

Article

Whole-Genome Sequence Analysis of Italian Honeybees (*Apis mellifera*)

Giulietta Minozzi ^{1,*}, Barbara Lazzari ^{2,†}, Maria Grazia De Iorio ¹, Cecilia Costa ³, Emanuele Carpana ³, Paola Crepaldi ⁴, Rita Rizzi ¹, Elena Facchini ⁵, Gustavo Gandini ¹, Alessandra Stella ² and Giulio Pagnacco ²

- ¹ Department of Veterinary Medicine DIMEVET, University of Milan, 20133 Milan, Italy; mariagrazia.deiorio1993@gmail.com (M.G.D.I.); rita.rizzi@unimi.it (R.R.); gustavo.gandini@unimi.it (G.G.)
- ² IBBA-CNR, 20133 Milano, Italy; barbara.lazzari@gmail.com (B.L.); stella@ibba.cnr.it (A.S.); giulio.pagnacco@unimi.it (G.P.)
- ³ CREA-Research Centre for Agriculture and Environment, via di Corticella 133, 40128 Bologna, Italy; cecilia.costa@crea.gov.it (C.C.); emanuele.carpana@crea.gov.it (E.C.)
- ⁴ Department of Agricultural and Environmental Science-Production, Landscape, Agroenergy DISAA, University of Milan, 20133 Milan, Italy; paola.crepaldi@unimi.it
- ⁵ Hendrix Genetics Research, Technology & Services B.V., Spoorstraat 69, 5831 CK Boxmeer, The Netherlands; elena.facchini@hendrix-genetics.com
- * Correspondence: Giulietta.minozzi@unimi.it
- † These authors have equally contributed to this work.

Simple Summary: The purpose of this study was to (i) explore the population structure of the *A.m. ligustica* which is widely distributed along the entire Italian peninsula, (ii) quantify the introgression of *A.m. carnica*, Buckfast, and *A.m. mellifera* bees in the two autochthonous Italian subspecies *A.m. ligustica* and *A.m. sicula*, and to (iii) to propose conservation strategies for the two autochthonous subspecies. Whole-genome sequencing was performed by Illumina technology obtaining a total of 4,380,004 single nucleotide polymorphisms (SNPs). Results of the analysis of the patterns of genetic variation allowed us to identify and subgroup bees according to their type. Morphometric analysis of 5800 worker bees was in agreement with genomic data. The investigation revealed the genetic originality of the Sicula, and in *A.m. ligustica* limited genetic introgression from the other breeds.

Abstract: At the end of the last glaciation, *Apis mellifera* was established in northern Europe. In Italy, *Apis mellifera ligustica* adapted to the mild climate and to the rich floristic biodiversity. Today, with the spread of *Varroa destructor* and with the increasing use of pesticides in agriculture, the Ligustica subspecies is increasingly dependent on human action for its survival. In addition, the effects of globalization of bee keeping favored the spread in Italy of other honeybee stocks of *A. mellifera*, in particular the Buckfast bee. The purpose of this study was to characterize the Italian honeybee's population by sequencing the whole genome of 124 honeybees. Whole genome sequencing was performed by Illumina technology, obtaining a total coverage of 3720.89X, with a mean sample coverage of 29.77X. A total of 4,380,004 SNP variants, mapping on Amel_HAv3.1 chromosomes, were detected. Results of the analysis of the patterns of genetic variation allowed us to identify and subgroup bees according to their type. The investigation revealed the genetic originality of the Sicula, and in *A.m. ligustica* limited genetic introgression from the other breeds. Morphometric analysis of 5800 worker bees was in agreement with genomic data.

Keywords: whole genome sequence; honeybee; SNPs; biodiversity; conservation; morphometry



Citation: Minozzi, G.; Lazzari, B.; De Iorio, M.G.; Costa, C.; Carpana, E.; Crepaldi, P.; Rizzi, R.; Facchini, E.; Gandini, G.; Stella, A.; et al. Whole-Genome Sequence Analysis of Italian Honeybees (*Apis mellifera*). *Animals* **2021**, *11*, 1311. <https://doi.org/10.3390/ani11051311>

Academic Editor: Michael E. Davis

Received: 15 March 2021
Accepted: 29 April 2021
Published: 2 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

At the end of the last glaciation, *Apis mellifera* (A.m.) was established in northern Europe spreading in numerous subspecies [1]. The long climate optimum during the last 10,000 years has consolidated their adaptation to different eco-geographical areas.

The *A.m.* presents over 30 subspecies described in the 1980s by Friedrich Ruttner on a morphometric basis and by Father Adam (Karl Kehrle) of Buckfast Abbey, on production and behavioral traits [2,3].

The native distribution of *A. mellifera* encompasses, in addition to Europe, Western Asia, Africa, and the Middle East. The different subspecies have been divided into 4 main branches: A (African subspecies), C (eastern European subspecies), M (western and northern European subspecies) and O (Middle Eastern subspecies) [2]. The high level of honeybee genetic diversity reflects differentiation and adaptation to different climatic and ecological conditions [4].

In Italy, the *Apis mellifera ligustica*, adapted to the temperate climate and to the rich floristic biodiversity of the territory. Thanks to its favorable characteristics, this subspecies has spread since 1800 in many other regions of the world by human mediated action. Today, specimens of *A.m. ligustica* can be found in North America, in northern and central Europe [5,6], in Reunion island [7], in Kangaroo Island in southern Australia [8], in China [9], in Mauritius island [10] and in Brazil [11].

Furthermore, another autochthonous subspecies is present in Italy, in Sicily, the *A.m. sicula/siciliana*, a variety close to the *A.m. intermissa* subspecies, endemic in the South side of the Mediterranean Sea [12]. Additionally, we observe two allochthone subspecies, *A. m. mellifera* along the western Ligurian coast and *A.m. carnica* on the northeastern side of the Alps [13]. Before the introduction of modern beekeeping practices, which affected the natural presence of local strains of bees, the conservation of genetic diversity of the two autochthonous subspecies in Italy was not an issue. Geographic distribution had, in fact, limited overlapping areas.

