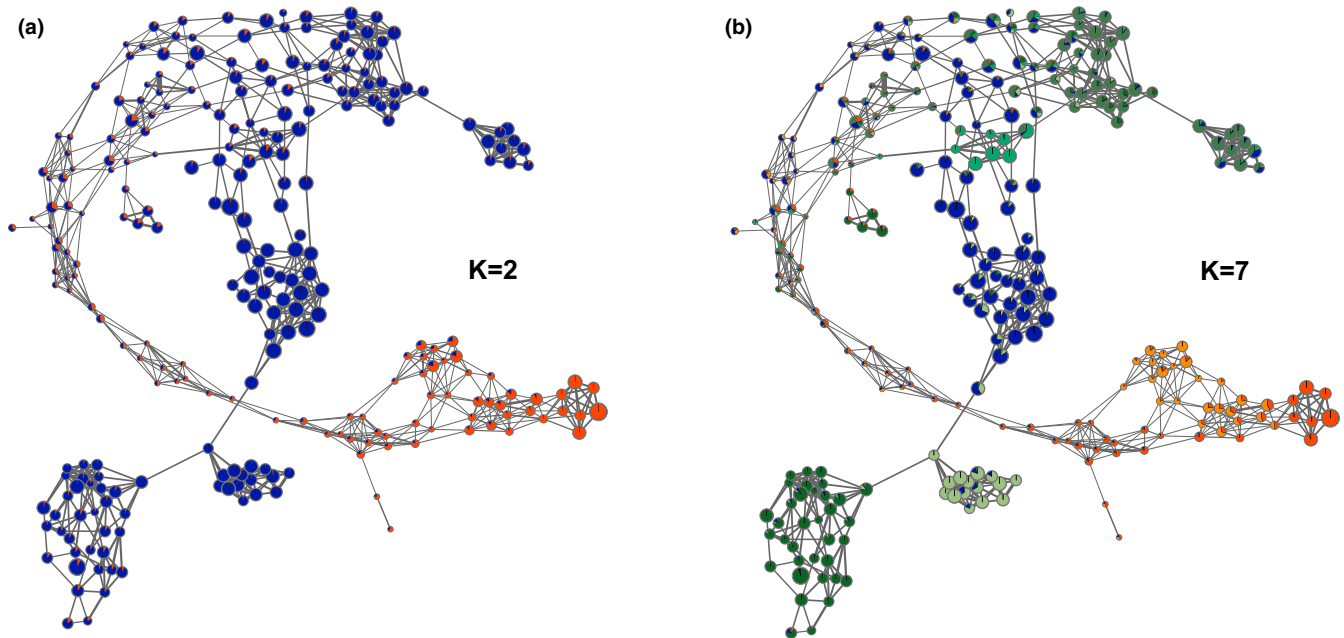
The  $F_{PED}$  values, of 74 SL\_CH queens including all strains except P4, ranged from 0.00 to 5.18%, with a mean of  $1.65\% \pm 1.41$ , whereas  $F_{ROH}$  ranged from 0.00 to 4.62%, with a mean of  $2.08\% \pm 0.95$ . The correlation between  $F_{PED}$  and  $F_{ROH}$  was slightly negative ( $r = -.22$ ).

### 3.4 | Homozygosity islands

After exclusion of all MEL queens with an admixture level greater than 10% at  $K = 2$ , we considered two subpopulations, SL\_CH $_{<10\%}$  ( $n = 94$ ), representing the selection lines, and CS\_CH $_{<10\%}$  ( $n = 43$ ) from the Swiss conservation area. We identified 15 SL\_CH $_{<10\%}$ -specific homozygosity islands distributed over nine chromosomes (Table S1). Considering only the CS\_CH $_{<10\%}$  queens there were 19 islands, with more on chromosomes 9 and 11, but fewer on chromosomes 6 and 10 than for SL\_CH $_{<10\%}$  queens, and one on chromosome 15 (Table S2). Twelve islands were overlapping, on chromosomes 1, 2, 3, 5, 6, 8, 9, and 11. Interestingly, there were islands on chromosomes 8, 9, 10, and 11 appearing in SL\_CH $_{<10\%}$  queens that were not present in CS\_CH $_{<10\%}$  queens.

