Sequence-based genome-wide association studies reveal the polygenic architecture of *Varroa destructor* resistance in Western honey bees *Apis mellifera*

Sonia E. Eynard[†1,2,11], Fanny Mondet[†3,11], Benjamin Basso[3,5,11], Olivier Bouchez[6], Tabatha Bulach[6], Yves Le Conte[3,11], Benjamin Dainat[8], Axel Decourtye[5,11], Lucie Genestout[2], Matthieu Guichard[8,9], François Guillaume[10], Emmanuelle Labarthe[1.11], Barbara Locke[7], Rachid Mahla[2], Joachim de Miranda[7], Markus Neuditschko[9], Florence Phocas[4,11], Yannick Poquet[2], Christina Sann[2], Rémi-Félix Serre[6], Kamila Tabet[1,11], Alain Vignal[§1,11], and Bertrand Servin[§1]

[1] GenPhySE, Université de Toulouse, INRAE, INP, ENVT, Castanet-Tolosan, 31320, France
[2] LABOGENA DNA, Palaiseau, 91120, France
[3] INRAE, UR 406 Abeilles et Environnement, Avignon, 84914, France
[4] Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, 78350, France
[5] ITSAP, Avignon, 84914, France
[6] GeT-PlaGe, Genotoul INRAE, Castanet-Tolosan, 31320, France
[7] Department of Ecology, Swedish University of Agricultural Sciences, 750-07, Uppsala, Sweden
[8] Agroscope Swiss Bee Research Centre, Bern 3003, Switzerland
[9] Animal GenoPhenomics, Agroscope, Posieux 1725, Switzerland
[10] Innoval, Noyal-sur-Vilaine 35538, France
[11] UMT PrADE, Avignon, 84914, France

Corresponding authors: sonia.eynard@inrae.fr, bertrand.servin@inrae.fr
†These authors contributed equally to the work
§These authors jointly supervised the work

Abstract

Honey bees, *Apis mellifera*, have experienced the full impacts of globalisation, including the recent invasion by the parasite *Varroa destructor*, hereafter called varroa. This deadly mite has rapidly become one of the main causes of colony losses worldwide, with a huge economic impact on beekeeping. Despite the lethal effects of varroa, some colonies survive infestation. In such colonies, honey bees are able to develop defence strategies that limit mite spread and confer colony resistance to varroa. In some populations, resistance traits were found to be sufficiently heritable to offer sustainable opportunities to fight varroa infestations through selective breeding of naturally resistant honey bees. Recent advances in genomics allow for the development of new tools to assess the importance of genetic mechanisms involved in varroa resistance. Here we report on the largest genome-wide association study performed on honey bees to understand the genetic basis of multiple phenotypes linked to varroa resistance in *Apis mellifera*. Genome wide association studies for varroa resistance traits were performed on whole genome sequencing of more than 1,500 colonies belonging to three different genetic backgrounds and combined in a meta-analysis. Results show that varroa resistance is largely polygenic in nature, involving many genes with small effects. Still, a total of 60 genetic markers were identified as having a significant impact on varroa resistance in at least one of the honey bee populations tested. These associations pinpoint several regions of the honey bee genome that are linked with varroa resistance phenotypes and increase our understanding of the mechanisms underlying honey bee resistance to its main parasite. Our results will also support the building of strategies for genomic selection in honey bee breeding.

Apis mellifera; *Varroa destructor*; varroa resistance; GWAS; mite non reproduction; recapping; pool sequencing; host-parasite interaction

Introduction

The honey bee, *Apis mellifera*, is a crucial contributor to sustainable food production [1]. However, for the past two decades beekeepers have been experiencing dramatic colony losses [2, 3]. Such losses are not sustainable for the beekeepers and the agroecosystems relying on services provided by honey bees. Extensive research has shown that honey bees are threatened by multiple factors: both abiotic factors with the loss of natural resources and the impact of pesticides due to agriculture intensification, and biotic factors with the infection by a diversity of pests and parasites that impair their survival [4]. Among biotic factors, the ectoparasite *Varroa destructor* is currently considered as the main threat to honey bee health and beekeeping worldwide [5]. In most regions of the world, colony losses have dramatically increased since its introduction in *Apis mellifera* populations in the early 80s [5, 6]. Originating from Asia, where a stable host-parasite relationship exists with its historical host *Apis cerana*, varroa now infests most *Apis mellifera* colonies worldwide. Varroa infests multiple compartments of the honey bee colony: it reproduces in the brood, feeds on adult honey bee haemolymph and fat body and favours virus infections [5, 7, 8]. Combined, these effects on individual bees lead to colony collapse within a few months if no actions are applied to control mite infestations [9]. To date, managing varroa infestation presents many constraints that offer beekeepers only a few unsustainable solutions to fight the deadly mite [7, 10].

However, since the beginning of the 1990's, colonies naturally surviving varroa infestation without treatment have been observed in several regions of the world and raised hope for beekeepers to overcome the problems caused by varroa infestation. In these surviving colonies, undergoing beekeeping activities, honey bees often display behavioural and physiological defences against the varroa parasite. These collective responses are expected to contribute to the limitation of parasite population growth and provide colonies with social immunity [11], a sustainable long-term adaptation to counter the immense damage caused by varroa. However, thus far sucj long term adaptation is not always reached, even in colonies expressing resistance traits (e.g. colonies in Gotland, Sweden [12, 13]). The defence repertoire against varroa includes hygiene behaviour targeted towards varroa parasitised brood cells, in the form of varroa sensitive hygiene (VSH [14]) or recapping behaviour ([15, 16]), and mechanisms suppressing mite reproduction (SMR [17]). Expression of these different traits lead to an increase in mite non reproduction within brood cells (MNR [15, 17], also called Decreased Mite Reproduction DMR [18]). This triggers lower mite population growth and infestation [10]. Such mechanisms have gained a major interest within the beekeeping sector and interest is growing to help decipher the genetic mechanisms underlying varroa resistance in the honey bee.

As for common livestock species, the most obvious way to induce varroa resistance into the honey bee population appears to be through selection and spreading of the most resistant lines. Since the late 90's efforts have been put into the selective breeding of such resistant honey bee lines [19, 20]. However, selecting for complex traits in honey bees has been hindered by features

that strongly differentiate honey bees from other typical livestock species. At the phenotyping level, honey bees being social insects, many traits of beekeeping interest, including varroa resistance traits, are expressed at the group level (the colony) and thus not directly unraveled by classical studies organized at the individual level (the individual bees). In addition, many of these traits, are linked to group behaviour and are thus difficult to phenotype and can display rather low repeatability and heritability [21, 22, 23, 24, 25]. At the breeding level, managing honey bee reproduction is difficult due to polyandry [26] (queen bees mate with multiple males) and to the fact that sex determination is governed by an haplodiploid mechanism at a single locus [27, 28] at which diploid homozygosity is effectively lethal [29], limiting drastically the possibility of inbreeding. Finally, honey bee populations bred for beekeeping encompass an important genetic diversity, with strong regional clustering, which impedes the identification of genetic markers that are valid outside the population where they are identified. Developing molecular tools to assess the genetic make-up of honey bee colonies could help in honey bee breeding by coping with some of these issues. To list a few possibilities, genomic-enabled prediction of resistance traits could reduce the amount of complex phenotyping required by identifying promising colonies early in life and genetic assessment of colonies could help identifying the genetic group of colonies [30] and allow to design crosses limiting inbreeding in selection programs.

In addition to their use in selection programs, molecular tools can help to identify pathways involved in a specific mechanism by performing genome-wide association studies, assessing the statistical effects of polymorphisms on the variability of complex phenotypes. This offers opportunities to develop new approaches that take into account the specific genetic determinism of honey bees and to open avenues for genomic selection on traits such as varroa resistance. However, tools to perform genetic association studies are so far not tailored to encompass such genetic characteristics, which limits the power of genomic studies performed on honey bees and their transferability into breeding tools. Some markers of interest have been identified (see [15] for a review) but most genomic studies performed on honey bee traits so far were built on a limited number of samples (10 to 200 hundred individuals or colonies [31, 32, 33, 34, 35, 36, 37, 38]), restricting the power of the analyses to come out as a general breeding tool. To date, the use of such markers has been limited, mostly due to a lack of easily accessible genotyping tools, leaving beekeepers with very limited access to varroa resistant stock.

In this study, we took advantage of the unique situation represented by France. Geographical crossroad, the French territory has the advantage to present a large variety of landscapes, environments and ecosystems, where honey bee populations with different genetic background coexist, together with a large variety of hybrid colonies [39]. To overcome the limitations in sample size, we performed one of the largest genomic study applied to honey bees, with the phenotyping and complete genome sequencing of more than 1,500 colonies. Using uniquely tailored genetic and genomic tools, such as queen genotype reconstruction from pool sequence data [30] and GWAS and meta-GWAS analysis ([40], [41]), we investigated the genetic bases of

three major traits linked to varroa resistance: overall varroa infestation of the colony, mite non reproduction (MNR) and recapping of varroa infested brood cells. We identified multiple genetic markers of interest, spread out on the whole genome and heterogeneous across the different populations. This large-scale effort provides a new understanding of the genetic mechanism underlying honey bee resistance to its main parasite, the *Varroa destructor* mite.

Results

Genetic and phenotypic diversity of honey bee colonies

Using allele frequencies estimated from pool sequence data for the 1513 sampled honey bee colonies, we identified three groups of different genetic ancestries: 703 colonies were identified as having more than 80% *Apis mellifera ligustica & carnica* genetic background, 407 having more than 80% *Apis mellifera mellifera* genetic background and 382 as hybrids (Figure 1). An additional 21 colonies were found to be of pure *Apis mellifera caucasia* ancestry, but due to this reduced sample size of this category they were not analysed further.

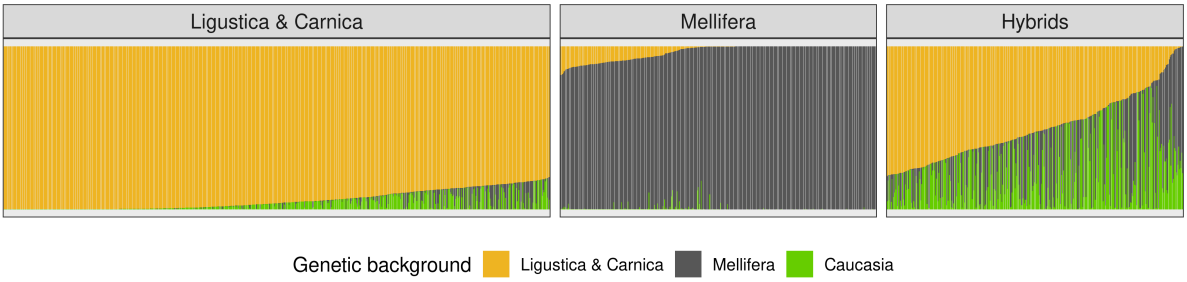

Figure 1: **Genetic background for each colony, per group.**Proportion of the three main genetic background for each of the group analysed in our study *Apis mellifera ligustica & carnica*, *Apis mellifera mellifera* and hybrids.

Out of the six phenotypes initially available four were associated with varroa infestation (*v_pho*, *v_brood*, *v_load* and *v_mito*). They were highly positively correlated with each other and drove the first dimension of the principal component analysis (PCA) across all colonies (Figure 2), and also within each group (supplementary figure SM1-1), with about 60% of the variance explained in each case. Therefore, coordinates of colonies on the first component of the PCA were used as the varroa infestation phenotype in the genome wide association study (GWAS) (*varroa_inf*). Mite non reproduction (*MNR*) was slightly positively correlated with the recapping of infested brood cells, as expected under the assumption that *recap* contributes to overall *MNR*. Both *MNR* and *recap* were slightly negatively correlated with varroa infestation. This observation is consistent with mechanisms being linked to a reduction in varroa infestation within the colony. *MNR* and *recap* contributed both to the second dimension of the PCA, explaining about 20% of the variance. They are separated on the third axis of the PCA, which

explains about 12% of the variance (Figure 2). GWAS were perfomed on each of these three phenotypes (*varroa_inf*, *MNR* and *recap*).