With the spread of *Varroa destructor* in the 1980s, and with the increasingly dramatic use of pesticides in agriculture, the Italian bee is increasingly dependent on human action for its conservation, both for the control of parasites and for the threat posed by pesticides. In addition, the effects of globalization have favored the spread throughout Italy of other honeybee stocks of *A. mellifera*, commonly retained to be better performing in terms of productivity and resistance to pathogens, in particular the Buckfast bees. In fact, during the last 30 years, the importation of non-ligustica bees in Italy has significantly increased [1]. Similarly, in the rest of Europe the regular practice of importing exotic strains, from distant evolutionary lineages, has exposed native honeybees to progressive hybridization [9]. In this scenario, it becomes urgent to provide protection of the local varieties from the phenomena of genetic erosion and admixture, and to define strategies for the conservation of local breeds and their valorization on a productive scale.

To differentiate honeybee subspecies several tools are available and widely used in different countries. Some are based on the use of morphometric analyses on the wing venation [11]. Concerning DNA analysis, studies have been conducted with microsatellites [6,13,14], and mitochondrial DNA [7,10,15–17]. Early studies conducted on the Italian populations found hybridization events between the *A.m. ligustica* and *A.m. mellifera* subspecies on the Ligurian coast and along the western Alpine arch [18] by analyzing allozymes at the Mdh-1 locus. Furthermore, hybrid populations in the Friuli area between *A.m. ligustica* and *A.m. carnica* were identified [19]. A more recent study based on microsatellite loci confirms the introgression of foreign alleles into the *A.m. ligustica* population and suggests an absence of (specific ecotypes) regional genetic substructure within *A.m. ligustica* [13]. More recently, studies have been based on whole-genome sequences conducted on a regional and worldwide scale [4,5,9,20]. Whole genome sequencing provides the opportunity of investigating genetic variation at millions of loci, while providing, at the same time, relevant information on the function of these variants. The availability of genome information and accurate phenotypes can be used to identify genes and pathways involved in fertility, adaptation, and social and immune-related traits [21] as well as other economically important traits [9].

The aim of this study was to characterize the Italian honeybee's population, using next-generation sequencing (NGS) techniques. In particular, we wanted to quantify the

introgression of *A.m. carnica*, Buckfast, and *A.m. mellifera* bees in the two autochthonous Italian subspecies *A.m. ligustica* and *A.m. sicula*, and to explore the population structure of the *A.m. ligustica* which is widely distributed along the entire Italian peninsula, and finally to propose conservation strategies for the two autochthonous subspecies.

2. Materials and Methods

2.1. Sampling

During summer 2018 and spring 2019, 124 worker bees were sampled in 12 Italian regions (Figure 1). Some of the beekeepers involved in this study were members of the Ligustica and Sicilian herd book, held and created in 1997 by the CREA-Agriculture and Environment Research Centre for Ligustica and Sicula subspecies, and some did not have a registered affiliation. During collection, all beekeepers were asked about the type of bees they breed.

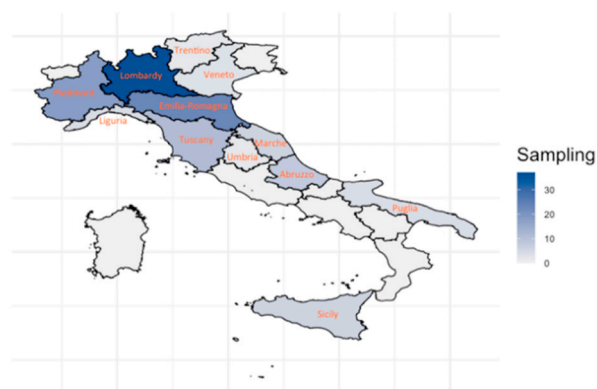


Figure 1. Map of Italy with density of bees sampled per region.

According to genomic information the samples consisted of:

- 61 Ligustica honeybees (*A.m. ligustica*) from Piedmont (2), Lombardy (6), Emilia-Romagna (23), Tuscany (11), Umbria (2), Marche (6), Abruzzo (8) and Puglia (3). We will refer to them as Ligustica.
- 6 Sicilian bees (*A. m. sicula*). We will refer to them as Sicula.
- 8 Carnica bees (*A. m. carnica*) from Lombardy (5), Piedmont, Trentino and Veneto (1 each). We will refer to them as Carnica.
- 4 Mellifera bees (*A.m. mellifera*) along the western Riviera coast (Liguria). These samples actually represent a population located at the intersection of the Italian Ligustica and the French *A.m. mellifera*. Local beekeepers describe it as a typical stabilized hybrid well fitted to the local environment. We will refer to them as Mellifera.
- 2 hybrid specimens Carnica by Ligustica (one in Lombardy and one in Trentino). We will refer to them as Carnica x Ligustica.
- 43 Buckfast bees sampled in Lombardy (25), Veneto (2), Piedmont (15) and in Apulia (1). We will refer to them as Buckfast.

For each of the 124 colonies, one newly hatched worker bee was sampled for DNA analysis. The bee was individually placed in a 1.5 mL Eppendorf tube containing 95% ethanol and kept at -8°C until samples were processed for DNA analysis.

Fifty worker bees from each of 116 colonies for a total of 5800 bees were enrolled for morphometric analysis. The 50 worker bees for each colony were stored in 50 mL tubes containing 95% ethanol and kept at -8°C until processing.

Morphometric analyses were performed at the CREA – Agriculture and Environment Research Centre laboratory according to Bouga et al. [22], based on 30 phenotypic measurements and tergite color.

2.2. DNA Extraction

Samples were processed by LCG genomics (LGC Genomics GmbH, Germany) from DNA extraction to NGS. Total DNA was extracted from 4 legs of each worker honeybee per colony. DNA isolation was performed with the sbeadexTM Livestock kit following the manufacturer's recommendations with few modifications. In detail, the lysis protocol D with lysis buffer P without PVP (polyvinylpyrrolidone) was used and DNA was eluted in 30 µL elution buffer AMP.