Chromosomes 4, 7, 12, 13, 14, and 16 did not bear any homozygosity islands for either MEL subpopulation. Five homozygosity islands were located near the starting end of the chromosomes. The largest homozygosity island was on chromosome 2 for CS\_CH and covered 943 kb (common with SL\_CH $_{<10\%}$  over two smaller regions of 16 and 36 kb, respectively). The shortest was for MEL $_{<10\%}$  on chromosome 3 and spanned 15 kb. There were substantially more uncharacterized genes than annotated genes within the ROH islands (i.e., 264 uncharacterized loci and 12 annotated genes in MEL $_{<10\%}$ , respectively, 447 and 10 for CS\_CH). Table 3 summarizes the annotated genes embedded in the ROH islands.

The genes *Ndufs1*, *Phrf1*, *Chmp1*, *Grp*, *Rep*, and *Snf* were in homozygosity islands specific to CS\_CH queens. The genes *Rpl35*, *Ctl5*, *Crh-BP*, *ATP5G2*, *Tmem98*, *Twi*, and *Uvop* were in a homozygosity island specific to SL\_CH queens.



**FIGURE 2** Dynamic network visualizations of honey bees (*Apis mellifera*) associated with admixture proportions. Each queen is illustrated by a node, with individual node size proportional to  $F_{ROH}$ , while the node color represents the individual levels of admixture at  $K = 2$  (a) and  $K = 7$  (optimal cluster solution (b)). The thickness of edges varies in the proportion of the genetic distance to visualize individual relationship between the colonies. The topology of both networks illuminates that in general highly admixed queens also show low  $F_{ROH}$ .

**TABLE 2** Mean values, 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 *Apis mellifera carnica* (CAR) and *Apis mellifera mellifera* (MEL)

Subspecies	Sample size	Mean	SD	Min	Max
CAR	49				
$N_{ROH}$		7.88 <sup>a</sup>	5.64	0.00	23.00
$S_{ROH}$ (Mb)		2.85 <sup>a</sup>	2.27	0.00	10.97
$L_{ROH}$ (Mb)		0.34	0.10	0.00	0.51
$F_{ROH}$ (%)		1.29 <sup>a</sup>	1.03	0.00	4.97
MEL	216				
$N_{ROH}$		13.43 <sup>a</sup>	7.21	0.00	30.00
$S_{ROH}$ (Mb)		4.84 <sup>a</sup>	2.70	0.00	10.60
$L_{ROH}$ (Mb)		0.34	0.10	0.00	0.66
$F_{ROH}$ (%)		2.19 <sup>a</sup>	1.22	0.00	4.80

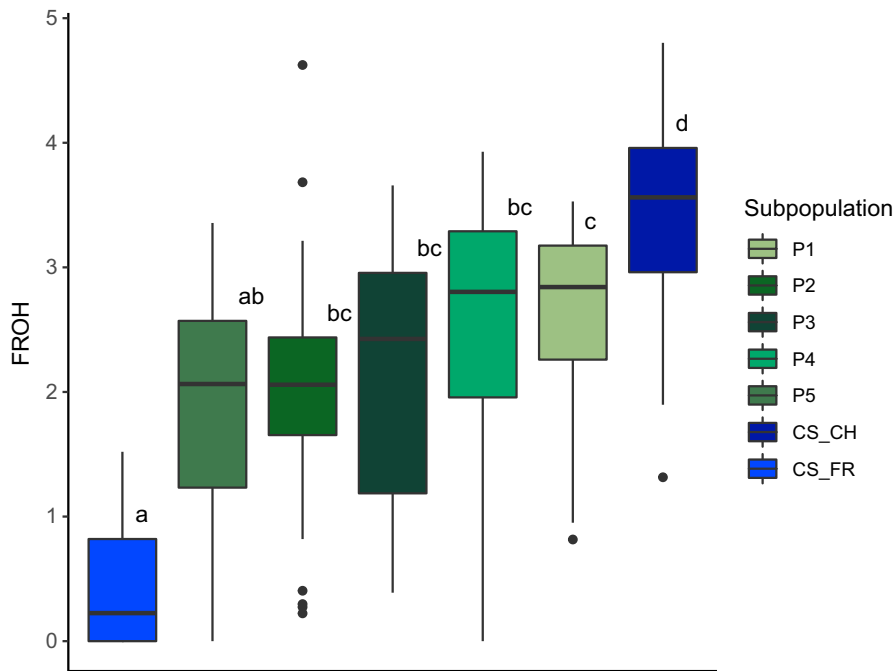
<sup>a</sup>Shows that the mean is different at a significance level of  $\alpha = 0.05$  based on a *t*-test.