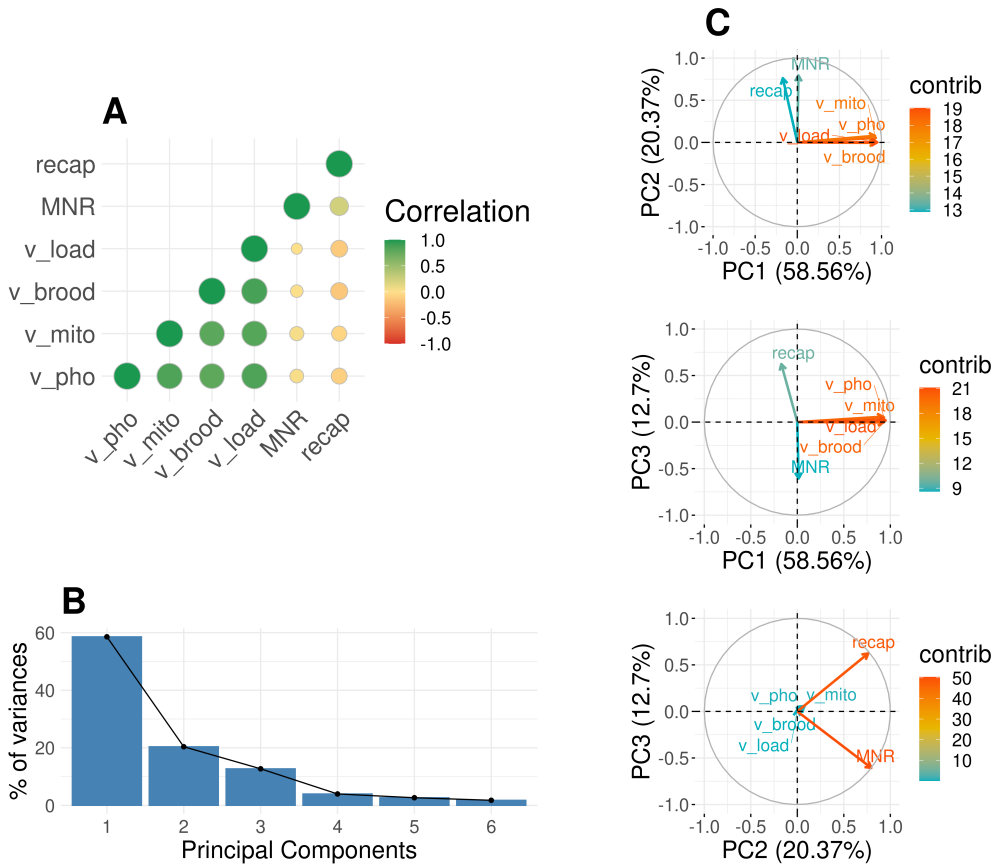

Figure 2: **Correlation and principal component analysis.** Description of the correlation between phenotypes and principal component analysis. (A) gives the correlation between our original phenotypes. (B) summarises the percentage of variance explained by each of the principal component analysis from axis 1 to 5. (C) shows our phenotype on principal component analysis for axis 1, 2 and 3, the colour gives the contribution of each variable to the axis, the closer to red, the higher. The correlation and PCA estimates are based on an analysis of the whole dataset, including all colonies from the three groups.

Meta-analysis of varroa resistance

We aim to identify markers significantly associated with our traits of interest across the three main genetic backgrounds found in Europe. Within population we used standard linear mixed model equations, implemented in GEMMA [42], significant markers were based on their local false discovery and false sign rates estimates from adaptive shrinkage [43]. Population level analyses were combined into a meta-analysis, using ctwo Bayesian methods: a co-ancestry specific program MANTRA [40] and the more generalist Mash [41], to increase power to detect significant markers across the three genetic backgrounds.

Associated variants

Individual GWAS for each group and each phenotype allowed to identify 8 regions (n SNPs=9). In details:
(i) For *varroa_inf*, we found one variant (1:10080627:C>T) in *A. m. ligustica & carnica*, with a positive alternative allele effect, and one (4:11665460:G>A) in the hybrid group, with a negative alternative allele effect (a detailed list of the single nucleotide polymorphisms, SNPs, can be found in supplementary table ST-R1).
(ii) For *recap* we identified four variants significant in *A. m. mellifera*, two of which the alternative alleles had negative effects on the trait (2:2729874:T>C and 3:1059430:T>C) and two having positive effects (4:4327611:G>A, 15:8485332:G>A). In *A. m. ligustica & carnica* a region, with two SNPS, both of which alternative alleles had positive effects (2:12025610:A>G, 2:12025647:A>G) were found. Finally in the hybrid group, one variant for which the alternative allele had a negative effect on the trait (13:9483955:C>T) was found.
Out of the 9 SNPs found significant in one trait for one group, eight fell inside genes of the honey bee annotation [44] and one fell 7kb upstream of its closest gene (for more details see supplementary table ST-R1).

The meta-analysis permitted to identify 51 regions, containing 56 significant SNPs across the three traits of interest. From these 14 regions (n SNPs=14) were significant for *varroa_inf*, 14 regions (n SNPs=15) for *MNR* and 23 regions (n SNPs=27) for *recap*. They distributed across the whole genome, on almost every chromosome (a detailed list of the SNPs can be found in supplementary table ST-R2).
In details:
(i) For the *varroa_inf* trait: the 14 SNPs having a significant effect on the trait were distributed on the chromosomes 7 (n=4), 1, 5 and 8 (n=2) and 4, 6, 11 and 12 (n=1). Thirteen of these SNPs were considered significant in the MANTRA meta-analysis with log10(BF) ranging from 7.59 to 5.09 and one was significant in the mash analysis with log10(BF)=1.16. For six of these variants, the alternative allele had a positive effect on the trait, in at least one of the group (5:9190579:A>G, 7:5772089:A>T, 7:6738985:T>A, 7:11806658:G>A, 8:9799408:C>T and 12:10734707:A>G), for seven alternative allele had a negative effect on the trait, in at least one of the group (1:20960056:C>T, 1:25184394:C>T, 4:11665460:G>A, 5:75369:C>T, 6:10450971:C>T, 7:5762037:T>C and 8:2468335:C>T) and for one the effect depended on the group (11:9369229:T>C). Eight SNPs fell inside genes found in the honey bee annotation, six of them are intronic variation, one leads to a change in 5' UTR and one in 3' UTR. Three fell downstream, within 54kb of their closest genes. The remaining three fell up upstream, within 11kb of their closest genes.
(ii) For *MNR*, the 15 SNPs having a significant effect on the trait were distributed on the chromosomes 1 (n=4), 8, 10 and 12 (n=2) and 2, 3, 5, 11 and 15 (n=1). All these SNPs were considered

significant in the MANTRA meta-analysis with log10(BF) ranging from 5.12 to 6.44. In three of these variants, the alternative allele had a positive effect on the trait, in at least one of the group (2:4437645:G>A, 12:10153855:A>G and 15:4853529:C>T), for seven it had a negative effect on the trait, in at least one of the group (1:16327085:C>T, 1:21374478:G>A, 1:24201224:C>T, 3:6206342:C>T, 8:1150346:C>T, 11:9527267:G>A and 12:136634:G>C) and for five the effect depended on the group (1:2891204:G>A, 5:2008472:A>C, 8:9557205:C>T, 10:5359169:T>A and 10:5359173:C>T). Out of the 15 significant SNPs, 12 fell inside genes, 10 are intronic variation, one into 3' UTR and one causes a missense variation. The three remaining SNPs fell within 17kb downstream of their closest genes.

(iii) Finally, for the *recap* trait SNPs were distributed on the chromosomes 2 and 7 (n=4), 1, 5 and 14 (n=3), 4 and 15 (n=2) and 3, 8, 9, 10, 11 and 16 (n=1). From these SNPs 26 were considered significant based on their log10(BF) values for the MANTRA meta-analysis with log10(BF) ranging from 5.01 to 7.61 and 6 SNPs add log10(BF) values from mash higher than one, their log10(BF) ranged between 1.41 and 2.53. Differently from the other trait five SNPs were found significant in both MANTRA and mash meta-analysis for *recap*. Their log10(BF_MANTRA) ranged between 5.03 and 7.61 and their log10(BF_mash) between 1.41 and 2.53. They were located on the chromosomes 2 (n=2), 3, 11 and 15. For eight of these variants, the alternative allele had a positive effect on the trait, in at least one of the group (1:7448807:A>T, 1:7448811:T>C, 4:7321246:T>A, 4:7321247:G>T, 14:6686131:A>G, 14:8481541:A>G, 15:2081876:A>G and 15:8485332:G>A), for 18 the alternative allele had a negative effect on the trait, in at least one of the group (1:15280956:G>A, 2:2729874:T>C, 2:8350714:G>A, 2:12025610:A>G, 2:16060868:G>A, 3:1059430:T>C, 5:6736534:T>C, 5:6761414:T>A, 5:8737386:G>A, 7:7028040:G>A, 7:7051965:A>G, 7:7078376:C>T, 8:1551638:C>T, 9:11564671:A>C, 10:2026877:C>G, 11:14369154:G>C, 14:3782741:G>A and 16:1812909:C>T) and for one the effect depended on the group (7:8466948:A>G). For this trait, 18 variants fell inside genes with 16 intronic variation, one in a region coding for long non coding RNA and one in 5' UTR, being also a missense variant. Three SNPs fell between 52 and 102kb downstream of their closest genes. Six SNPs fell between 5 and 129kb upstream of their closest genes (details are available in supplementary tables ST-R2 and ST-R3).

We identified chromosome regions smaller than 1Mb, sharing significant SNPs between multiple traits. Between *varroa_inf* and *MNR* we identified two regions on chromosome 1 (20.9-21.4Mb and 24.2-25.2Mb), a region on chromosome 8 (9.5-9.8Mb), a region on chromosome 11 (9.3-9.6Mb) and a region on chromosome 12 (10.1-10.8Mb) having significant SNPs for both traits. Between *varroa_inf* and *recap* we identified a region on chromosome 5 (8.7-9.2Mb), a region on chromosome 7 (6.7-7.1Mb) and a region on chromosome 8 (1.5-2.5Mb). Finally between *MNR* and *recap* we identified a region on chromosome 8 (1.1-1.6Mb) (Figure 3, for the details of each region see supplementary table ST-R4).

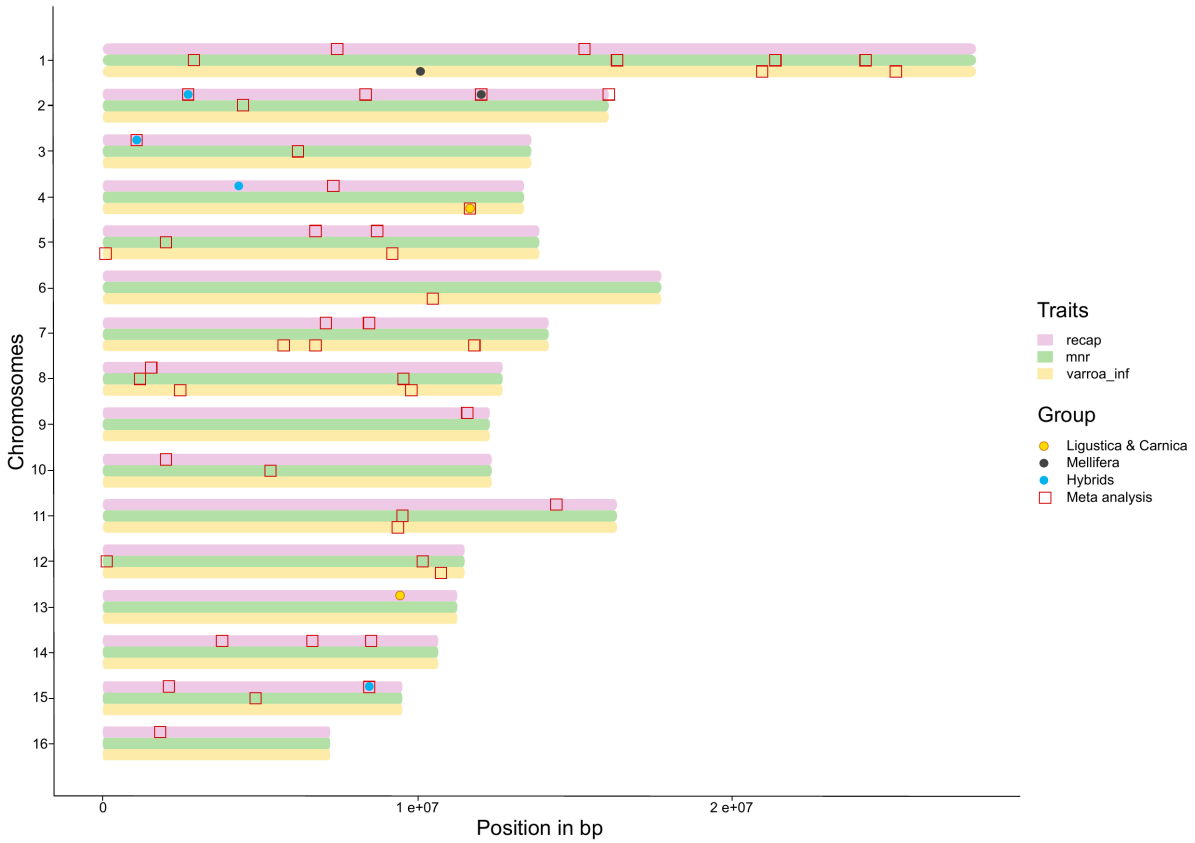

Figure 3: **Overlap between traits across significant regions.** Position of each of the significant SNPs on their chromosomes. A coloured dot means that this position has been identified in either *A. m. ligustica & carnica*, *A. m. mellifera* or hybrids. A red square means that it has been identified in the meta analysis. The colour bars represent the phenotypes of interest *varroa_inf*, *MNR* and *recap*. This figure allows to see overlapping window containing significant marker across the phenotypes.

Heterogeneity of effects

As a reminder, three genetic groups were analysed in this study, two groups having 'pure' genetic background corresponding to two of the three main honey bee subspecies in Europe and a hybrid group. The Fst, measure of population differentiation, between groups were 0.26, between *A. m. mellifera* and *A. m. ligustica & carnica*, 0.20 between *A. m. mellifera* and hybrids and 0.08 between *A. m. ligustica & carnica* and hybrids.

As detailed earlier, from the meta analysis we observed 17 variants (n *varroa_inf*=6, n *MNR*=3 and n *recap*=8) for which alternative allele had significant positive effects for at least one of groups, 32 (n *varroa_inf*=7, n *MNR*=7 and n *recap*=18) having significant negative effects for at least one of the group, and seven (n *varroa_inf*=1, n *MNR*=5 and n *recap*=1) having divergent effect depending on the group (supplementary table ST-R3). Correlations between the effects of significant SNPs between different groups were null therefore showing great heterogeneity in SNPs having an effect, their impact on the trait of interest and their magnitude across the different genetic groups, especially for the *MNR* trait.