2.3. Library Preparation

Fragmentation of extracted DNA samples was performed using Covaris instrument (Covaris, Woburn, MA, USA) to obtain a desired fragment size of 300 bp. Sample purification was done using MiniElute Columns (QIAGEN, Redwood City, CA, USA), eluted in 20 µL EB buffer. Illumina library construction was performed with the Ovation Rapid DR Multiplex System 1–96 (NuGEN, San Carlos, CA, USA) according to the manual. Libraries were amplified for 10 cycles using MyTaq (Bioline) and standard Illumina primers. Illumina libraries were pooled and size selection was undertaken by gel electrophoresis selecting a range between 300 and 500 bp. Final library purification steps and quality control of DNA libraries was carried out via BioAnalyzer (AGILENT, Santa Clara, CA, USA). Illumina Nextera libraries were prepared for all samples and sequenced with a 150 bp paired-end read module (Illumina NextSeq 500/550 V2) following manufacturer recommendations. In total, more than 2760 million quality-trimmed read pairs were produced.

2.4. Sequence Processing and Alignment

Demultiplexing of all libraries as well as clipping of sequencing adapter remnants and Nextera mate pair linkers were performed with the Illumina bcl2fastq 2.17.1.14 software. Sequence quality control was assessed with FastQC [23]. Clipped Illumina reads were quality-trimmed at the 3'-end to get a minimum average Phred quality score of 20 over a window of 10 bases with Trimmomatic [24], and reads with final length < 20 bases were discarded. Trimmed reads were then mapped to the reference genome (Amel_HAv3.1: GCF_003254395.2) with BWA version 0.7.12 using the mem algorithm with default parameters [25]. The resulting BAM files were sorted with Samtools [26] and duplicate reads were marked using Picard [27].

2.5. Single Nucleotide Polymorphism (SNP) Calling, Annotation and Filtering

Variant discovery was performed with FreeBayes v1.0.2-16 [28], with standard filters. A total of 4,380,004 single nucleotide polymorphism (SNP) variants were discovered, mapping on Amel_HAv3.1 chromosomes (linkage groups LG1 to LG16 and mitochondrion). Variants mapping on unplaced scaffolds were discarded. Breed-specific SNP files were obtained for the six subspecies from the vcf file containing the total SNPs with vcftools [29], and filtered according to Kumar et al. [30]. Each breed-specific file was filtered with plink2 (-geno 0.05 -max-alleles 2 -maf 0.36) [31,32]. Plink2 was further used to exclude markers that fail the Hardy-Weinberg test (-hwe 0.001) and to perform linkage disequilibrium (LD) pruning (-indep-pairwise 50 5 0.01). SNPs from the six subspecies were then merged and pruned with the same parameters. The resulting dataset contains 44,811 SNPs.

2.6. SNP Analysis

The filtered dataset of 44,811 SNPs was used to perform principal component analysis (PCA) with plink, to build a phylogenetic tree, and to investigate population structure with the Admixture software [33]. Admixture from possible source populations was explored with K from 2 (lowest CV error) to 5. Admixture was further used to estimate pair wise Fst values, and the software Distruct [34] was used for graphical display of the genetic clustering.

The phylogenetic tree was built using a custom script, which builds a virtual sequence for each sample composed of variant positions. The sequences were aligned, and the align-

ment was input to the Phylip [35] programs seqboot (100 bootstraps), dnadist, neighbor and consensus. iTOL [36,37] was used to display the tree.

3. Results

3.1. Sequencing and Read Mapping

In total, 124 samples were sequenced with a total coverage of 3720.89X, obtaining a mean sample coverage of 29.77X (minimum coverage 15.55X and maximum coverage 56.27X). Mean number of raw, adapter and quality trimmed, and percentage of high quality trimmed reads are given in Table S1. A total of 4,380,004 SNP variants, mapping on Amel_HAv3.1 chromosomes, was detected (Table 1).

Table 1. Number of variable single nucleotide polymorphisms (SNPs) identified in each population.

Population	% SNP	Number of Samples	N° SNP
<i>A.m. ligustica</i>	64.19%	61	2,811,524
<i>Buckfast</i>	69.76%	43	3,055,491
<i>A.m. carnica</i>	46.67%	8	2,044,148
<i>Hybridcarnica</i>	29.89%	2	1,309,183
<i>A.m. mellifera</i>	49.74%	4	2,178,614
<i>A.m. sicula</i>	59.68%	6	2,613,986
All		124	4,380,004

3.2. Percentage of Specific and Common SNPs

The distribution across subspecies of SNPs in the filtered dataset is shown in Figure 2. In total, 42.78% of the selected SNPs are shared by all the subspecies. The *A.m. sicula* exhibits the highest number of private SNPs (2287), followed by the *A.m. mellifera* (120) and *A.m. ligustica* × *A.m. carnica* (105). This confirms that *A.m. mellifera* and especially *A.m. sicula* differ from the other subspecies. This result is in agreement with the subsequent genomic analyses and the fact that they derive from two different evolutionary lineages.