For the 12 purebred CAR queens (CAR<sub><10%</sub>) we identified 11 ROH islands, all on chromosome 11, roughly spanning from 4,235,653 to 7,082,258 bp (Table S3). The shortest island was 25 kb and the longest island was 499 kb. The second longest, with 462 kb, contained five genes coding for cuticular proteins: *Cpr1*, *Cpr2*, *Cpr3*, and *Cpr4*, as well as a regulator of gene activity protein (*Rga*, Table 3). Similarly, to MEL, the homozygosity islands contained mainly uncharacterized loci (105) compared to the five annotated genes.

## 4 | DISCUSSION

We demonstrated that queen genotypes derived from pooled honey bee workers can be successfully applied to ascertain high-resolution

population structures, including the computation of  $F_{ROH}$  and the detection of breed- and subpopulation-specific ROH islands. However, it should be noticed, that the applied queen reconstruction procedure assumes that queens are mated to drones of similar ancestry. Our pool-seq data, where the majority of colonies were derived from breeders and conservatories fulfilled this prerequisite, while for other data settings (e.g., artificial insemination), the applicability of the queen reconstruction procedures needs to be further investigated. The applied population structure analyses clearly differentiated MEL from CAR ( $F_{ST} = 0.45$ ), despite the occurrence of highly admixed MEL and CAR queens and simultaneously highlighted the challenges to conserve native honey bees due to lack of control over mating. Therefore, the dynamic population network illustrated that a successful honey bee conservation program requires an appropriate



**FIGURE 3** Comparison of genomic inbreeding ( $F_{ROH}$ ) between the different *A. mellifera* subpopulations. Boxplot of the genomic inbreeding  $F_{ROH}$  in percent (%) for each *A. mellifera* (MEL) subpopulation. The horizontal line shows the median, the box extends from the lower to the upper quartile, and the whiskers to 1.5X the interquartile range above the upper quartile or below the lower quartile. Means not sharing any letter are significantly different based on Tukey's honest different means test with a  $p$ -value  $<.05$  adjusted for multiple testing.

**TABLE 3** Location and length of homozygosity islands (runs of homozygosity shared by  $>50\%$  of individuals) for *Apis mellifera mellifera* and *Apis mellifera carnica* containing characterized genes

Chr.	Start	End	Length (kb)	Subpopulation	Genes
2	2,030,227	2,065,958	35.73	CS_CH <sub>&lt;10%</sub> & SL_CH <sub>&lt;10%</sub>	<i>Tert</i>
3	309,688	692,594	382.91	CS_CH <sub>&lt;10%</sub>	<i>Ndufs1</i>
5	3723	446,000	442.28	CS_CH <sub>&lt;10%</sub>	<i>Phrf1</i>
5	483,461	656,850	173.39	CS_CH <sub>&lt;10%</sub>	<i>Chmp1</i>
8	2,038,809	2,354,163	315.35	SL_CH <sub>&lt;10%</sub>	<i>Rpl35, Ctl5, Crh-BP, ATP5G2</i>
8	2,355,518	2,416,454	60.94	SL_CH <sub>&lt;10%</sub>	<i>Tmem98, Twi</i>
8	11,659,873	11,857,385	197.51	CS_CH <sub>&lt;10%</sub> & SL_CH <sub>&lt;10%</sub>	<i>Hex70a, Hex70b</i>
8	11,857,385	11,857,837	0.45	CS_CH <sub>&lt;10%</sub>	<i>Hex70b</i>
9	842,007	1,263,372	421.37	CS_CH <sub>&lt;10%</sub>	<i>Grp, Rep</i>
9	1,812,584	1,983,997	171.41	CS_CH <sub>&lt;10%</sub> & SL_CH <sub>&lt;10%</sub>	<i>WRNexo</i>
10	189,237	414,181	224.94	SL_CH <sub>&lt;10%</sub>	<i>Uvop</i>
11	6,294,671	6,757,043	462.37	CAR <sub>&lt;10%</sub>	<i>Cpr1, Cpr2, Cpr3, Cpr4, Rga</i>
15	2,635,510	2,869,345	233.84	CS_CH <sub>&lt;10%</sub>	<i>Snf</i>