About 80% of the SNPs identified significant fell into intronic regions, there was no differ-

ences between the annotation of the significant SNPs on the genome and the annotation of all tested SNPs. Five SNPs were found significant in both individual GWAS and meta-analysis, one for *varroa_inf* identified significant for the hybrid group, and four for *recap*, three for *A. m. mellifera* and one for *A. m. ligustica & carnica*. They were located on chromosome 2 (n=2), 3 (n=1), 4 (n=1) and 15 (n=1). They all located inside the genes, LOC102655235, LOC410853 (chromosome 2), LOC409402 (chromosome 3), LOC408787 (chromosome 4) and LOC726948 (chromosome 15).

Example of associations

As an example of a region associated with two traits, we focus on the region between 9.5 and 9.8Mb on the chromosome 8 where two SNPs appear significant in the meta analysis, one for *MNR* at 9,557,205 bp and one for *varroa_inf* at 9,799,408 bp.

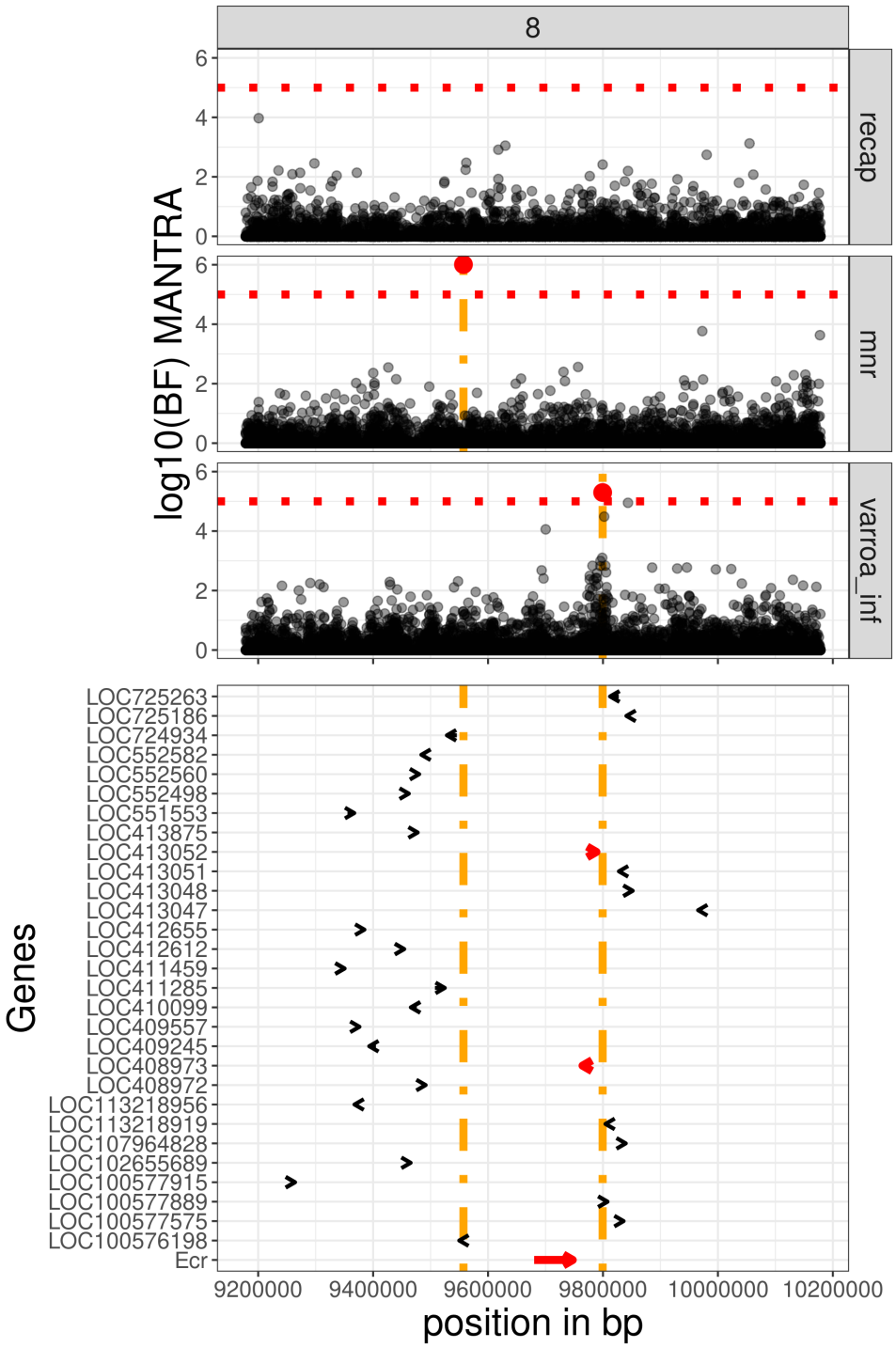

Figure 4: **Region on chromosome 8.** From top to bottom we represented results from genome wide association studies for chromosome 8, region going from 9.2Mb to 10.2Mb, for *varroa_inf*, *MNR* and *recap*. We represented log10 bayes factor for MANTRA, with the significant threshold as a dotted red line, under which we listed the genes identified within this region of chromosome 8. The orange lines represent the positions of the significant markers, the genes falling in the region between these markers are highlighted in red.

In this region we identified a couple of SNPs in close vicinity to the gene Ecr, ecdysone receptor (Figure 4). In addition, we also identified a SNP, located at 9,696,277bp so within Ecr, in high linkage disequilibrium (LD) with the significant SNP 8:2468335:C>T also found in the

meta-analysis for *varroa_inf* (supplementary table ST-R6).

Next, we turn to chromosomes 15, where a significant SNP for *recap* was found at position 2,081,876. Nearby we also observed two suggestive SNPs, close to the significant threshold set for this analysis, in positions 2,021,142 and 2,081,914 bp. The markers 15:2081876:A>G and 15:2081914:A>G were in full linkage disequilibrium in *A. m. ligustica & carnica* and in the hybrid group and in high linkage disequilibrium ($r^2 > 0.8$) in *A. m. mellifera* and they fell within the same haplotype block of 1.6kbp (as identified by Wragg et al. (2022) [39], supplementary table ST-R7). This haplotype block did not seem to contain annotated genes.

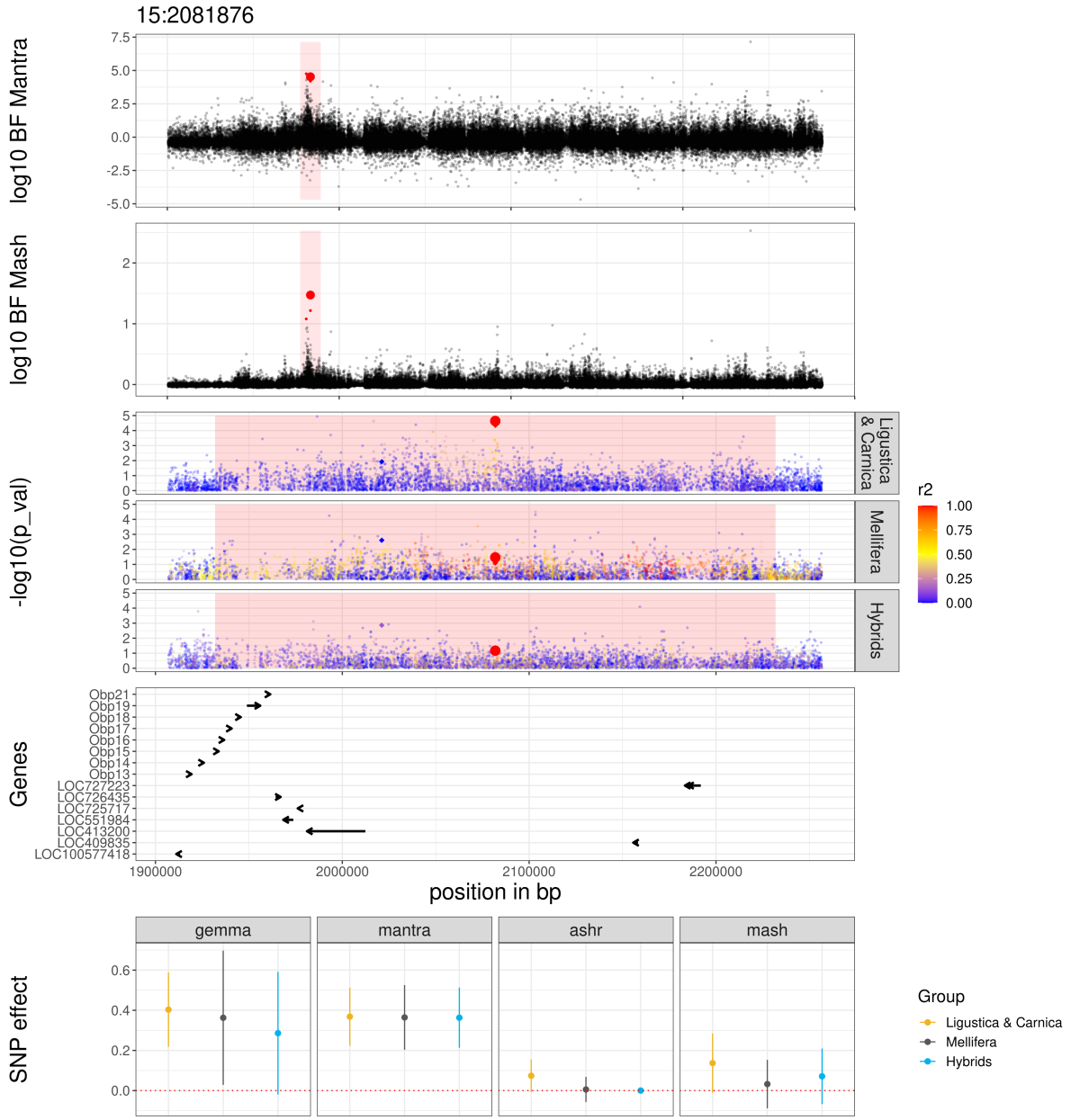

Figure 5: **Region on chromosome 15.** We plotted manhattan plots for the marker positioned at 2,081,876 bp on chromosome 15 and the surrounding 0.3Mbp region. The marker 2,081,876 is represented by a large red dot when the close by markers identified as near significant thresholds are represented as smaller. From top to bottom we plotted log10(bayes factor) for MANTRA, then for mash, then -log10(p-values), estimated using GEMMA, for the groups *A. m. ligustica & carnica*, *A. m. mellifera* and hybrids. Each marker is coloured based on its linkage disequilibrium r^2 with the marker of interest. The genes falling in the region of interest are plotted and then the effects of the significant marker extracted from genome wide association studies ran with GEMMA, `ash`, MANTRA and mash.

The marker 15:2021142:C>T fell in a short haplotype block (0.175kbp) containing the gene LOC413200 which has been identified as putative immune related gene by Ryabov et al. (2014) [45]. Interestingly they were located within less than 1Mbp downstream from a group of eight genes coding for odorant binding proteins. For the first SNP (15:2021142:C>T), alternative allele had a negative effect for all groups whereas the second (15:2081876:A>G),

mostly significant in *A. m. ligustica & carnica*, the alternative allele had a positive effect for all groups (Figure 5). Marker effects estimated with GEMMA lack precision, whereas the `ash` and mash methods apply a strong shrinkage to the estimates. Going from individual analysis to meta-GWAS improved the power to detect associations and improve our ability to estimate markers effects.

Polygenic architecture of varroa resistance

The program GEMMA [42], used to perform individual GWAS, provides estimates for the proportion of phenotypic variance explained by the SNPs (*pve*) and proportion of genetic variance explained by the sparse effects (*pge*) for each trait in each group (Table 1 and Figure 6). PVEs ranged between 0.20 and 0.82 (se=[0.09 ; 0.19]). PVEs, estimated using the Bayesian Sparse linear mixed model (`bslmm`), were close to the `lmm` estimates, and ranged between 0.14 and 0.82 (se=[0.07; 0.15]). The 95% confidence and credible intervals for PVEs from both `lmm` and `bslmm` appeared exclusively positive, except for *MNR* in the hybrids group. However, PGEs were much lower, ranging between 0.06 to 0.33 (se=[0.08; 0.25]). Their 95% confidence/credible intervals often included zero. The only traits and groups having PGEs different from zero were *MNR* and *recap* for the group of colonies of *A. m. ligustica* type. The estimates for the GWAS on hybrids and *A. m. mellifera* always showed larger standard error, as expected due to the smaller sample sizes of these groups. Interestingly it appears that PVE estimate is slightly higher for *A. m. mellifera* for the *MNR* phenotype compared to the two other groups, whereas they seem similar between the three groups for the two other phenotypes, *varroa_inf* and *recap* (a complete summary can be found in supplementary table ST-R8).