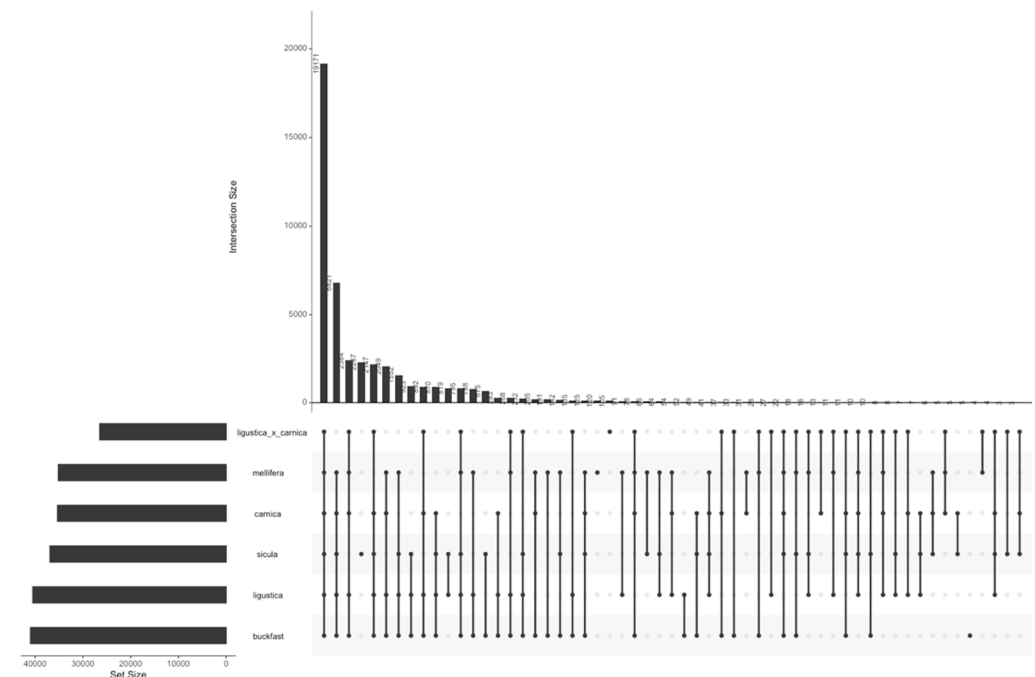


Figure 2. Distribution of specific and common SNPs across subspecies.

3.3. Principal Component Analysis (PCA)

The principal component analysis (PCA) of the 124 sampled honeybees based on the selected 44,811 SNPs described above was performed to assess the Italian population struc-

ture and genetic relationship among the different genetic groups (Figure 3). The first two components explain 38.1% of the total variance (21.5% and 16.6%, respectively), while the third component (9.1%) increases the distance of the Sicula samples and among the Mellifera samples.

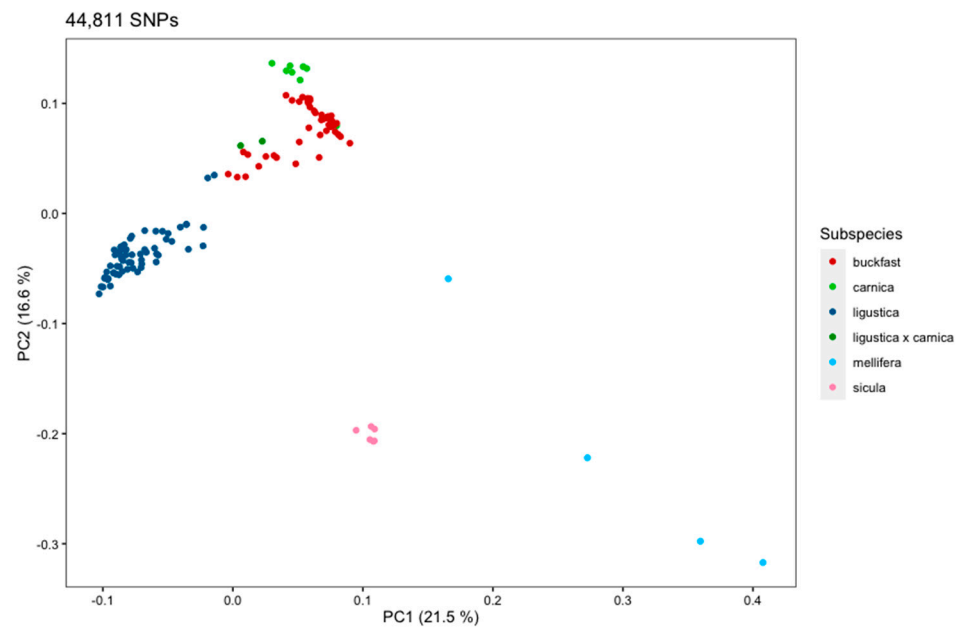


Figure 3. Principal component analysis (PCA) based on 2 dimensions (PC1 and PC2). Individual bee samples are indicated with colored dots corresponding to their subspecies.

The results show a clear separation among the sampled groups. The two dimensional PCA (Figure 3) displays this clear separation across genetic groups, with Ligustica, Buckfast, and Carnica, belonging to the common evolutionary C group, in the upper left quadrant. The clustering of Buckfast bees between Ligustica and Carnica testifies to their composite origin. This strain is known to have a highly admixed genome [38] as a number of different breeds were included in a long process of breeding for favorable traits. Nevertheless, at the Buckfast Abbey the selective practice, driven by Brother Adam along many decades in the 20th century, was grafted on a Ligustica strain imported in England to mend the damages produced to the English black bee population by the internal parasite *Acarapis woodi*. Furthermore, as foreseeable, the two Carnica \times Ligustica (crosses of Carnica queens with Ligustica drones) are situated between the two groups of origin.

Figure 3 also shows how the other two genetic groups belonging to different evolutionary lineages, Mellifera (group M) and Sicula (group A), are well separated from each other. The PCA revealed that the Mellifera samples show a wide internal variability probably because of their hybrid origin, also considering samples were taken at the periphery of its distribution area. By contrast, the Sicula samples are tightly clustered.

A limited number of samples had the breed declared by their beekeepers to be not coherent with their clustering position, such was the case of five colonies declared Ligustica, sampled in Apulia, Lombardy and Piedmont, in apiaries surrounded by Buckfast beekeepers. In reality, breeders were aware of a possible hybrid origin due to uncontrolled mating of the queen with males of Buckfast kept in nearby colonies, and in fact the five samples clustered in the Buckfast group.

3.4. Phylogenetic Tree

A phylogenetic tree of all the samples based on the filtered dataset of 44,811 SNPs is shown in Figure 4. Results display a clear concordance of the clustering identified previously by PCA analysis (Figure 3). The *A.m. sicula* and *A.m. mellifera* samples create two initial clearly separated branches. By contrast, the subspecies of group C are phyloge-

netically closer, and show a common root. Furthermore, a detailed analysis of the different branches of the tree that correspond to the Buckfast bees indicates that the Buckfast samples are split into four branches. The information provided by beekeepers allowed us to identify a possible explanation: the breeder of origin. Indeed, the samples forming the four branches correspond to four different breeders of origin. The first Buckfast group descended from a Luxemburg breeder. The second and the third groups have Danish and Austrian origins, respectively, whereas the fourth group is from a German breeder.