Note: SL\_CH<sub><10%</sub> are all *A. mellifera* queens with less than 10% admixture proportions from the selection lines, CS-CH<sub><10%</sub> are MEL queens from the Swiss conservation area with less than 10% admixture proportions, CAR<sub><10%</sub> are the *A. m. carnica* queens with less than 10% admixture proportions.

*Tert*: Telomerase Reverse Transcriptase, *Ndufs1*: NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase), *Phrf1*: PHD and RING finger domain-containing protein 1, *Chmp1*: chromatin modifying protein 1, *Rpl35*: ribosomal protein L35, *Ctl5*: C-type lectin 5, *Crh-BP*: corticotropin-releasing hormone binding protein, *ATP5G2*: ATP synthase H<sup>+</sup> transporting mitochondrial F<sub>0</sub> complex, subunit C2 (subunit 9), *Tmem98*: transmembrane protein 98, *Twi*: Twist, *Hex70A*: Hexamerin 70A, *Hex70B*: Hexamerin 70B, *Grp*: glycine-rich cuticle protein, *Rep*: Rab escort protein, *WRNexo*: WRN exonuclease, *Uvop*: ultraviolet-sensitive opsin, *Cpr*: cuticular protein, *Rga*: regulator of gene activity protein, *Snf*: U1 small nuclear ribonucleoprotein A.

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 highly 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 selected P1

queens might introduce foreign genetic variants to the CS\_CH gene pool.

Compared to CS\_CH, the origin of CS\_FR queens was only sporadically assessed in the past based on wing vein measurements, which simultaneously explains the highly observed diversity of the queens, whereas six queens showed a high genetic relatedness with

CAR. The population structure of CS\_FR and the genetic origin of some SL\_CH 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 et al., 2013), could be a more efficient alternative to maintain the gene pool of native honey bees.

In spite of the fact that SL\_CH are carefully selected to contribute to the local genetic diversity, MEL showed significantly higher  $F_{ROH}$  than CAR. However, this comparison must be moderated, as the majority of CAR queens originated from open mating not following a clear selection. This characteristic of CAR samples was also evident in the dynamic population network, which illuminated the high admixture levels and low  $F_{ROH}$  of CAR queens, whereas it was also possible to identify some purebred CAR, which showed similar  $F_{ROH}$  compared to MEL. Hence, for a comprehensive comparison between MEL and CAR population structure, further investigations are needed, especially involving more selected CAR queens.

The ROH results, according to the observed population structure (Figure 1a), confirmed the direct inverse relationship ( $r = -.75$ ) between admixture and ROH length in honey bees: there were fewer and shorter ROHs in queens with higher admixture proportions, concurrent to previous findings on the individual level in other livestock 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  $F_{PED}$ , with a slightly negative correlation ( $r = -.22$ ) indicating poor agreement between the two methods, compared to commonly observed values in livestock, such as goats ( $r = .50$ ; Burren et al., 2016) and sheep ( $0.18 < r < .70$ ; Purfield et al., 2017). In absence of instrumental insemination, the paternal origin of honey bees is not precisely known, as honey bee queens naturally mate in flight with 10 to 20 drones (polyandrous mating system; Estoup et al., 1994; Neumann et al., 1999; Tarpay 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 queens from insufficiently isolated mating stations (with higher admixture proportions) is 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 or by performing artificial insemination.