Table 1: **Summary of PVE (Phenotypic Variance Explained) and PGE (Proportion of Genetic variance Explained).**

		Varroa infestation	**Resistance to varroa infestation**	
		Varroa_inf	***MNR***	***Recap***
PVE(se) from `lmm`	***Ligustica & Carnica***	0.705 (0.091)	0.396 (0.113)	0.632 (0.100)
	Mellifera	0.731 (0.160)	0.625 (0.191)	0.534 (0.165)
	Hybrids	0.816 (0.124)	0.197 (0.178)	0.570 (0.160)
PVE[95% ci] from `bslmm`	***Ligustica & Carnica***	0.820 [0.683; 0.953]	0.475 [0.314; 0.640]	0.759 [0.602; 0.910]
	Mellifera	0.719 [0.410; 0.982]	0.718 [0.433; 0.981]	0.663 [0.420; 0.893]
	Hybrids	0.821 [0.570; 0.998]	0.145 [0.006; 0.393]	0.580 [0.285; 0.847]
PGE[95% ci] from `bslmm`	***Ligustica & Carnica***	0.277 [0.024; 0.555]	0.143 [0.001; 0.503]	0.238 [0.008; 0.575]
	Mellifera	0.065 [0; 0.286]	0.233 [0.005; 0.666]	0.331 [0.010; 0.751]
	Hybrids	0.128 [0; 0.569]	0.238 [0; 0.880]	0.115 [0.001; 0.425]

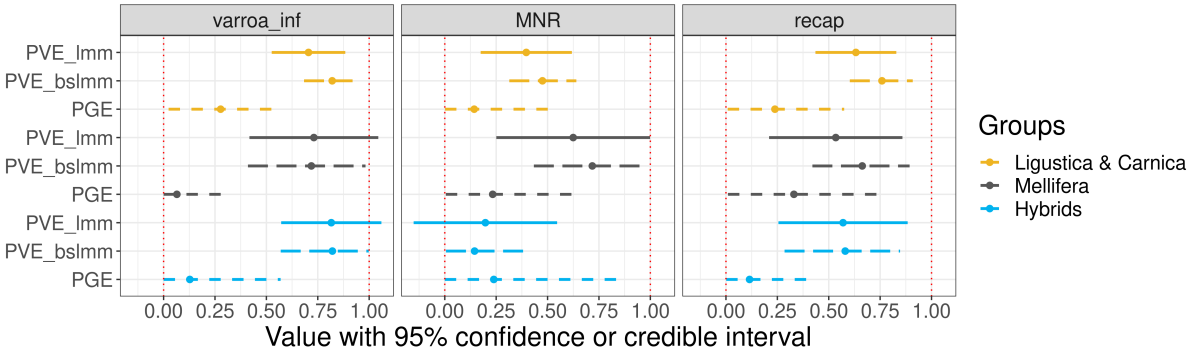

Figure 6: **Genome wide association estimates.** Confidence and credible intervals for percentage of variance explained by linear mixed model (full line), Bayesian sparse linear mixed model (long dotted line) and percentage of genetic variance explained by Bayesian sparse linear mixed model (short dotted line) for the three group analysed, in yellow *Apis mellifera ligustica & carnica*, in dark grey for *Apis mellifera mellifera* and in blue for hybrid.

Correlation between SNP effects, across the whole genome, estimated using GEMMA and `ash`, for individual GWAS analysis, on the one hand and MANTRA and mash, for meta-GWAS, on the other hand were positive. However, as expected, between the different groups there was no correlation when estimated from individual GWAS whereas there was some positive correlation when estimated from meta-GWAS. Across the different phenotypes, *recap* and *MNR* appear to be slightly positively genetically correlated, as one could expect knowing that recapping behaviour, performed by adult bees, is potentially a component of the *MNR* phenotype. The phenotypes *varroa_inf* and *MNR* were not genetically correlated. Finally, *varroa_inf* and *recap* were mostly negatively correlated for *A. m. mellifera* and *A. m. ligustica & carnica* whereas there was no to very little positive genetic correlation between these two phenotypes in the hybrids group (Table 2, the detailed correlations can be found in Supplementary table ST-R5).

Table 2: **Genetic correlations.** Range of the estimates from Pearson correlations on the different GWAS analyses methods used in our study.

		Varroa_inf	***MNR***	***Recap***
	Varroa_inf	1		
Ligustica & Carnica	***MNR***	[0.042; 0.049]	1	
Mellifera		[-0.023; 0.005]		
Hybrids		[0.050; 0.073]		
Ligustica & Carnica	***Recap***	[-0.304; -0.213]	[0.131; 0.176]	1
Mellifera		[-0.132; -0.058]	[0.088; 0.197]	
Hybrids		[0.021; 0.173]	[0.021; 0.119]	

Discussion

In this study we performed the largest genome wide association study on the resistance of honey bees to their current biggest biotic threat, the parasite *Varroa destructor*. We combined an extensive genotyping and phenotyping effort with meta-analyses methods to identify genetic markers and associated genes harbouring a significant effect on varroa resistance. This leveraged both multiple traits associated to varroa resistance and the complex genetic structure found in French honey bee colonies. Our results show that we were able to pinpoint significant effects of some regions of the honey bee genome, within specific genetic types and across the whole meta-population, offering insights into the biological mechanisms involved in varroa resistance. These genomic regions only explain a small portion of the genetic variation which remains mostly polygenic. However we found the contribution of genetics to varroa resistance to be substantial, offering positive perspectives to a possible adaptation and selection of honey bee populations to this relatively recent threat.

With about 1,500 sampled honey bee colonies, sequenced in pool and genotyped for around 3 million SNPS, this study is the most global association study for honey bee varroa resistance traits to date. In fact, this study benefits from the largest sample size (phenotyped and sequenced) known thus far to perform genome wide association study in honey bees. In addition, the set of colonies studied is highly representative of many honey bee populations worldwide, measured in terms of the genetic backgrounds described in the diversity panel of Wragg et al. (2022) [39]. This study stands out from the previous honey bee quantitative genetic studies that mostly focused on specific genetic backgrounds, often in small experiments, not representative of the field situation and for really specific phenotypes ([37] for a review). In addition to the experimental effort to gather the raw data, this analysis benefited from dedicated statistical methods for reconstructing the honey bee queen genotype from a pool sequencing experiment [30] and meta-analyses approaches to increase statistical power to detect significant associations [40, 41].

Phenotyping varroa resistance

Resistance to varroa is a complex trait, involving many different aspects of the biology of honey bee colonies. A recent extensive review of varroa resistance traits [25] reveals that there is no clear evidence for significant correlations between the standard traits measured as proxy for varroa resistance, such as varroa sensitive hygiene (VSH), grooming, mite non reproduction (MNR), hygienic behaviour or uncapping-recapping as cleaning behaviour. Depending on the population on which the trait has been measured, in particular their evolutionary history (*e.g.* natural or artifical selection) the correlations between these traits range from 'apparent link' to 'no link'.

Here, we observe a slightly positive correlations between traits linked to varroa resistance,

MNR and *recap* overall or at the population level, and a small negative correlation between these two traits and *varroa_inf* (Figure 2 and in supplementary figure SM1-1). This finding fits with the hypothesis that there is a panel of mechanisms allowing honey bee to resist varroa and that these mechanisms do not seem to be completely shared across populations or genetic ancestries.

An important experimental aspect that limit varroa resistance studies, including this one, is that most currently applied measures of varroa resistance are difficult to scale to large samples. For example, they can imply tedious and potentially subjective scoring, induced or artificial varroa infestation, multiple measures in time, estimation of ratios and applying heuristic thresholds for minimum detection. However, direct estimates of varroa infestation are the most simple traits to measure to summarise varroa resistance. Indeed, a low varroa infestation can be explained by multiple phenomena either linked to the environment, beekeeping practices, varroa biology or an action from the honey bee colony itself. In this study, besides classical measures of varroa infestation, we proposed to measure it using the ratio of reads mapped to the varroa mitochondrial DNA over those mapped to the honey bee genome (*varroa_mito*). We believe this new measure offers specific advantages for the study of varroa resistance: (i) there is a high correlation between this estimate and the phoretic varroa infestation (supplementary methods 1), a trait that is more complex to measure, (ii) both colony sequence and varroa infestation information come from a unique biological sample, (iii) there is no potential bias due to the collector, (iv) no specific technical skills are needed on the field and (v) it is comparable across studies. Using it as a standard phenotype for varroa infestation would facilitate the establishment of a large collection of phenotypic records, necessary to be able to build up information across studies, follow phenotypic progress in a surveyed population and potentially breed honey bee populations for varroa resistance.

Insights into biological mechanisms underlying varroa resistance

Varroa resistance mechanisms can be partitioned into two types of traits: first, traits related to hygiene (including VSH, recapping and MNR, but also more broadly grooming behaviour) that involve the accurate detection by workers of varroa infested cells and second, their subsequent inspection/destruction. It has been shown that VSH bees target more specifically cells with highly compromised brood, which is related to the level of infestation in the cells ([46], [14]). As a result, cells with less mites or mites that are not effectively reproducing are more likely to stay intact, thus increasing the level of mite non reproduction in the colony (MNR). The second type of trait is a trait expressed by either the workers or the brood, that would disrupt mite reproduction within capped cells (and thus increase MNR). Both trait types can reduce mite infestation in the colony, thus increasing varroa resistance of honey bee colonies. Interestingly, in this study we found markers associated with genes that relate to these two categories.

Impairment of mite reproduction First of all, we took advantage of the recent review by Mondet et al. (2020) [15] that summarises past results from association study on varroa resistance. We performed a liftover to obtain markers and regions coordinates over the latest genome assembly (HAv3.1, [44]) (supplementary table ST-M3). Two of our significant SNPs, for *recap* fell into a region around 7Mb on chromosome 1 described, in an association study, by Tsuruda et al. (2012) [47] as suggestive quantitative trait loci (QTL) for VSH. Two other significant markers 1:2891204:G>A and 1:21374478:G>A fell inside genes LOC410758 and LOC413968 respectively. These genes have been identified in a study by Saelao et al. (2020) [48] looking for selection signal in hygienic honey bee populations of the US. In addition, the marker 11:14369154:G>C, significant for *recap*, is located close to the TpnCi gene, coding for troponin C type. This gene has been shown to be over-expressed in non-hygienic Africanized bee lines, when compared with hygienic lines, by Teixeira et al. (2021) [49]. Finally markers 4:10789077:T>C and 8:1551638:C>T fell into genes found by Ament et al. (2011) [50] as linked to protein abundance in fat body and haemolymph of the adult honey bee. We know that varroa, while infesting the colony, survives by feeding on these adult bee biological fluids. These markers might sign an impact of varroa infestation, or the resilience to such infestation, on biological pathways linked to composition of the honey bee heamolymph and fat body.

In the region on chromosome 8, we identified multiple significant markers close to the gene Ecr, ecdysone receptor. This gene has been identified as crucial for the reproduction of *Varroa destructor*, within the vitellogenin production pathway, and only produced by honey bees. Hence it stands a key in the interaction between honey bees and varroa ([51], [34], [52]).

Detection of varroa infested cells If we look more into the general biology of the honey bee, we saw that two markers, significant for *MNR*, on chromosome 10, fell into the 5-HT2beta gene. This gene is known to be a serotonin receptor, thus involved in olfactory pathways in a large number of insects ([53], [54]). The biological hypothesis for resistance to varroa infestation in honey bee is that some chemical mechanisms might be involved and will give cues for the adult bees to perform behaviours leading to better colony resistance to the mite. Olfactory biological pathways thus appear crucial to such response to the infestation. We also noticed that the significant markers 3:12973246:A>G and 12973248:A>G fell into LOC413503 (alias GB41230) also found by Mondet et al. (2015) [46], as being differentially expressed in antennae of honey bees expressing VSH behaviour.

In the region on chromosome 15, we identified multiple significant markers located within less than 1 Mb downstream from a group of genes coding for odorant binding proteins. These genes are found in two major clusters on the honey bee genome, seven on chromosome 9, nine on chromosome 15 (monophyletic group called C-minus subfamily); and two on chromosome 10 and one respectively on chromosome 2 and 12, summing to 21 genes ([55], [56]). The cluster located on chromosome 15 contains OBPs 13 to 21, 6 of which have already been mentioned in genomic studies looking at varroa resistance related traits (obp 14, 15, 16, 17, 18 and 21; [15],

[57]). In particular, Obp14 has been identified as up-regulated in two studies looking at gene and protein expression in VSH bees ([46], [58], [59]). Obp18 has been identified by two proteomic studies looking at the VSH and hygienic behaviour traits ([60], [59], [58], [61]). Even though these genes were not declared significant in the dedicated analysis we can still hypothesise that they might be involved in some resistance mechanism targeting varroa infestation as they play a major role in sensory pathways. These genes might be relevant for marker assisted selection, as suggested by Guarna et al. (2017) [62] in the case of a tool dedicated to Canadian honey bees selection for hygienic behaviour.

Genetic architecture of varroa resistance

The review from Guichard et al. (2020) [25] reported heritabilities ranging from close to 0 to up to 0.85, with large standard errors, for traits associated with resistance to varroa. More recently Gabel et al. (2023) [63] estimated the heritability of *MNR* to be close to 0.4, which is in the same range as our estimates. Most 95% confidence/credible intervals for these estimates included zero and the ones who did not were mostly modest estimates (< 0.2). In addition, repeatability estimates of these traits are low, for example they range between 0.2 (long term) and 0.4 (short term) [24] or virtually zero [20] for *MNR*. More importantly, estimates based on different populations, e.g. *A. m. mellifera* or *A. m. ligustica & carnica* show discrepancy. We expect such variability to be due to different genetic mechanisms involved in these traits.