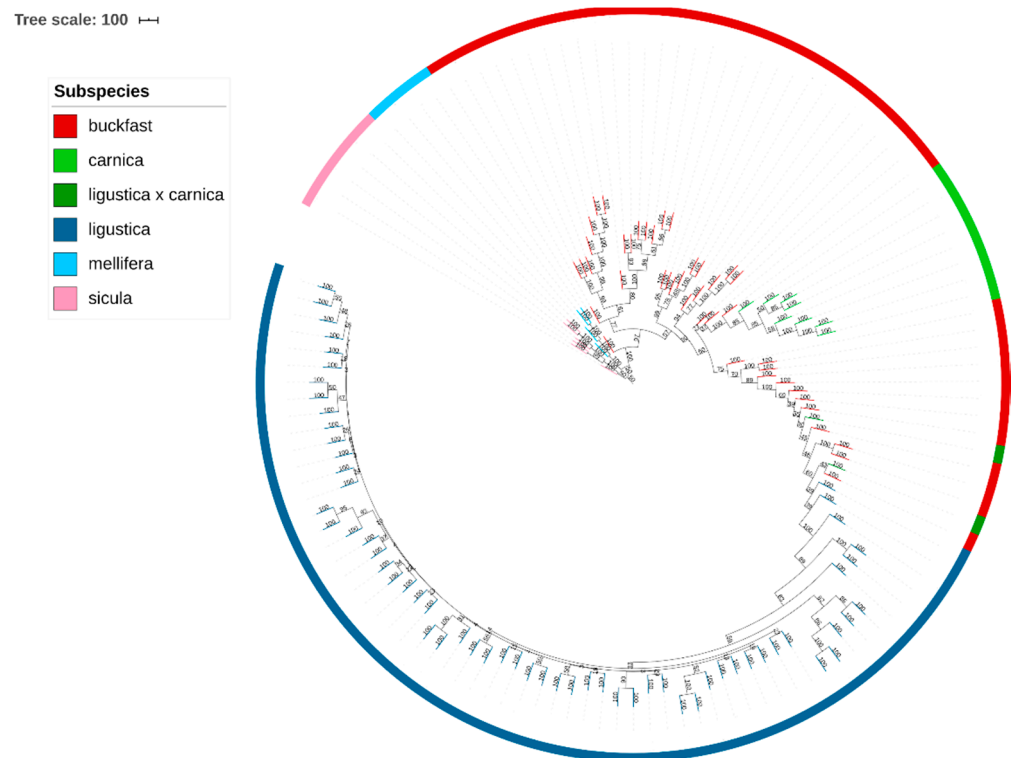


Figure 4. Phylogenetic tree based on 44,811 SNPs and 124 honeybee samples.

It is interesting to note that some Buckfast individuals are situated closer to the *A.m. carnica* samples and others nearer to *A.m. ligustica*, indicating the existence of different breeding strategies used by the different breeders. Finally, the *A.m. ligustica* samples are grouped in one branch, at the opposite side of the *A.m. sicula* samples, and next to the Buckfast.

3.5. Admixture

Population structure analyzed by Admixture software of the 124 individuals based on 44,811 SNPs, with K from 2 to 5, is shown in Figure 5. The relative CV errors steadily increasing from 0.21 to 0.22, are given in Table 2. With K > 5 (data not shown) the interpretation of the Admixture analysis becomes confusing. At K = 2, Ligustica shows an ancestry different from the other bee populations, but 45 individuals with limited proportions of the second ancestry. Increasing the source populations (K) up to 5, we observe the differentiations of Sicula and Mellifera and the marked admixture from at least three sources of the Buckfast bees. Finally, at K = 5, corresponding to five clusters identified by PCA (PC1 + PC2), we have Sicula and Carnica showing a single ancestry (for the Sicula not shared with other populations), the Buckfast with significant admixtures from the Carnica ancestry and secondly from the Ligustica ancestry. Carnica × Ligustica samples take an intermediate position between the two respective breeds. Ligustica maintains its originality, with 28 samples of >90% pure origin. The Q matrix of the proportion of membership of each predefined population in each of the 5 clusters based on K = 5 is shown in Table 3. Pairwise Fst values estimated by Admixture software are shown in Table 4. Fst ranges from

0.149 (*A.m. carnica* and Buckfast) to 0.389 (*A.m. carnica* and *A.m. mellifera*), with *A.m. mellifera* showing the highest average value (0.357), followed by *A.m. sicula* (0.248).

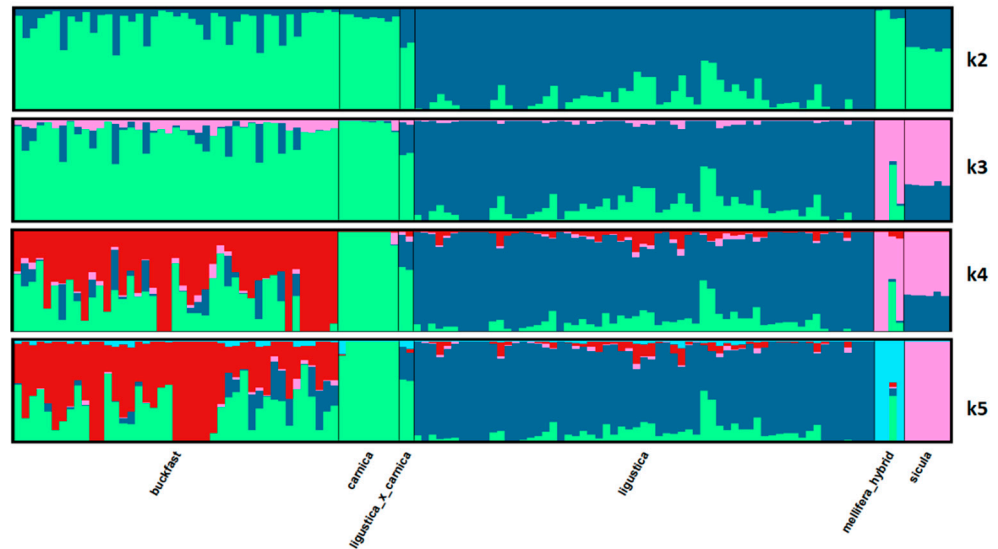


Figure 5. Admixture results of the 124 honeybee samples based on 44,811 SNPs with number of ancestral backgrounds (K = 5).