Within MEL-specific homozygosity islands, we identified two genes that are directly associated with the current applied selection traits, including increased productivity and swarming drive (Bouga et al., 2011; Guichard et al., 2021). Based on highly selected *A. m. ligustica* strains, it has already been demonstrated that *RpL35*, identified in the SL\_CH<sub><10%</sub>-specific island, controls royal jelly production and larval growth (Ararso et al., 2018). Furthermore, the differential expression of *Ndufs1*, found in a CS\_CH<sub><10%</sub>-specific ROH island, may also increase foraging behavior (Guo et al., 2019), and consequently, productivity.

The gene *Crh-BP* embedded in a ROH island for SL\_CH<sub><10%</sub>, is involved in the resistance to ultraviolet (UV) exposure, and therefore suggests adaptive mechanisms due to the ancestral geographical origin of the subspecies. The gene *Crh-BP* was shown to be upregulated in honey bees in response to UV exposure and heat stress (Even et al., 2012). Therefore, homozygosity in this gene could indicate local adaptation to lower sun exposure and temperatures of Northern Europe by potential downregulation of this gene. The homozygous state of the *Uvop* gene in SL\_CH<sub><10%</sub> is associated with retinal development and the circadian rhythm (Lichtenstein et al., 2018), which may enable SL\_CH<sub><10%</sub> to deal with seasonally more variable sun exposure. Diurnal mammal species also produce different quantities of UV-sensitive pigments depending on their ecological niche (Emerling et al., 2019). Furthermore, it has recently been demonstrated that genes involved in the response to UV exposure are associated with the local adaptation of horse breeds (Grilz-Seger, Neuditschko, et al., 2019).

Another indication of signatures of selection related to the geographical distribution of MEL can be found in the homozygosity of the genes *Hex70a* and *Hex70b*, encoding for two hexamerin proteins of the same name. Similar to vitellogenin, hexamerins bind to juvenile hormone (JH) and are storage proteins in the larval fat body, providing amino acids for the development into the adult stage (Martins et al., 2010; Telfer & Kunkel, 1991). Hexamerins also appear to be involved in ovary and testes development, and spermatogenesis in drones (Martins et al., 2011). More storage proteins in the larval fat body imply both larger and more long-lived bees, essential for colony survival during the winter. Although the quantity of storage proteins mostly depends on the pollen supply and quality (Basualdo et al., 2013; Frias et al., 2016), differences between queen strains have been observed under comparable feeding conditions, suggesting a genetic component (DeGrandi-Hoffman et al., 2021). We stipulate here that the ROH islands containing *Hex70a* and *Hex70b* could be linked to protein conversion efficiency, body size, and longevity of queens, allowing MEL to survive a longer winter period. This would be consistent with Bergman's rule, predicting that larger animals are better adapted to colder conditions (see Chole et al. (2019) for a review on bee size). The evolutionary emergence of longer-lived workers accumulating vitellogenin, another JH-binding protein, in MEL and CAR subspecies compared to the subtropical *A. m. scutellata* subspecies would suggest a similar adaptation in the quantity of accumulated hexamerin proteins in MEL (northern origin) compared to CAR (southern origin; Seehuus et al., 2006). Ruttner described both MEL and CAR as "large" (Ruttner, 1988), therefore objective studies measuring multiple workers from diverse *Apis mellifera* subspecies are necessary to confirm this hypothesis.

Several characterized genes in a homozygous state for MEL shared functions associated with stress response (*ATP5G2* (Watts et al., 2018), *Crh-BP* (Even et al., 2012), *Hex70b* (Aronstein et al., 2010)), and immunity (*Ct15* (Lin et al., 2020)). Two genes, *Wrmexo* and *Tert*, are involved in DNA structure and integrity, and therefore are thought to be associated with longevity (Hornstein, 2008; Robertson & Gordon, 2006; Rossi et al., 2010). 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 direction of selection within the MEL subspecies.