The heritability estimates from this study, for honey bee resistance traits, seem high compared to standard traits measured on livestock species. In fact it is somehow possible that we fail to fully disentangle genetic from environmental stratification. When intending to estimate heritabilities in species like honey bee one faces the challenge to integrate the potentially large impact of environmental variation. Part of the population structure is likely associated with such variation, in addition to genetic variation. In this study, we aim to correct for such structure in our sample by thoroughly accounting for covariates, axis of the principal component of the genomic relationship matrix. In addition, when computing correlations (using Mantel test) between genetic, geographic and environmental distances between colonies we see that environmental variables are slightly correlated with population structure (positively for *A. m. mellifera* and negatively for *A. m. ligustica & carnica*), whereas their correlation with our phenotypes of interest is not significantly different from zero (supplementary table ST-R9). Environmental variables seemed to have some impacts that we do not take into account in our analysis (supplementary table ST-R10), and that might have affected slightly our heritability estimates (potentially deflating). However, most of these variables are highly confounded and nested in our experimental design (same observer, on the same day, for the same apiary of a specific genetic background) and if corrected for, prior to GWAS estimation, might lead to over-correction and thus drastically decrease our detection power in the association studies.

For most of our traits of interest we obtain a large, positive, phenotypic variance explained

(PVE) by the markers. Such phenotypic variance is likely overestimated due to limitations in the way we accounted for environmental fixed effects. However the proportion of genetic variance explained by large effects (PGE), estimated using the `bslmm` model in GEMMA [42] was generally low and included 0, except for *varroa_inf* in *A. m. ligustica & carnica*. These estimates support our hypothesis that such traits are highly polygenic and not solely driven by a few markers with large effects. In addition we observed a small difference between the explained variance from the two models (`lmm` and `bslmm` from GEMMA) due to not correcting for fine population structure, principal components from the relationship matrix, in `bslmm`. We can see that taking into account fine population structure affects slightly the estimated heritabilities.

The traits linked to varroa infestation and resistance seem to have a small yet significant part of genetic heritability, and thus can be passed on from one generation to next through selection. However, and even though we identified genetic markers with significant effects, it is unlikely that causal mutation, explaining the largest part of the phenotypic variance can be identified and then selected for.

The future of genome-wide association studies in honey bees

The honey bee genome harbours some specificities. First, it is known to experience a large number of recombination events, with an average recombination rate of 37cM/Mbp [44]. Second, as a result of the peculiar natural mating observe in honey bee the effective population size, the number of actively reproducing animal in a population, of each population or colony is expected to be large as colonies come from the mating of a queen with multiple males, avoiding the over-representation of a specific animal. Such particularities cause low linkage disequilibrium, as seen in the supplementary methods 2, between genomic markers, making it harder to identify candidate loci (QTLs) linked to specific traits and to further select for them.

In addition, the honey bee population exhibits a complex genetic diversity. In this study we provide a better understanding of the genetic background behind varroa infestation and resistance in honey bees in general and the French population in particular. Many honey bee colonies are hybrids of various proportions from the three main *Apis mellifera* subspecies found in Europe, *ligustica & carnica*, *mellifera* and *caucasia*. In this study we took advantage of this admixed population to identify genetic markers linked with our traits of interest within genetic types, in hybrids and across these populations, making it possible to see differences in significance and effect depending on the genetic type. Knowledge of linkage disequilibrium (LD) and associated estimation of local haplotypes for regions of interest combined with knowledge on SNP effects for each sub-species can increase our prediction accuracy for different traits. A better understanding of the local genetic background of the hybrid population could help predict effects of specific SNPs. Studies focusing on hybrid colonies, describing their genetic background throughout the genome and comparing different genetic make up could be highly valuable to identify relevant genetic patterns, especially in the context of genomic selection. As

an example, multiple genome regions were flagged with more than one significant marker for the trait *recap* but literature linking these regions to honey bee biology is lacking. In addition, we identified some markers with opposing effects across the different genetic backgrounds, especially *A. m. ligustica & carnica* and *A. m. mellifera*. These SNPs can be of interest to explain differences between subspecies as these seem to be signs of local phase rearrangement. It appears relevant to short-list these regions as potential regions of interest for future studies geared towards a better honey bee genome annotation and understanding of underlying biological pathways.

Genomic selection One candid perspective of our work would be to integrate the identified variants into genomic selection program aiming to breed for resistant honey bee colonies. Genomic selection is commonly used in mainstream livestock species. In some countries, such as Germany, beekeepers are grouped in breeding organisations, they highly use artificial insemination to track their mating making the construction of breeding schemes easier [64]. In this context, some studies ([65], [66], [67], [68], [69]) described the statistical models and sampling strategies that can be successfully applied to implement genomic selection in honey bees. In contrast in other countries, as in France, there is limited consensus between beekeepers on breeding practices, hence making selection schemes for beekeeping difficult to describe and manage. Moreover, one limitation to the widespread use of these methods is that so far genomic selection has been proven successful when the focus is on a unique honey bee subspecies, e.g *A. m. carnica*, and concentrate their efforts on phenotypes linked to production. In the context of the French honey bee population, as described in this study, the hybrid stock as well as the complex phenotypes of interest, make it less straightforward to directly apply genomic selection. Highly polygenic traits come with challenges when considering selection and we expect that the markers identified in this study will not be sufficient to contribute alone to the establishment of a genomic selection scheme even though the estimated non null phenotypic variances still supports the possibility of selection, as it has been pursued in the US [17]. For future studies, a primary focus should be put on increasing the sample size, in terms of number of phenotyped and genotyped colonies, to boost the precision, detection power and replication capacity of association studies. A way to improve the robustness of our marker contributions to selection decisions would be, for example, to deepen our knowledge on pedigree, as done in Bernstein et al. (2023) [69] as well as have access to a large number of individual queen genotypes, rather than reconstructed queen genotypes, as explained in Bernstein et al. (2021) [70]. In addition, there is a need for standardised biological samples in terms of genotype and phenotype. The genotypes could either come from large SNP panels, covering the whole genome, such as proposed by Jones et al. (2020) [71] and used by Bernstein et al. (2023) [69] or come from whole genome sequencing experiments characterised using genetic diversity panel [39]. The phenotypes could be more 'automated' and alleviated from some sampling variability, as we propose with *varroa_mito*.

Finally, to fully understand the genetic architecture behind varroa resistance one needs to broaden his horizon and look into additional phenomena, not only associated with genetics. For example, it would be necessary to better describe maternal effects, disentangling queen from drones genetic contribution, to our traits. Another interesting source of variation could be associated with epigenetics, as it is known to drive cast determination and colony composition ([72], [73], [74], [75]).

Materials and Methods

Honey bee colonies and sampling strategy

The sampling strategy was established to represent the diversity (both in terms of genetic background and beekeeping practices) of honey bee colonies maintained by French beekeepers. A total of 97 beekeepers, located in France, Switzerland, Luxembourg, the Netherlands, Sweden and New Zealand participated in this study (Figure 7). Foreign colonies were included as they show similar genetic background to the French honey bee population, mostly because of historical or ongoing trade between beekeepers. A total of 1,513 colonies were sampled to go through sequencing. Out of these 1,513, 1,441 *Apis mellifera* colonies were phenotyped, under the condition that each beekeeper contributed at least 6 colonies (from 6 to 125 with on average 19.25 colonies). These colonies were phenotyped for multiple traits known to be related to varroa resistance (described in details below), once per colony at the end of the beekeeping season (summer and fall in Europe, *i.e.* typically between July and September).

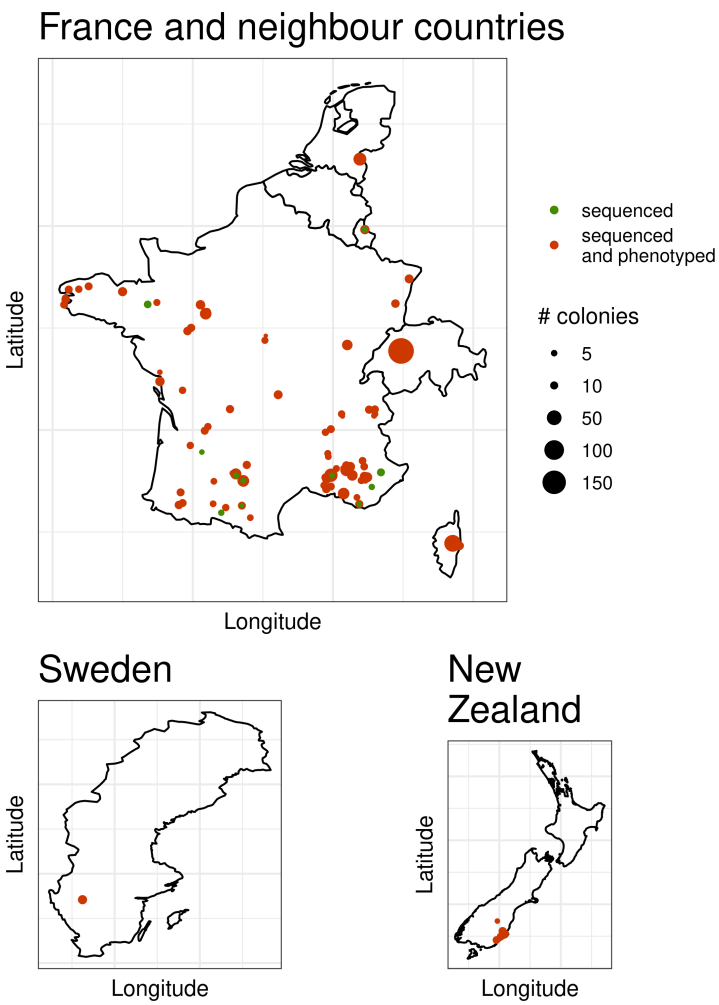

Figure 7: **Geographic distribution of the sampled colonies.** Geographical locations of colonies that were whole genome sequenced (in green) and both phenotyped and sequenced (in red). The size of the dot represents the number of honey bee colonies per location and category.

Genetic characterisation

All colonies were genotyped from whole genome sequencing of pools of workers using the strategy described in Eynard et al. (2022) [30]. This strategy was decomposed in three steps: first allele counts at Single Nucleotide Polymorphisms (SNPs) were obtained from whole genome sequence, then the genetic background of each colony was estimated from a set of well differentiated markers and finally the genotype of the queen was predicted among colonies of similar genetic background. These different steps are detailed below.

Whole Genome Sequencing For each colony approximately five hundred honey bee workers were ground in 100 mL of TNE buffer. Fifteen mL of ground sample was then collected and centrifuged for 15 min at 3400 rcf (relative centrifugal force). A volume of 200 μL of supernatant was lysed overnight at 56°C, with a solution of proteinase K (Eurobio GEXPRK01-B5) and DTT (1,4 Dithiothreitol). Automated DNA extraction was done with a Qiasymphony® -Qiagen.

DNA concentrations for each sample were estimated with Infinit200®-Tecan. Thereafter, pool sequencing was done on a NovaSeq6000 platform in order to obtain about 30X genome-wide raw sequencing coverage. Sequencing reads were aligned on the honey bee reference genome Amel HAv3.1 ([44], Genbank accession GCA_003254395.2), using BWA-MEM [76]. In addition, it was expected that the biological sample contained adult varroa, present in phoretic phase on the bees, within the pool of sequenced honey bee workers. The obtained reads were therefore also aligned on the varroa mitochondria (genome Vdes_3.0, Genbank accession GCA_002443255.1, [77]) using the same procedure.

SNP Genotyping Genotypes were estimated at each of the 7,023,976 SNPs identified in Wragg et al. (2022) [39]. Pool sequences were analysed using Samtools mpileup [78] with the recommended parameters: -C minimum mapping quality for reads with excessive mismatches of 50, -q minimum mapping quality for an alignment of 20, -Q minimum base quality of 20. Then Pileup files were interpreted by the PoPoolation2 utility mpileup2sync [79], with a minimum quality of 20 and were finally converted to allele counts and sequencing depth files, filtering out real tri-allelic and potential sequencing error. This procedure led to a set of 6,831,074 SNPs that were used in all downstream analyses. Colonies were sequenced on average with 27.4X coverage, each SNP was on average sequenced with 29.9X coverage, as planned during experimental design (Supplementary table ST-M1).

Population Structure For each colony, we ran the model presented in Eynard et al. (2022) [30] to estimate the genetic background on 48,589 SNPs selected so that they differentiate the three main genetic background of honey bees in Europe [80, 39, 81]): the **C** lineage comprising the lowly differentiated *Apis mellifera ligustica* and *Apis mellifera carnica*, the **M** lineage of Western Europe *Apis mellifera mellifera* and the **O** lineage of Eastern Europe/ South-Western Asia *Apis mellifera caucasia*. Specifically, the SNPs were chosen based on the following criteria: (i) there is a maximum of two polymorphic sites within a 100 base pair window, (ii) only one representative marker per linkage disequilibrium block with r^2 higher than 0.8, (iii) the variance between allele frequencies in the three main European genetic background is higher than zero, to allow for population identification and (iv) so that minor allele frequencies (MAF) within the selected markers follows a uniform distribution. The list of selected markers is provided in supplementary table ST-M2. For step (ii) above, LD was estimated from the reference diversity panel from Wragg et al. (2022) [39] using the plink software version 1.9[82, 83] with options –r2 –ld-window 100 –ld-window-kb 10 –ld-window-r2 0.8 and a unique marker was selected manually as the median point for each LD block. The Admixture model from Eynard et al. (2022) [30] was used to estimate the genetic background of each colony (*i.e.* the admixture proportions [84] of each of the three genetic backgrounds).