Table 2. CV error of the Admixture analysis of the 124 honeybee samples based on 44,811 SNPs based on 2 to 6 ancestral backgrounds (K).

Ancestral Background (K)	CV Error
2	0.206
3	0.210
4	0.214
5	0.223

Table 3. Results of clustering of the 124 honeybees sampled in 5 clusters (K = 5) according to Admixture analysis.

Given Population	Clusters					n
	1	2	3	4	5	
Buckfast	0.018	0.318	0.083	0.010	0.572	43
Carnica	0.016	0.982	0.000	0.000	0.002	8
Carnica × ligustica	0.067	0.608	0.306	0.000	0.019	2
Ligustica	0.008	0.079	0.863	0.011	0.039	61
Mellifera	0.853	0.113	0.017	0.003	0.013	4
Sicula	0.000	0.000	0.000	1.000	0.000	6

Table 4. Fst divergence between honeybee populations.

Population	<i>A.m. mellifera</i>	<i>A.m. carnica</i>	<i>A.m. ligustica</i>	<i>A.m. sicula</i>
<i>A.m. mellifera</i>				
<i>A.m. carnica</i>	0.389			
<i>A.m. ligustica</i>	0.383	0.153		
<i>A.m. sicula</i>	0.308	0.259	0.207	
Buckfast	0.346	0.149	0.153	0.217

3.6. Morphometric Analyses

A comparison was made between the breed assessed by the genomic approach and the breed declared by the breeder. For the 124 specimens, the asserted subspecies was confirmed in 119 cases with a 96% agreement. Five cases were initially asserted as *Ligustica*, while the SNP panel assigned them to the Buckfast breed. Beekeepers, in these few cases, were aware of the possible hybrid situation of their bees, which was in fact verified by the genomic analyses.

Among 116 samples that underwent morphometric analysis, inclusive of the tergite color test, 96 were in agreement with the SNP panel (82.8%).

Excluding two cases where a *Carnica* and a *Mellifera* were incorrectly classified as a *Ligustica* and a Sicilian, respectively, and three *Ligustica* samples erroneously qualified as Buckfast, the major challenge for morphometric analysis was the detection of the Buckfast breed, which was identified correctly only 25 times out of 38 (65.8%). Indeed, Buckfast are frequently regarded as somehow intermediate, or hybrid forms, between *Ligustica* and *Carnica*.

4. Discussion

In this study, we analyzed the genetic diversity of honeybee populations in Italy, giving particular attention to detect and understand the hybridization with the Buckfast honeybees imported in the last few decades. The overall aim was to provide knowledge to support the implementation of genetic conservation and selection programs. The availability of the honeybee genome allowed us to use next-generation sequencing methods to better define the honeybee subspecies present in Italy.

One hundred and twenty-four colonies were sampled in different regions of Italy and sequenced with a mean coverage of 29.77X, expanding and focusing in Italy the study conducted by Wallberg et al. [4] who sampled 140 colonies belonging to 14 different populations of *A. mellifera* from all over the world, including samples from the A, O, C and M groups. In total, we obtained 4,380,004 SNPs on 124 samples belonging mainly to *Ligustica* (61) and Buckfast samples (43) and including individuals of other subspecies. Wallberg et al., [4] identified a total of 8,282,459 SNPs, 1,745,809 of which were from the *Ligustica* samples. The higher number of total SNPs identified by Wallberg compared to our study is due to the inclusion of a higher number of lineages, i.e., the A and O groups. Regarding *Ligustica*, we identified 2,811,524 SNPs compared to the reference genome, which are almost double the number found by Wallberg et al. [4]. The two studies are based on different sample sizes, as we sampled 6 times the specimens of *Ligustica* used by Wallberg, and sequences with higher coverage. Furthermore, the *Ligustica* bees included in this study were collected from many geographical regions along the Italian peninsula.

The Buckfast bees in our study exhibited a high level of variability compared to the *Ligustica*, with 3,055,491 SNPs, in accordance with the composite nature of the population and its breeding structure, and in agreement with previous studies that identified a higher level of genetic diversity in Buckfast samples compared to the ancestral populations [4,5,38].

Despite the limited number of samples sequenced, the *Sicula* and *Mellifera* had a higher number of specific SNPs compared to the other subspecies included in the study. In the *Sicula* this is probably due to the contribution of the A lineage (African) already identified by microsatellite analysis in previous studies [1].

In general, 9.23% of SNPs (404,274) were polymorphic in all samples compared to the *Amel_Hav3.1* reference sequence [39].

The Buckfast samples, known to have a highly introgressed genome represented by their Eastern and Western European progenitor populations [38], clustered together, as previously observed by Parejo et al. [5] by means of PCA and high-resolution population network analysis, although their study contained samples of *A.m. mellifera*, *carnica* and Buckfast. In accordance with our clustering, the Buckfast were close to the *Carnica* samples [5].

Based on our results, the *Sicula* bees can be considered a pure breed, despite the limited sample size. All the bees clustered closely in the PCA analysis and created a separate branch

in the phylogenetic analysis. Furthermore the results of admixture confirmed at $K = 5$ the lack of recent hybridization events.

PCA analysis showed that despite Ligustica and Buckfast bees being present all over the country, their breeding management has not caused high levels of hybridization. The two populations create two distinct clusters on the PCA, which are further confirmed by the phylogenetic tree. The *A.m. ligustica* bees did not reveal geographic specific ecotypes based on our analysis, confirming previous results [13]. These outcomes indicate a high potential of adaptation of the Italian bee to the different climatic conditions present in Italy, confirmed by the practice of intensive nomadism that beekeepers undertake with their colonies in our country from the North to the South and back.

Lastly, referring to the 2 Ligustica \times Carnica hybrid bees, results of Admixture analysis confirmed PCA and phylogenetic results, allowing us to clearly disentangle the samples as Carnica \times Ligustica crosses. In general, all results of Admixture confirmed PCA and phylogenetic results.

The morphometric analysis for bee subspecies detection was reasonably good especially when the color of tergites was included. Nevertheless, some improvement could be achieved for Buckfast detection with a wider or different reference sample of Buckfast honeybees necessary for statistical comparison. Interestingly, in only five cases was the subspecies declared by the beekeeper an unsuspected and perhaps undesired genetic type: a Buckfast.