We also identified genes related to the exterior phenotype used to distinguish the two subspecies. The gene *Chmp1* identified in a CS\_CH<sub><10%</sub> ROH island is known to influence the veining pattern in *Drosophila* (Valentine et al., 2014), which might explain the morphological differences in vein patterns used to classify individuals into subspecies (Bouga et al., 2011; Ruttner, 1988). Several cuticular protein-coding genes (*Cpr3* and *Cpr4* in particular), present in the private ROH island of CAR<sub><10%</sub> may be involved in the CAR-specific morphotype of broader hairy stripes (Figure 1a), as they affect the thickness and coloring of the exoskeleton (Costa et al., 2016; Soares et al., 2013). We found another gene related to the cuticle (*Grp*, glycine-rich cuticle protein) in a CS\_CH<sub><10%</sub>-specific island, although its functions have not been formally described in the literature as far as we could discern. The *Twf* gene is involved in establishment of both the anterior-posterior and dorsal-ventral axes during embryogenesis, including the segmentation (i.e., stripes) of the abdomen (Wilson et al., 2014). Whether species-specific homozygosity in this gene could affect broadness of stripes was not specified. The identified genes discussed here are only a fraction of the loci found in the homozygosity islands for either MEL or CAR. The poor annotation of the current reference genome does not allow for a more thorough interpretation of our results.

In summary, we have described a number of novel aspects to investigate the genetic diversity of honey bees that are of potential interest. First, the application of queen genotypes derived from pooled honey bee workers to ascertain fine-scale population structures. Second, the identification of ROH segments to compute genomic inbreeding of honey bee queens. Finally, the identification of genes associated with geographic adaptation and human-mediated selection by means of ROH islands. Therefore, we believe that ROH derived from whole-genome sequencing data will be of invaluable benefit to investigate complex population structures in honey bees and other insects.

## AUTHOR CONTRIBUTIONS

**Annik Imogen Gmel:** Formal analysis (equal); visualization (equal); writing – original draft (equal). **Matthieu Guichard:** Data curation (equal); formal analysis (equal); investigation (equal); visualization (equal); writing – original draft (equal). **Benjamin Dainat:** Conceptualization (equal); data curation (equal); supervision (equal); writing – review and editing (equal). **Geoffrey Rhys Williams:** Data curation (equal); writing – review and editing (equal). **Sonia Eynard:** Data curation (equal); methodology (equal); writing – review and editing (equal). **Alain Vignal:** Data curation (equal); methodology (equal); writing – review and editing (equal). **Bertrand Servin:** Data curation (equal); methodology (equal); writing – review and editing

(equal). **the Beestrong Consortium:** Data curation (equal); writing – review and editing (equal). **Markus Neuditschko:** Conceptualization (lead); data curation (equal); formal analysis (lead); funding acquisition (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

## ACKNOWLEDGMENTS

We thank Lucie Genestout (Labogena) for coordinating the Beestrong project, Christina Sann and Yannick Poquet (Labogena) for organizing the sampling in Savoie, and Alice Revel (Agroscope) for her help in organizing 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-Pyrénées ([www.bioinfo.genotoul.fr](http://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 Génétiques pour les Espèces Animales, Domaine de Vilvert, 78353 Jouy-en-Josas, France), François 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 (Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France), Olivier Bouchez (GeT-PlaGe, Genotoul, INRAE Auzeville, 31320, Castanet-Tolosan, France).

## FUNDING INFORMATION

Financial support for this study was provided by Bundesamt für 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.

## DATA AVAILABILITY STATEMENT

Swiss, Swedish, Norwegian, and American bee sequence data will be deposited at the European Nucleotide Archive (ENA: <http://www.ebi.ac.uk/ena>), while French bee sequence data remain the property of the Beestrong Consortium. However, data are available from the authors upon reasonable request.