Queen genotype reconstruction Following the procedure described in Eynard et al. (2022) [30] we grouped the colonies based on their genetic background (*A. m. ligustica & carnica*, *A. m. mellifera* and hybrid) to obtain homogeneous populations. Colonies were if they harboured more than 80% of the same genetic background. Colonies that could not be assigned an homogeneous background colonies were assigned to the 'hybrid' group. Due to a lack of pure *A. m. caucasia* colonies this group was not considered further in the study. Once homogeneous groups are defined it is possible to perform honey bee queen genotype inference using the homogeneous model described in Eynard et al. (2022) [30]. In short, the method is based on the likelihood of the queen genotype, written as

$$x_l^c | d_l^c, f_l, g_l^c \sim \text{Binomial}(\frac{f_l + g_l^c}{2}, d_l^c) \tag{1}$$

where g_l^c is the (unknown) queen genotype, f_l is the unknown reference allele frequency in the population, d_l^c and x_l^c the sequencing depth and allele counts obtained from pool sequencing experiments for locus l and colony c. By considering all colonies of the same genetic background jointly, f_l can be estimated by maximum likelihood and the posterior probabilities of the three possible genotypes of the queen computed.

Varroa resistance phenotypes

Varroa Infestation Varroa infestation was quantified with four different measures: phoretic mite infestation (on adult bees using two different methods), brood infestation, and total mite load.

Phoretic mite infestation (*v_pho*) was measured using the detergent method [85]. In brief, a sample of approximately 300 adult honey bees was collected in each colony, on a frame containing uncapped brood. After weighing of this sample, the number of mites falling as a consequence of washing with a detergent solution was counted, and the proportion of mites within the sample expressed as the number of varroa per 100 honey bees (assuming the weight of 1 single bee to be 140 mg). An alternative measure of phoretic varroa (*v_mito*) was obtained from the pool sequencing data by calculating the ratio of the number of reads mapping varroa mitochondria on the number of reads mapping the honey bee genome.

Brood infestation (*v_brood*) was expressed as the proportion of varroa infesting brood cells in the colony. This proportion was estimated among 300 randomly sampled brood cells on a single frame containing capped brood aged 7 to 11 days post-capping (P5 to P8 stages).

Total mite infestation (*v_load*) was estimated by combining the phoretic and brood infestations:

$$v_load = v_brood \times n_{brood} + v_pho \times \frac{n_{bee}}{100} \tag{2}$$

where the total number of brood cells (n_{brood}) and adult bees n_{bee} in the colonies were

583 estimated using the ColEval method [86].

584 **Recapping of infested cells** The uncapping and further recapping of varroa infested brood
585 cells by adult honey bees is a behavioural trait that has been shown to be associated with
586 varroa resistance [16]. It can be estimated by dissecting brood cells to detect the presence
587 (non-recapped cell) or absence (recapped cell) of the silk cocoon built by the larvae [87]. This
588 was measured on the colony at the same time as measuring mite non reproduction (MNR). This
589 trait is expressed as the proportion of recapped cells within the infested cells.

590 **Mite non reproduction** Mite Non Reproduction (MNR), originally known as Suppressed
591 Mite reproduction (SMR), was estimated as detailed in Mondet et al. 2020 [15]. In brief,
592 this estimates infers varroa reproductive status for each brood cell infested by a single varroa
593 foundress and provides a proportion of reproductive mites in the colony. It was estimated on
594 about 300 brood cells (some also used to determine mite brood infestation) with the aim to
595 reach at least 35 single mite infested cells.
596 All phenotypes were recorded by technicians having followed an extensive training period
597 prior to sampling. Moreover, for statistical analysis they were corrected to fit the assumption of
598 Normality underlying genome wide association study models. Details on the transformations
599 can be found in Supplementary methods 1.

600 **Phenotypic Characterisation of colonies** The correlation between varroa-associated phe-
601 notypes within and across groups were estimated using the traditional Pearson's method. A
602 Principal Component Analysis (PCA) was performed using the R package FactoMineR [88],
603 after imputation of the missing phenotypes using the R package MissMDA [89]. Missing
604 phenotypes appeared due to sampling difficulties occurring when performing the scoring in
605 the field. The PCA was used to extract uncorrelated synthetic phenotypes to test for genetic
606 association.

607 Genome wide association studies and meta-analyses

608 **Genomic relationship matrix** For each group of genetic background identified earlier, *A.*
609 *m. ligustica & carnica*, *A. m. mellifera* and the hybrids, only SNPs with a MAF above
610 0.01 and missing rate below 5% were kept. A genomic relationship matrix (GRM) between
611 colonies of the group was estimated on pool sequencing experiment allele frequencies taking
612 SNP linkage disequilibrium into account through the SNP weights produced by LDAK [90],
613 see supplementary methods 2 for more details. Additionally, in order to describe further genetic
614 structure within the group, a PCA on the GRM was performed, using LDAK [91]. The Horn's
615 parallel analysis [92] was used to decide on the number of principal components kept to explain

the variance. 20, 12 and 16 first components were kept from this PCA for *A. m. ligustica & carnica*, *A. m. mellifera* and the hybrids respectively.

Genome wide association Genome wide association studies (GWAS) were performed for three traits: varroa infestation (thereafter called *varrao_inf*), MNR and recapping of varroa infested cells (*recap*). Each GWAS tested the association between the reconstructed queen genotypes and the phenotype using the univariate linear mixed model (`lmm`) as proposed in GEMMA [42] at each SNP in turn, resulting for each SNP in an estimate of its effect and associated standard error, as well as a p-value. In the GWAS, polygenic effects were accounted for with the GRM described above. In addition, further correction was performed by adding the principal components from the PCA on GRM, selected as explained above, as covariates. This was done to correct for the effects of unmeasured confounders with the genetic structure on the phenotype variation (such as apiaries, beekeeper, year ... effects). In supplementary methods 2 we illustrate how the structures of the GRMs correlate somewhat to different environmental structures in the data. Association studies were run for all markers initially available, after filtering for minor allele frequency (MAF) above 0.01 and missing rate below 5%. This led to retain a total of 3,084,335; 2,729,072 and 3,185,994 SNPs for the *A. m. ligustica & carnica*, *A. m. mellifera* and hybrid individuals respectively.

To assess the effectiveness of the correction for population structure, the genomic inflation factor λ_{gc} was estimated as the median of the chi-squared test statistics divided by the expected median of the chi-squared distribution under the null hypothesis. λ_{gc} ranged between 1.02 and 1.08 for the GWAS on *A. m. ligustica & carnica*, between 0.98 and 1.03 for the GWAS on *A. m. mellifera* and between 0.99 and 1.04 for the GWAS on the hybrid colonies therefore showing really little inflation or deflation of the p-values associated to the tested SNPs (supplementary figures SF1).

To call SNPs significant, a False Discovery Rate procedure was applied, using the adaptive shrinkage method [43] as implemented in the `ashr` R package. Specifically, SNPs with a local false discovery rate (lfdr) and local false sign rate (lfsr) < 0.1 were deemed significant. The lfdr is the probability, knowing the observed data, that an effect is declared significant erroneously and lfsr is the probability, knowing observed data, that the sign of an effect declared significant is wrong [43]. The proportion of phenotypic variance explained by the SNPs (*pve*), and its standard error, was estimated by the univariate linear model and, using the Bayesian Sparse linear mixed model (`bslmm`), with default parameters of 0 t 300 SNPs, 1,000,000 sampling steps and 100,000 burn-in iterations, proposed by GEMMA [93]. This model was fitted with the GRM and associated covariates and we estimated the proportion of genetic variance explained by the sparse effects (*pge*) of the trait as well as 95% credible interval from empirical estimates.

Meta-analysis GWAS results for the same phenotype on the three genetic backgrounds were combined with two meta-GWAS methods: (i) MANTRA [40] a meta-analysis method dedicated

to combine GWAS results from different genetic ancestries (ii) Mash [41] a general purpose data-driven Bayesian meta-analysis method modeling SNP effects with a mixture of multivariate gaussian distributions with different correlation matrices.

MANTRA and mash were run on all SNPs, using effects (β) and associated standard errors estimated with the `lmm` model of GEMMA. For mash inferences, canonical and data-driven covariance matrices were used. The canonical matrices were estimated automatically by mash. Data-driven matrices were: (i) estimated based on extreme deconvolution from PCA matrices, (ii) based on Fst values between populations, similar to MANTRA and (iii) based on correlation between SNP or gene effects in the different groups. Mash includes an estimation of the residual correlation. In this analysis the simple residual correlation estimation model was preferred, as it outperformed more complex residual correlation estimation models. SNPs with a log10(Bayes Factor) > 5 with MANTRA were called significant, a threshold which was shown to be conservative by Wang et al. 2013 [94]. Mash automatically assigns significance to each marker. In our study the corresponding log10(BF) threshold varied from 1.16 to 1.4 depending on the trait.

Genetic correlations were estimated using Pearson's correlation coefficients on the allele effects, for each SNP, calculated by our different GWAS methods (individually with GEMMA and `ash`, and across co-ancestries with MANTRA and mash).

Gene prioritisation The variant effect predictor (VEP) tool from Ensembl [95] was used to identify, based on the honey bee genome annotation, for each of the significant SNPs, its impact on the annotation (stop, gained or lost, missense, frameshift...), its closest genes and their locations (upstream, downstream, intronic region ...). Additionally, we identified genes, located elsewhere in the genome in LD regions, containing variants in high LD ($r^2 > 0.8$) with the significant SNP. Linkage disequilibrium was computed, for each group, using data from the genetic diversity reference panel [39], for each significant SNP with all other variants on the same chromosome using the plink software version 1.9[82, 83] with options –ld-window-r2 0.8 –r2 inter-chr.

Data availability statement

Raw sequences are made available under the project SRxxxx on NCBI (link). Additional data such as the mpileup files, depth and count files, SNP weights, genomic relationship matrices, PCA, phenotypes, summary statistics from GWAS analysis ... are available in the following depository: link. The scripts to perform the analysis are available on github (https://github.com/seynard/gwas_beestrong)

Competing interests

The authors declare that they have no competing interests.

Author's contributions

AV, FM, LG, YLC, BB, BS, FP and AD designed the study and obtained the funding for the BeeStrong project. LG led the BeeStrong project while FM, YLC, AD, AV and BS supervised this study. SEE, FM, AV and BS coordinated the execution of this project. AV, FM, YLC, BB, BS, and AD designed the data collection while FM, BB, YP, CS, BD, MG and the BeeStrong technicians, performed the data collection. BL and JM provided part of the data within this collaboration. KT, EL, RM performed the DNA extraction, library preparation, OB and sequencing platform GeT-PlaGe employees, TB and RFS, performed the sequencing. SEE, FM, BB, BD, MG, MN, YP, CS and BS perform data curation. BS, SEE and AV developed the methodology. SEE, BS, AV and FG designed and performed the association studies. SEE, BS, AV, FM, FG, FP, and YLC interpreted the results. SEE, FM, AV and BS drafted the original manuscript and contributed to the evolution of the draft. BB, BL, OB, YLC, BD, AD, LG, MG, JM, FG, MN, FP, KT reviewed the latest version. SEE, AV and BS created the visuals of this paper. All authors have read and approved the manuscript.

Acknowledgements

This study was performed thanks to the consent of more than a hundred private beekeepers and beekeeping groups to take part in this project, including Arista Bee Research (NL) and Betta Bees Research (NZ). This study was performed with the support of the ITSAP team (Anne-Laure Guirao and Antoine Sudan) for the maintenance of the honey bee colonies from the ITSAP experimental station and Benoît Droz for the maintenance of the Swiss colonies, the multiple technicians employed on the project (Baptiste Biet, Marius Bredon, Samuel Favier, Héléne Gerwis, Coralie Guerry, Albéric Delamotte, Pierre Lamy, Antonin Leclercq, Julie Petit, Diane Roriz, Emilien Rottier, Justine Rougé, Agathe Valette, Chloé Vasseur and Alice Revel) for visiting beekeepers and collecting the data for the project. We thank the sequencing platform GeT-PlaGe, Toulouse (France), a partner of the National Infrastructure France Génomique, thanks to support by the Commissariat aux Grands Invetissements (ANR-10-INBS-0009), for the sequencing. Bioinformatics analyses were performed on the computing facility Genotoul. This research was funded by FranceAgriMer, the Ministère de l'Agriculture de l'Agroalimentaire et de la Forêt within the framework of MOSAR RT 2015-776 project and the Ministère de l'Agriculture de l'Agroalimentaire et de la Forêt and Investissement d'avenir for BeeStrong PIA P3A and Bundesamt für Landwirtschaft BLW (Swiss Federal Office for Agriculture; grant no. 627000708). The authors thank LABOGENA DNA for carrying this project. Thanks to

François Gerster for initial discussion and his help to build the consortium and target appropriate funding.

References

[1] A.M. Klein et al. "Importance of pollinators in changing landscapes for world crops". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1608 (2006), pp. 303–313. DOI: 10.1098/rspb.2006.3721. URL: https://doi.org/10.1098/rspb.2006.3721.

[2] S. Bruckner et al. "A national survey of managed honey bee colony losses in the USA: results from the Bee Informed Partnership for 2017–18, 2018–19, and 2019–20". In: *Journal of Apicultural Research* (2023), pp. 1–15. ISSN: 0021-8839. DOI: 10.1080/00218839.2022.2158586. URL: https://doi.org/10.1080/00218839.2022.2158586.