Since 2018, the Food and Agriculture Organization (FAO) has recognized the domesticated bee as a component of the animal genetic resources for food and agriculture, and the process to include the bee in DAD-IS, the FAO information System on breed-related information, has started [40]. In this direction, the conservation and valorization of the two Italian autochthonous breeds, the Ligustica and the Sicula, deserve specific attention like for any other domestic breed. Our investigation has provided a conservation status of the two breeds being somehow different.

For the Sicula, all analyses underlined its genetic originality and absence of genetic introgression from the other analyzed breeds. Therefore, it is imperative to guarantee in the future the identity of the Sicula from hybridization risks associated to nomadism and commerce.

A.m. ligustica samples did not show on average high levels of introgression from other breeds, however with the current development of bee breeding in Italy the picture could deteriorate. The widespread use by beekeepers of commercial queens purchased from trusted breeders, together with the practice of nomadism, poses the need to improve, for this species, a mating control strategy that is almost absent in the country today. The only way to achieve an efficient control of reproduction is through the identification of isolated areas for the reproduction of these two varieties in pureness. The identification of these areas is not easy given the high density of apiaries distributed throughout the national territory. Furthermore, the use of Italian islands is hampered by difficulties of connection and the short duration of the useful season, which changes quickly in arid summers of drought with a total disappearance of all flora. In addition, instrumental insemination cannot be seen as a solution due to the limited number of queens that this procedure could be applied to. A crucial element in the definition of the mating areas is that their location should be chosen in places not appealing to beekeepers looking for rich floristic pastures. Furthermore, specific municipal ordinances should be defined to limit these areas to the sole presence of apiaries intended for the production of specific drones. From the genetic point of view, the colonies that should produce the drones present in these areas must guarantee the widest genetic variability of the variety they intend to protect and must also represent the best lines of selection of beekeepers who will make them available. The assessment of whether these colonies belong to the required variety can be based on traditional morphometric evaluations, which we have shown here deserve a more precise definition, or on more advanced molecular tools such as the SNPs presented in this paper. With this approach, the control of mating will ensure a pure fertilization of the virgins that

will also merge with the commitment of achieving genetic improvement that sometimes makes non-native varieties more competitive than native ones.

Then, conservation and selection must flow together under the mandatory conditions to increase the mating control through artificial insemination and/or the creation of well-ordered mating stations. We believe that even a limited increase of mating control could help in creating a wider perception among beekeepers of the meaning of terms like pure breed, hybrid, and biodiversity conservation.

5. Conclusions

In conclusion, the results of the analysis based on whole genome sequencing of 124 Italian honeybees allowed us to identify and subgroup bees according to their type. For *A.m sicula*, all analyses underlined its genetic originality and absence of genetic introgression from the other analyzed breeds. *A.m. ligustica* samples did not show on average high levels of introgression from other breeds. However with the current development of bee breeding in Italy the picture could deteriorate. One possibility that could be applied to improve genetic conservation of local breeds in Italy is the creation of isolated mating sites for the reproduction of these two varieties in pureness.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ani11051311/s1>, Table S1: title: Total number of raw, adapter and quality trimmed, and percentage of quality trimmed reads per sample.

Author Contributions: Conceptualization, G.M. and G.P.; data curation: B.L.; formal analysis: B.L.; funding acquisition: G.P., G.M. and E.F.; sampling: P.C., G.P., G.M. and E.F.; morphometric analysis: E.C., C.C.; methodology: A.S.; writing original draft: G.M.; G.P.; M.G.D.I. and B.L.; writing, review and editing: A.S., G.G., R.R., E.F. and P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the BEENOMIX and BEENOMIX 2.0 projects funded by the Lombardy Region (FEASR program), PSR 2014–2020 (grant number 2016/00361532-G42F16000540002) and PSR (grant number 201801057971—G44I19001910002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: These data (sequences and phenotypes) are part of a reference populations used for selection by commercial breeders and have commercial value. Therefore, restrictions apply to the availability of these data, which are not publicly available. The authors can be contacted for a specific requests.

Acknowledgments: We thank all beekeepers and bee breeders for providing samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Franck, P.; Garnery, L.; Celebrano, G.; Solignac, M.; Cornuet, J.M. Hybrid origins of honeybees from Italy (*Apis mellifera ligustica*) and Sicily (*A. m. sicula*). *Mol. Ecol.* **2000**, *9*, 907–921. [[CrossRef](#)]
2. Ruttner, F. *Biogeography and Taxonomy of Honeybees*; Springer: Berlin/Heidelberg, Germany, 1988.
3. Adam, B. *Breeding the Honeybee*; Northern Bee Books: Mytholmroyd, UK, 1987.
4. Wallberg, A.; Han, F.; Wellhagen, G.; Dahle, B.; Kawata, M.; Haddad, N.; Simões, Z.L.P.; Allsopp, M.H.; Kandemir, I.; De La Rúa, P.; et al. Worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* **2014**, *46*, 1081–1088. [[CrossRef](#)] [[PubMed](#)]
5. Parejo, M.; Wragg, D.; Gauthier, L.; Vignal, A.; Neumann, P.; Neuditschko, M. Using whole-genome sequence information to foster conservation effort for the European dark honey bee, *apis mellifera mellifera*. *Front. Ecol. Evol.* **2016**. [[CrossRef](#)]
6. Péntek-Zakar, E.; Oleksa, A.; Borowik, T.; Kusza, S. Population structure of honeybees in the Carpathian basin (Hungary) confirms introgression from surrounding subspecies. *Ecol. Evol.* **2015**, *5*, 5456–5467. [[CrossRef](#)]
7. Wragg, D.; Techer, M.A.; Canale-Tabet, K.; Basso, B.; Bidanel, J.P.; Labarthe, E.; Bouchez, O.; Le Conte, Y.; Clémencet, J.; Delatte, H.; et al. Autosomal and mitochondrial adaptation following admixture: A case study on the honeybees of Reunion Island. *Genome Biol. Evol.* **2018**, *10*, 220–238. [[CrossRef](#)] [[PubMed](#)]