## ORCID

Markus Neuditschko  <https://orcid.org/0000-0001-7824-701X>

## REFERENCES

- Adam, B. (1983). *In search of the best strains of bees*. Northern Bee Books.
- 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., Meng, L., & 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.
- Aronstein, K. A., Murray, K. D., & Saldívar, E. (2010). Transcriptional responses in honey bee larvae infected with chalkbrood fungus. *BMC Genomics*, 11(1), 1–12.
- Basualdo, M., Barragán, S., Vanagas, L., García, C., Solana, H., Rodríguez, E., & Bedascarrasbure, E. (2013). Conversion of high and low pollen protein diets into protein in worker honey bees (hymenoptera: Apidae). *Journal of Economic Entomology*, 106(4), 1553–1558.
- 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. The R Project Retrieved June 2019.
- Bouga, M., Alaux, C., Bienkowska, M., Büchler, R., Carreck, N. L., Cauia, E., Chlebo, R., Dahle, B., Dall'Olio, R., de la Rúa, P., Gregorc, A., Ivanova, E., Kence, A., Kence, M., Kezic, N., Kiprijanovska, H., Kozmus, P., Kryger, P., Conte, Y. L., ... Wilde, J. (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.
- Bubnič, J., Mole, K., Prešern, J., & Moškrič, A. (2020). Non-destructive genotyping of honeybee Queens to support selection and breeding. *Insects*, 11(12), 896.
- 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., Drögemüller, C., & 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), 220–234.
- 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), 7.
- 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.
- Chole, H., Woodard, S. H., & Bloch, G. (2019). Body size variation in bees: Regulation, mechanisms, and relationship to social organization. *Current Opinion in Insect Science*, 35, 77–87.
- Cobey, S. W., Tarpay, 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 Rúa, P., Jaffé, R., Dall'Olio, R., & Muñoz, I. (2009). Biodiversity, conservation and current threats to European honeybees. *Apidologie*, 40(3), 263–284.
- DeGrandi-Hoffman, G., Corby-Harris, V., Carroll, M., Toth, A. L., Gage, S., Watkins, J. E., Graham, H., Chambers, M., Meador, C., & Obernesser, B. (2021). The importance of time and place: Nutrient composition and utilization of seasonal pollens by European honey bees (*Apis mellifera* L.). *Insects*, 12(3), 235.
- Dietemann, V., Pirk, C. W. W., & Crewe, R. (2009). Is there a need for conservation of honeybees in Africa? *Apidologie*, 40(3), 285–295.
- Druml, T., Neuditschko, M., Grilz-Seger, G., Horna, M., Ricard, A., Mesarič, M., Cotman, M., Pausch, H., & 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. (1819). 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*, 2015(282), 20151817.
- Estoup, A., Solignac, M., & Cornuet, J.-M. (1994). Precise assessment of the number of patrilineal 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.
- Eynard, S. E., Vignal, A., Basso, B., Canale-Tabet, K., Le Conte, Y., Decourtye, A., Genestout, L., Labarthe, E., Mondet, F., & Servin, B. (2022). Reconstructing queen genotypes by pool sequencing colonies in eusocial insects: Statistical methods and their application to honeybee. *Molecular Ecology Resources*, 22, 3035–3048.
- Frias, B. E. D., Barbosa, C. D., & Lourenço, A. P. (2016). Pollen nutrition in honey bees (*Apis mellifera*): Impact on adult health. *Apidologie*, 47(1), 15–25.
- Graves, S., Piepho, H.-P., & Selzer, M. L. (2015). Package 'multcompView'. Visualizations of Paired Comparisons.
- 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., Cotman, M., Horna, M., Dobretsberger, M., Brem, G., & Druml, T. (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., Beestrong Consortium, & Neuditschko, M. (2021). Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences. *Animal Genetics*, 52(4), 472–481.
- Guichard, M., Neuditschko, M., Soland, G., Fried, P., Grandjean, M., Gerster, S., Dainat, B., Bijma, P., & Brascamp, E. W. (2020). Estimates of genetic parameters for production, behaviour, and health traits in two Swiss honey bee populations. *Apidologie*, 51, 876–891.
- Guo, Y., Fu, B., Qin, G., Song, H., Wu, W., Shao, Y., Altaye, S. Z., & 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.
- Hornstein E. (2008). *Longevity in the Honeybee (Apis mellifera): Expression of Telomerase and Insulin Signaling Pathway Genes in Queen and Worker Bees*. *Journal of Young Investigators*. <https://www.jyi.org/2008-february/2017/11/12/longevity-in-the-honeybee-apis-mellifera-expression-of-telomerase-and-insulin-signaling-pathway-genes-in-queen-and-worker-bees#>. Accessed 10th January 2023.
- 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., Wu, J., & Brewer, L. (2020). Homozygosity mapping reveals population history and trait architecture in self-incompatible pear (*Pyrus* spp.). *Frontiers in Plant Science*, 11, 11.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- 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., Xia, X., & You, M. (2020). Immune responses to bacillus thuringiensis in the midgut of the diamondback moth, *Plutella xylostella*. *Developmental & Comparative Immunology*, 107, 103661.
- Madella, S., Grubbs, K., & Alburaki, M. (2020). Non-invasive genotyping of honey bee Queens *Apis mellifera* L.: Transition of the Dral mtDNA COI-COII test to in silico. *Insects*, 12(1), 19.
- Martins, J. R., Anhezini, L., Dallacqua, R. P., Simoes, Z. L., & Bitondi, M. M. (2011). A honey bee hexamerin, HEX 70a, is likely to play an intranuclear role in developing and mature ovarioles and testicles. *PLoS One*, 6(12), e29006.
- Martins, J. R., Nunes, F. M., Cristino, A. S., Simões, Z. L., & Bitondi, M. M. (2010). The four hexamerin genes in the honey bee: Structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. *BMC Molecular Biology*, 11(1), 1–20.
- 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.
- McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., ... Wilson, J. F. (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., Gut, I. G., & 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., Farajzadeh, L., Bendixen, C., Căuia, E., Charrière, J. D., Coffey, M. F., Costa, C., Dall'Olio, R., de la Rúa, P., Drazic, M. M., Filipi, J., Galea, T., Golubovskii, M., Gregorc, A., ... Estonba, A. (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.
- Oldroyd, B. P., Rinderer, T. E., Harbo, J. R., & Buco, S. M. (1992). Effects of intracolony 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.
- Pavan, S., Delvento, C., Mazzeo, R., Ricciardi, F., Losciale, P., Gaeta, L., D'Agostino, N., Taranto, F., Sánchez-Pérez, R., Ricciardi, L., & Lotti, C. (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. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Robertson, H. M., & Gordon, K. H. (2006). Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera*. *Genome Research*, 16(11), 1345–1351.
- 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., Sherman, A., & 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.
- Saelao, P., Simone-Finstrom, M., Avalos, A., Bilodeau, L., Danka, R., de Guzman, L., Rinkevich, F., & Tokarz, P. (2020). Genome-wide