[3] S. G. Potts et al. "Global pollinator declines: trends, impacts and drivers". In: *Trends in Ecology and Evolution* 25.6 (2010), pp. 345–353. ISSN: 0169-5347. DOI: https://doi.org/10.1016/j.tree.2010.01.007. URL: https://www.sciencedirect.com/science/article/pii/S0169534710000364.

[4] D. Goulson et al. "Bee declines driven by combined stress from parasites, pesticides, and lack of flowers". In: *Science* 347.6229 (2015), p. 1255957. DOI: 10.1126/science.1255957. URL: https://doi.org/10.1126/science.1255957.

[5] K. Traynor et al. "*Varroa destructor*: A Complex Parasite, Crippling Honeybees Worldwide". In: (2020). DOI: 10.20944/preprints202002.0374.v1.

[6] Y. Le Conte, M. Ellis, and W. Ritter. "Varroa mites and honey bee health: can Varroa explain part of the colony losses?*". In: *Apidologie* 41.3 (2010), pp. 353–363. URL: https://doi.org/10.1051/apido/2010017.

[7] P. Rosenkranz, P. Aumeier, and B. Ziegelmann. "Biology and control of *Varroa destructor*". In: 103.Supplement (2010), S96–S119. URL: http://www.sciencedirect.com/science/article/pii/S0022201109001906.

[8] J. de Miranda et al. "Honey bee viruses and their effect on bee and colony health". In: 2011, pp. 71–102.

[9] I. Fries, A. Imdorf, and P. Rosenkranz. "Survival of mite infested (*Varroa destructor*) honey bee (*Apis mellifera*) colonies in a Nordic climate". In: *http://dx.doi.org/10.1051/apido:2006031* 37 (2006). DOI: 10.1051/apido:2006031.

[10] A. Noël, Y. Le Conte, and F. Mondet. "*Varroa destructor*: how does it harm *Apis mellifera* honey bees and what can be done about it?" In: *Emerging Topics in Life Sciences* 4.1 (2020), pp. 45–57. ISSN: 2397-8554. DOI: 10.1042/ETLS20190125. URL: https://doi.org/10.1042/ETLS20190125.

[11] S. Cremer, S. A. O. Armitage, and P. Schmid-Hempel. "Social Immunity". In: *Current Biology* 17.16 (2007), R693–R702. DOI: https://doi.org/10.1016/j.cub.2007.06.008. URL: https://www.sciencedirect.com/science/article/pii/S0960982207015035.

[12] S. Thaduri et al. "Disentangling host-parasite-pathogen interactions in a varroa-resistant honeybee population reveals virus tolerance as an independent, naturally adapted survival mechanism". In: *Scientific Reports* (2019). ISSN: 2045-2322. DOI: 10.1038/s41598-019-42741-6. URL: https://doi.org/10.1038/s41598-019-42741-6.

[13] J. J. M. van Alphen and B. J. Fernhout. "Natural selection, selective breeding, and the evolution of resistance of honeybees (*Apis mellifera*) against Varroa". In: *Zoological Letters* 6 (2020). ISSN: 2056-306X. DOI: 10.1186/s40851-020-00158-4. URL: https://doi.org/10.1186/s40851-020-00158-4.

[14] J. R. Harbo and J. W. Harris. "Suppressed mite reproduction explained by the behaviour of adult bees". In: *Journal of Apicultural Research* 44.1 (2005), pp. 21–23. ISSN: 0021-8839. DOI: 10.1080/00218839.2005.11101141. URL: https://doi.org/10.1080/00218839.2005.11101141.

[15] F. Mondet et al. "Evaluation of Suppressed Mite Reproduction (SMR) Reveals Potential for Varroa Resistance in European Honey Bees (*Apis mellifera L.*)" In: *Insects* 11.9 (2020). ISSN: 2075-4450. DOI: 10.3390/insects11090595.

[16] M. Oddie, B. Dahle, and P. Neumann. "Reduced Postcapping Period in Honey Bees Surviving *Varroa destructor* by Means of Natural Selection". In: *Insects* 9.4 (2018).

[17] J. R. Harbo and J. W. Harris. "Heritability in honey bees (Hymenoptera: Apidae) of characteristics associated with resistance to *Varroa jacobsoni* (Mesostigmata: Varroidae)". In: 92.2 (1999), pp. 261–265. DOI: 10.1093/jee/92.2.261. URL: http://dx.doi.org/10.1093/jee/92.2.261.

[18] A. von Virag et al. "Decreased Mite Reproduction to Select Varroa destructor (Acari: Varroidae) Resistant Honey Bees (Hymenoptera: Apidae): Limitations and Potential Methodological Improvements". In: *Journal of Economic Entomology* 115.3 (Apr. 2022), pp. 695–705. ISSN: 0022-0493. DOI: 10.1093/jee/toac022. eprint: https://academic.oup.com/jee/article-pdf/115/3/695/43977501/toac022.pdf. URL: https://doi.org/10.1093/jee/toac022.

[19] T. E. Rinderer et al. “Breeding for resistance to *Varroa destructor* in North America*”. In: *Apidologie* 41.3 (2010), pp. 409–424. URL: https://doi.org/10.1051/apido/2010015.

[20] R. Büchler et al. “Evaluation of Traits for the Selection of *Apis Mellifera* for Resistance against *Varroa Destructor*”. In: *Insects* 11.9 (2020). ISSN: 2075-4450. DOI: 10.3390/insects11090618.

[21] W. C. Rothenbuhler. “Behavior genetics of nest cleaning in honey bees. IV. Responses of F1 and backcross generations to disease-killed blood”. In: *Am Zool* 4 (1964), pp. 111–23. ISSN: 0003-1569 (Print) 0003-1569. DOI: 10.1093/icb/4.2.111.

[22] K. Bienefeld and F. Pirchner. “Heritabilities for several colony traits in the honeybee (*Apis mellifera carnica*)”. In: *Apidologie* 21.3 (1990), pp. 175–183. URL: https://doi.org/10.1051/apido:19900302.

[23] J. R. Harbo and J. W. Harris. “Responses to Varroa by honey bees with different levels of Varroa Sensitive Hygiene”. In: *Journal of Apicultural Research* 48.3 (2009), pp. 156–161. ISSN: 0021-8839. DOI: 10.3896/IBRA.1.48.3.02. URL: https://doi.org/10.3896/IBRA.1.48.3.02.

[24] S. E. Eynard et al. “Descriptive Analysis of the Varroa Non-Reproduction Trait in Honey Bee Colonies and Association with Other Traits Related to Varroa Resistance”. In: *Insects* 11.8 (2020). ISSN: 2075-4450 (Print) 2075-4450. DOI: 10.3390/insects11080492.

[25] M. Guichard et al. “Advances and perspectives in selecting resistance traits against the parasitic mite *Varroa destructor* in honey bees”. In: *Genetics Selection Evolution* 52.1 (2020), p. 71. ISSN: 1297-9686. DOI: 10.1186/s12711-020-00591-1. URL: https://doi.org/10.1186/s12711-020-00591-1.

[26] D. R. Tarpy, S. Hatch, and D. J. C. Fletcher. “The influence of queen age and quality during queen replacement in honeybee colonies”. In: *Animal Behaviour* 59.1 (2000), pp. 97–101. ISSN: 0003-3472. DOI: https://doi.org/10.1006/anbe.1999.1311. URL: https://www.sciencedirect.com/science/article/pii/S0003347299913118.

[27] A. Zayed. “Effective population size in Hymenoptera with complementary sex determination”. In: *Heredity* 93.6 (2004), pp. 627–630. ISSN: 1365-2540. DOI: 10.1038/sj.hdy.6800588. URL: https://doi.org/10.1038/sj.hdy.6800588.

[28] M. Beye et al. “The Gene csd Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein”. In: *Cell* 114.4 (2003), pp. 419–429. ISSN: 0092-8674. DOI: https://doi.org/10.1016/S0092-8674(03)00606-8. URL: https://www.sciencedirect.com/science/article/pii/S0092867403006068.

[29] J. Woyke. “Drones from fertilized eggs and the biology of sex determination in the honey bee, Apimondia, Prague 12-17.08.1963”. In: Aug. 1963. DOI: 10.13140/2.1.4890.8161.

[30] S. E. Eynard et al. “Reconstructing queen genotypes by pool sequencing colonies in eusocial insects: Statistical Methods and their application to honeybee”. In: *Molecular Ecology Resources* 22.8 (2022), pp. 3035–3048. ISSN: 1755-098X. DOI: https://doi.org/10.1111/1755-0998.13685. URL: https://doi.org/10.1111/1755-0998.13685.

[31] J. R. Shorter et al. “A Genetic Analysis of the Stinging and Guarding Behaviors of the Honey Bee”. In: *Behavior Genetics* 42.4 (2012), pp. 663–674. ISSN: 1573-3297. DOI: 10.1007/s10519-012-9530-5. URL: https://doi.org/10.1007/s10519-012-9530-5.

[32] Y. Liu et al. “Larva-mediated chalkbrood resistance-associated single nucleotide polymorphism markers in the honey bee *Apis mellifera*”. In: *Insect Molecular Biology* 25.3 (2016), pp. 239–250. ISSN: 0962-1075. DOI: https://doi.org/10.1111/imb.12216. URL: https://doi.org/10.1111/imb.12216.

[33] B. R. Southey et al. “Characterization of Genomic Variants Associated with Scout and Recruit Behavioral Castes in Honey Bees Using Whole-Genome Sequencing”. In: *PLOS ONE* 11.1 (2016), e0146430. DOI: 10.1371/journal.pone.0146430. URL: https://doi.org/10.1371/journal.pone.0146430.

[34] B. H. Conlon et al. “A gene for resistance to the Varroa mite (Acari) in honey bee (*Apis mellifera*) pupae”. In: *Molecular Ecology* 0.ja (2019). ISSN: 0962-1083. DOI: 10.1111/mec.15080. URL: https://doi.org/10.1111/mec.15080.

[35] A. Avalos et al. “Genomic regions influencing aggressive behavior in honey bees are defined by colony allele frequencies”. In: *Proceedings of the National Academy of Sciences* (2020), p. 201922927. DOI: 10.1073/pnas.1922927117. URL: http://www.pnas.org/content/early/2020/07/01/1922927117.abstract.

[36] M. Guichard et al. “Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences”. In: *Animal genetics* (2021). ISSN: 0268-9146. DOI: 10.1111/age.13070. URL: https://doi.org/10.1111/age.13070.

[37] M. Guichard et al. “Exploring Two Honey Bee Traits for Improving Resistance Against *Varroa destructor*: Development and Genetic Evaluation”. In: *Insects* 12.3 (2021). ISSN: 2075-4450. DOI: 10.3390/insects12030216.

[38] M. Guichard et al. “Two quantitative trait loci are associated with recapping of Varroa destructor-infested brood cells in *Apis mellifera mellifera*”. In: *Animal genetics* 1365-2052 (Electronic) (2022). DOI: 10.1111/age.13150.

[39] D. Wragg et al. "Complex population structure and haplotype patterns in the Western European honey bee from sequencing a large panel of haploid drones". In: *Molecular Ecology Resources* 22.8 (2022), pp. 3068–3086. ISSN: 1755-098X. DOI: https://doi.org/10.1111/1755-0998.13665. URL: https://doi.org/10.1111/1755-0998.13665.

[40] A. P. Morris. "Transethnic meta-analysis of genomewide association studies". In: *Genetic epidemiology* 35.8 (2011), pp. 809–822. ISSN: 1098-2272 0741-0395. DOI: 10.1002/gepi.20630. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3460225/.

[41] S. M. Urbut et al. "Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions". In: *Nature Genetics* 51.1 (2019), pp. 187–195. ISSN: 1546-1718. DOI: 10.1038/s41588-018-0268-8. URL: https://doi.org/10.1038/s41588-018-0268-8.

[42] X. Zhou and M. Stephens. "Genome-wide efficient mixed-model analysis for association studies". In: *Nature genetics* 44.7 (2012), pp. 821–824. URL: https://www.ncbi.nlm.nih.gov/pmc/PMC3386377/.

[43] M. Stephens. "False discovery rates: a new deal". In: *Biostatistics* 18.2 (2017), pp. 275–294. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxw041. URL: https://doi.org/10.1093/biostatistics/kxw041.

[44] A. Wallberg et al. "A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds". In: *BMC Genomics* 20.1 (2019), p. 275. ISSN: 1471-2164. DOI: 10.1186/s12864-019-5642-0. URL: https://doi.org/10.1186/s12864-019-5642-0.

[45] E. V. Ryabov et al. "A Virulent Strain of Deformed Wing Virus (DWV) of Honeybees (*Apis mellifera*) Prevails after Varroa destructor-Mediated, or In Vitro, Transmission". In: *PLOS Pathogens* 10.6 (2014), e1004230. DOI: 10.1371/journal.ppat.1004230. URL: https://doi.org/10.1371/journal.ppat.1004230.

[46] F. Mondet et al. "Antennae hold a key to Varroa-sensitive hygiene behaviour in honey bees". In: *Scientific Reports* 5.1 (2015), p. 10454. ISSN: 2045-2322. DOI: 10.1038/srep10454. URL: https://doi.org/10.1038/srep10454.

[47] J. M. Tsuruda et al. "High-resolution linkage analyses to identify genes that influence Varroa sensitive hygiene behavior in honey bees". In: *PLOS ONE* 7.11 (2012), e48276. URL: https://doi.org/10.1371/journal.pone.0048276.