8. Aminimoghadamfarouj, N.; Nematollahi, A. Structure elucidation and botanical characterization of diterpenes from a specific type of bee glue. *Molecules* **2017**, *22*, 1185. [[CrossRef](#)]
9. Wragg, D.; Marimon, M.M.; Basso, B.; Bidanel, J.P.; Labarthe, E.; Bouchez, O.; Le Conte, Y.; Vignal, A. Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly. *Sci. Rep.* **2016**, *6*, 27168. [[CrossRef](#)] [[PubMed](#)]
10. Techer, M.A.; Clémencet, J.; Simiand, C.; Turpin, P.; Garnery, L.; Reynaud, B.; Delatte, H. Genetic diversity and differentiation among insular honey bee populations in the southwest Indian Ocean likely reflect old geographical isolation and modern introductions. *PLoS ONE* **2017**, *12*, e0189234. [[CrossRef](#)] [[PubMed](#)]
11. Francoy, T.M.; Gonçalves, L.S.; De Jong, D. Rapid morphological changes in populations of hybrids between Africanized and European honey bees. *Genet. Mol. Res.* **2012**, *11*, 3349–3356. [[CrossRef](#)]
12. Arias, M.C.; Sheppard, W.S. Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence. *Mol. Phylogenet. Evol.* **1996**, *5*, 557–566. [[CrossRef](#)] [[PubMed](#)]
13. Dall’Olio, R.; Marino, A.; Lodesani, M.; Moritz, R.F.A. Genetic characterization of Italian honeybees, *apis mellifera ligustica*, based on microsatellite DNA polymorphism. *Apidologie* **2007**, *38*, 207–217. [[CrossRef](#)]
14. Rahimi, A.; Mirmoayedi, A.; Kahrizi, D.; Abdolshahi, R.; Kazemi, E.; Yari, K. Microsatellite genetic diversity of *apis mellifera medaskorikov*. *Mol. Biol. Rep.* **2014**, *41*, 7755–7761. [[CrossRef](#)]
15. Il’yasov, R.A.; Poskryakov, A.V.; Petukhov, A.V.; Nikolenko, A.G. Genetic differentiation of local populations of the dark European bee *apis mellifera mellifera*. *Genetika* **2015**, *51*, 792–798. [[CrossRef](#)]
16. Il’yasov, R.A.; Poskryakov, A.V.; Nikolenko, A.G. Seven genes of mitochondrial genome enabling differentiation of honeybee subspecies *Apismellifera*. *Genetika* **2016**, *52*, 1176–1184.
17. Utzeri, V.J.; Ribani, A.; Fontanesi, L. Dispersion of *Apis mellifera* subspecies in the Emilia Romagna region (North of Italy) estimated using honey DNA analysis. Presented at 23rd ASPA Congress Italy, Sorrento, Italy, 11–14 June 2019.
18. Badino, G.; Celebrano, G.; Manino, A. Population structure and MDh-1 locus variation in *apis mellifera ligustica*. *J. Hered.* **1983**, *74*, 443–446. [[CrossRef](#)]
19. Nazzi, F. Morphometric analysis of honeybees from an area racial hybridization in northeastern Italy. *Apidologie* **1992**, *23*, 89–96. [[CrossRef](#)]
20. Parejo, M.; Wragg, D.; Henriques, D.; Charrière, J.D.; Estonba, A. Digging into the Genomic Past of Swiss Honey Bees by Whole-Genome Sequencing Museum Specimens. *Genome Biol. Evol.* **2020**, *12*, 2535–2551. [[CrossRef](#)]
21. Grozinger, C.M.; Zayed, A. Improving bee health through genomics. *Nat. Rev. Genet.* **2020**, *21*, 277–291. [[CrossRef](#)] [[PubMed](#)]
22. Bouga, M.; Alaux, C.; Bienkowska, M.; Büchler, R.; Carreck, N.L.; Cauia, E.; Chlebo, R.; Dahle, B.; Dall’Olio, R.; De la Rúa, P.; et al. A review of methods for discrimination of honey bee populations as applied to European beekeeping. *J. Apic. Res.* **2015**, *50*, 51–84. [[CrossRef](#)]
23. Babraham Institute. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 2 September 2020).
24. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
25. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
26. SAMtools. Available online: <http://samtools.sourceforge.net/> (accessed on 12 September 2020).
27. Picard. Available online: <https://broadinstitute.github.io/picard/> (accessed on 12 September 2020).
28. Cornell University. Available online: <https://arxiv.org/abs/1207.3907> (accessed on 12 September 2020).
29. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; De Pristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
30. Kumar, H.; Panigrahi, M.; Chotaray, S.; Pal, D.; Bhanuprakash, V.; Saravanan, K.A.; Shandilya, R.; Parida, S.; Bhushan, B. Identification of breed-specific SNP panel in nine different cattle genomes. *Biomed. Res.* **2019**, *30*, 78–81. [[CrossRef](#)]
31. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bende, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)] [[PubMed](#)]
32. Mass General Brigham. Available online: <http://pngu.mgh.harvard.edu/purcell/plink/> (accessed on 12 September 2020).
33. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [[CrossRef](#)] [[PubMed](#)]
34. Rosenberg Lab at Stanford University. Available online: <http://web.stanford.edu/group/rosenberglab/distruct.html> (accessed on 12 September 2020).
35. Phylip. Available online: <http://evolution.genetics.washington.edu/phylip.html> (accessed on 12 September 2020).
36. Interactive Tree of Life. Available online: <https://itol.embl.de> (accessed on 12 September 2020).
37. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **2007**, *23*, 127–128. [[CrossRef](#)] [[PubMed](#)]
38. Harpour, B.A.; Minaei, S.; Kent, C.F.; Zayed, A. Management increases genetic diversity of honey bees via admixture. *Mol. Ecol.* **2012**, *21*, 4414–4421. [[CrossRef](#)] [[PubMed](#)]

-
39. 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. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genom.* **2019**, *20*, 275. [[CrossRef](#)] [[PubMed](#)]
 40. FAO. Available online: <http://www.fao.org/3/CA0103EN/ca0103en.pdf> (accessed on 11 February 2020).