- patterns of differentiation within and among US commercial honey bee stocks. *BMC Genomics*, 21(1), 1–12.
- Seehuus, S.-C., Norberg, K., Gimsa, U., Krekling, T., & Amdam, G. V. (2006). Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proceedings of the National Academy of Sciences*, 103(4), 962–967.
- Signer-Hasler, H., Burren, A., Ammann, P., Drögemüller, 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., Simões, 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-Reckweg, 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 SNPs. *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.
- Telfer, W. H., & Kunkel, J. G. (1991). The function and evolution of insect storage hexamers. *Annual Review of Entomology*, 36(1), 205–228.
- Uzunov, A., Brascamp, E., & Büchler, 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., Mikheyev, A. S., Robertson, H. M., Robinson, G. E., & 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.
- Wilson, M. J., Kenny, N. J., & Dearden, P. K. (2014). Components of the dorsal-ventral pathway also contribute to anterior-posterior patterning in honeybee embryos (*Apis mellifera*). *EvoDevo*, 5(1), 1–11.
- Wragg, D., Eynard, S. E., Basso, B., Canale-Tabet, K., Labarthe, E., Bouchez, O., Bienefeld, K., Bieńkowska, M., Costa, C., Gregorc, A., Kryger, P., Parejo, M., Pinto, M. A., Bidanel, J. P., Servin, B., le Conte, Y., & Vignal, A. (2022). Complex population structure and haplotype patterns in the Western European honey bee from sequencing a large panel of haploid drones. *Molecular Ecology Resources*, 22(8), 3068–3086.
- Wragg, D., Marti-Marimon, M., Basso, B., Bidanel, J.-P., Labarthe, E., Bouchez, O., le Conte, Y., & 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.
- 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.

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Gmel, A. I., Guichard, M., Dainat, B., Williams, G. R., Eynard, S., Vignal, A., Servin, B.; the Beestrong Consortium., & Neuditschko, M. (2023). Identification of runs of homozygosity in Western honey bees (*Apis mellifera*) using whole-genome sequencing data. *Ecology and Evolution*, 13, e9723. <https://doi.org/10.1002/ece3.9723>