[48] P. Saelao et al. "Genome-wide patterns of differentiation within and among U.S. commercial honey bee stocks". In: *BMC Genomics* 21.1 (2020), p. 704. ISSN: 1471-2164. DOI: 10.1186/s12864-020-07111-x. URL: https://doi.org/10.1186/s12864-020-07111-x.

[49] E. W. Teixeira et al. “Transcriptomic analysis suggests candidate genes for hygienic behavior in African-derived *Apis mellifera* honeybees”. In: *Apidologie* (2021). ISSN: 1297-9678. DOI: 10.1007/s13592-020-00834-6. URL: https://doi.org/10.1007/s13592-020-00834-6.

[50] S. A. Ament et al. “Mechanisms of stable lipid loss in a social insect”. In: *Journal of Experimental Biology* 214.22 (2011), pp. 3808–3821. ISSN: 0022-0949. DOI: 10.1242/jeb.060244. URL: https://doi.org/10.1242/jeb.060244.

[51] M. F. Feldlaufer and K. Hartfelder. “Relationship of the neutral sterols and ecdysteroids of the parasitic mite, *Varroa jacobsoni* to those of the honey bee, *Apis mellifera*”. In: *Journal of Insect Physiology* 43.6 (1997), pp. 541–545. ISSN: 0022-1910. DOI: https://doi.org/10.1016/S0022-1910(97)00005-X. URL: https://www.sciencedirect.com/science/article/pii/S002219109700005X.

[52] C. M. Aurori et al. “Juvenile hormone pathway in honey bee larvae: A source of possible signal molecules for the reproductive behavior of *Varroa destructor*”. In: *Ecology and Evolution* 11.2 (2021), pp. 1057–1068. ISSN: 2045-7758. DOI: https://doi.org/10.1002/ece3.7125. URL: https://doi.org/10.1002/ece3.7125.

[53] A. S. French et al. “The role of serotonin in feeding and gut contractions in the honeybee”. In: *Journal of insect physiology* 61.100 (2014), pp. 8–15. DOI: 10.1016/j.jinsphys.2013.12.005. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969292/.

[54] H. Verlinden, R. Vleugels, and J. Vanden Broeck. “Serotonin, serotonin receptors and their actions in insects”. In: *Neurotransmitter* 2 (2015), e314. DOI: 10.14800/nt.314.

[55] S. Forêt and R. Maleszka. “Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*)”. In: *Genome research* 16.11 (2006), pp. 1404–1413. DOI: 10.1101/gr.5075706. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1626642/.

[56] B. Mam, S. D. Karpe, and R. Sowdhamini. “Minus-C subfamily has diverged from Classic odorant-binding proteins in honeybees”. In: *Apidologie* 54.1 (2023), p. 12. ISSN: 1297-9678. DOI: 10.1007/s13592-022-00988-5. URL: https://doi.org/10.1007/s13592-022-00988-5.

[57] H. Gebremedhn et al. “Expression of Molecular Markers of Resilience against *Varroa destructor* and Bee Viruses in Ethiopian Honey Bees (*Apis mellifera simensis*) Focussing on Olfactory Sensing and the RNA Interference Machinery”. In: 14.5 (2023). DOI: 10.3390/insects14050436.

[58] S. Jiang et al. "Differential gene expression of two extreme honey bee (*Apis mellifera*) colonies showing varroa tolerance and susceptibility". In: *Insect Molecular Biology* 25.3 (2016), pp. 272–282. ISSN: 0962-1075. DOI: https://doi.org/10.1111/imb.12217. URL: https://doi.org/10.1111/imb.12217.

[59] H. Hu et al. "Proteome Analysis of the Hemolymph, Mushroom Body, and Antenna Provides Novel Insight into Honeybee Resistance against Varroa Infestation". In: *Journal of Proteome Research* 15.8 (2016), pp. 2841–2854. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.6b00423. URL: https://doi.org/10.1021/acs.jproteome.6b00423.

[60] M. Guarna et al. "A search for protein biomarkers links olfactory signal transduction to social immunity". In: *BMC Genomics* 16 (2015). DOI: 10.1186/s12864-014-1193-6.

[61] A. McAfee et al. "A death pheromone, oleic acid, triggers hygienic behavior in honey bees (*Apis mellifera L.*)" In: *Scientific Reports* 8.1 (2018), p. 5719. ISSN: 2045-2322. DOI: 10.1038/s41598-018-24054-2. URL: https://doi.org/10.1038/s41598-018-24054-2.

[62] M. Marta Guarna et al. "Peptide biomarkers used for the selective breeding of a complex polygenic trait in honey bees". In: *Scientific Reports* 7.1 (2017), p. 8381. ISSN: 2045-2322. DOI: 10.1038/s41598-017-08464-2. URL: https://doi.org/10.1038/s41598-017-08464-2.

[63] M. Gabel et al. "Heritability of *Apis mellifera* recapping behavior and suppressed mite reproduction as resistance traits towards *Varroa destructor*". In: *Frontiers in Insect Science* 3 (2023). ISSN: 2673-8600. URL: https://www.frontiersin.org/articles/10.3389/finsc.2023.1135187.

[64] S. Lorenz. "The endangerment of bees and new developments in beekeeping: a social science perspective using the example of Germany". In: *International Journal of Environmental Studies* 73.6 (2016), pp. 988–1005. ISSN: 0020-7233. DOI: 10.1080/00207233.2016.1220703. URL: https://doi.org/10.1080/00207233.2016.1220703.

[65] M. Du et al. "Consequences of incorrect genetic parameter estimates for single-trait and multi-trait genetic evaluations in honeybees". In: *Journal of Animal Breeding and Genetics* 139.6 (2022), pp. 666–678. ISSN: 0931-2668. DOI: https://doi.org/10.1111/jbg.12728. URL: https://doi.org/10.1111/jbg.12728.

[66] M. Du et al. "Influence of model selection and data structure on the estimation of genetic parameters in honeybee populations". In: *G3 Genes|Genomes|Genetics* (2022), jkab450. ISSN: 2160-1836. DOI: 10.1093/g3journal/jkab450. URL: https://doi.org/10.1093/g3journal/jkab450.

[67] A. Hoppe et al. "Heritability of disease resistance to chronic bee paralysis, chalkbrood and nosemosis in the honeybee (*A.m.carnica*)". In: *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Wageningen Academic Publishers, 2022.

[68] A. Uzunov et al. "606. Evaluating the potential for mating control in honey bee breeding in three SE European countries (preliminary results)". In: *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Wageningen Academic Publishers, 2022. Chap. 606, pp. 2508–2511. DOI: doi:10.3920/978-90-8686-940-4_60610.3920/978-90-8686-940-4_606. URL: https://doi.org/10.3920/978-90-8686-940-4_606.

[69] R. Bernstein et al. "First large-scale genomic prediction in the honey bee". In: *Heredity* (2023). ISSN: 1365-2540. DOI: 10.1038/s41437-023-00606-9. URL: https://doi.org/10.1038/s41437-023-00606-9.

[70] R. Bernstein et al. "Simulation studies to optimize genomic selection in honey bees". In: *Genetics Selection Evolution* 53.1 (2021), p. 64. ISSN: 1297-9686. DOI: 10.1186/s12711-021-00654-x. URL: https://doi.org/10.1186/s12711-021-00654-x.

[71] J. C. Jones et al. "Tool for genomic selection and breeding to evolutionary adaptation: Development of a 100K single nucleotide polymorphism array for the honey bee". In: *Ecology and Evolution* 10.13 (2020), pp. 6246–6256. ISSN: 2045-7758. DOI: 10.1002/ece3.6357. URL: https://doi.org/10.1002/ece3.6357.

[72] "Epigenetics: What makes a queen bee?" In: *Nature* 468.7322 (2010), pp. 348–348. ISSN: 1476-4687. DOI: 10.1038/468348a. URL: https://doi.org/10.1038/468348a.

[73] S. Foret et al. "Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes". In: *BMC Genomics* 10.1 (2009), p. 472. ISSN: 1471-2164. DOI: 10.1186/1471-2164-10-472. URL: https://doi.org/10.1186/1471-2164-10-472.

[74] H. Li-Byarlay et al. "Transcriptomic and Epigenomic Dynamics of Honey Bees in Response to Lethal Viral Infection". In: *Frontiers in Genetics* 11 (2020). ISSN: 1664-8021. URL: https://www.frontiersin.org/articles/10.3389/fgene.2020.566320.

[75] H. J. Penn et al. "Honey Bee Genetic Stock Determines Deformed Wing Virus Symptom Severity but not Viral Load or Dissemination Following Pupal Exposure". In: *Frontiers in Genetics* 13 (2022). ISSN: 1664-8021. URL: https://www.frontiersin.org/articles/10.3389/fgene.2022.909392.

[76] H. Li. "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv preprint arXiv:1303.3997* (2013).

[77] M. A. Techer et al. "Divergent evolutionary trajectories following speciation in two ectoparasitic honey bee mites". In: *Communications Biology* 2.1 (2019), p. 357. DOI: 10.1038/s42003-019-0606-0. URL: https://doi.org/10.1038/s42003-019-0606-0.

[78] H. Li and R. Durbin. "Fast and accurate short read alignment with Burrows–Wheeler transform". In: *Bioinformatics* 25.14 (2009), pp. 1754–1760. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp324. URL: https://doi.org/10.1093/bioinformatics/btp324.

[79] R. Kofler, R. V. Pandey, and C. Schlötterer. "PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq)". In: *Bioinformatics* 27.24 (2011), pp. 3435–3436. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr589. URL: https://doi.org/10.1093/bioinformatics/btr589.

[80] K. A. Dogantzis et al. "Thrice out of Asia and the adaptive radiation of the western honey bee". In: *Science Advances* 7.49 (2021), eabj2151. DOI: 10.1126/sciadv.abj2151. URL: https://doi.org/10.1126/sciadv.abj2151.

[81] J. Obšteter et al. "Inferring whole-genome tree sequences and population and demographic parameters of the Western honeybee". In: *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Wageningen Academic Publishers, 2022.

[82] Purcell S. and Chang C. *PLINK 1.9*. URL: www.cog-genomics.org/plink/1.9/.

[83] S. Purcell et al. "PLINK: a tool set for whole-genome association and population-based linkage analyses". In: *American journal of human genetics* 81.3 (2007), pp. 559–575. ISSN: 0002-9297 1537-6605. DOI: 10.1086/519795. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1950838/.

[84] J. K. Pritchard, M. Stephens, and P. Donnelly. "Inference of population structure using multilocus genotype data". In: *Genetics* 155.2 (2000), pp. 945–959. ISSN: 0016-6731. URL: %3CGo%20to%20ISI%3E://WOS:000087475100039.

[85] V. Dietemann et al. "Standard methods for varroa research". In: *Journal of Apicultural Research* 52 (2012), pp. 1–54. DOI: 10.3896/IBRA.1.52.1.09.

[86] J. Hernandez et al. "ColEval: Honeybee COLony Structure EVALuation for Field Surveys". In: *Insects* 11.1 (2020). DOI: 10.3390/insects11010041.

[87] R. Büchler et al. "Screening for low Varroa mite reproduction (SMR) and recapping in European honey bees". In: (2017).

[88] S. Lê, J. Josse, and F. Husson. "FactoMineR: An R Package for Multivariate Analysis". In: *Journal of Statistical Software* 25.1 (2008), pp. 1–18. DOI: 10.18637/jss.v025.i01. URL: https://www.jstatsoft.org/index.php/jss/article/view/v025i01.

[89] J. Josse and F. Husson. "missMDA: A Package for Handling Missing Values in Multivariate Data Analysis". In: *Journal of Statistical Software* 70.1 (2016), pp. 1–31. DOI: 10.18637/jss.v070.i01. URL: https://www.jstatsoft.org/index.php/jss/article/view/v070i01.

[90] D. Speed et al. "Reevaluation of SNP heritability in complex human traits". In: *Nature Genetics* 49.7 (2017), pp. 986–992. ISSN: 1546-1718. DOI: 10.1038/ng.3865. URL: https://doi.org/10.1038/ng.3865.

[91] D. Speed et al. "Improved heritability estimation from genome-wide SNPs". In: *American journal of human genetics* 91.6 (2012), pp. 1011–1021. ISSN: 1537-6605 0002-9297. DOI: 10.1016/j.ajhg.2012.10.010. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3516604/.

[92] A. Dinno. "Implementing Horn's Parallel Analysis for Principal Component Analysis and Factor Analysis". In: *Stata Journal* 9 (2009), pp. 291–298. DOI: 10.1177/1536867X0900900207.

[93] X. Zhou, P. Carbonetto, and M. Stephens. "Polygenic Modeling with Bayesian Sparse Linear Mixed Models". In: *PLOS Genetics* 9.2 (2013), e1003264. DOI: 10.1371/journal.pgen.1003264. URL: https://doi.org/10.1371/journal.pgen.1003264.

[94] X. Wang et al. "Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies". In: *Human Molecular Genetics* 22.11 (Feb. 2013), pp. 2303–2311. ISSN: 0964-6906. DOI: 10.1093/hmg/ddt064. eprint: https://academic.oup.com/hmg/article-pdf/22/11/2303/17258469/ddt064.pdf. URL: https://doi.org/10.1093/hmg/ddt064.

[95] W. McLaren et al. "The Ensembl Variant Effect Predictor". In: *Genome Biology* 17.1 (2016), p. 122. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0974-4. URL: https://doi.org/10.1186/s13059-016-0974-4.