Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences

Matthieu Guichard, Benjamin Dainat, Sonia Eynard, Alain Vignal, Bertrand Servin, The Beestrong Consortium, Markus Neuditschko, Rachid Mahla, Yannick Poquet, François Guillaume, et al.

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

FULL PAPER

doi: 10.1111/age.13070

Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences

M. Guichard*,†, B. Dainat*, S. Eynard‡,§, A. Vignal‡,§, B. Servin‡,§, the Beestrong Consortium¶, M. Neuditschko†, Sonia Eynard†, Bertrand Servin†, Rachid Mahla†, Yannick Poquet†, François Guillaume†, Yves Le Conte†, Benjamin Basso†, Florence Phocas† and Olivier Bouchez†

*Agroscope, Swiss Bee Research Centre, Schwarzenburgstrasse 161, Bern 3003, Switzerland. †Agroscope, Animal GenoPhenomics, Rte de la Tioleyre 4, Posieux 1725, Switzerland. ‡GenPhySE, INRA, INPT, INPENVT, Université de Toulouse, Castanet-Tolosan 31320, France. §UMT PrADE, Protection des Abeilles Dans l'Environnement, Avignon 84914, France. ¶Domaine de Vilvert, Bat 224, CS80009, Jouy-en-Josas CEDEX 78353, France.

Summary

The identification of quantitative trait loci (QTL) through genome-wide association studies (GWAS) is a powerful method for unravelling the genetic background of selected traits and improving early-stage predictions. In honey bees (*Apis mellifera*), past genetic analyses have particularly focused on individual queens and workers. In this study, we used pooled whole-genome sequences to ascertain the genetic variation of the entire colony. In total, we sampled 216 *Apis mellifera mellifera* and 28 *Apis mellifera carnica* colonies. Different experts subjectively assessed the gentleness and calmness of the colonies using a standardised protocol. Conducting a GWAS for calmness on 211 purebred *A. m. mellifera* colonies, we identified three QTL, on chromosomes 8, 6, and 12. The two first QTL correspond to *LOC409692* gene, coding for a disintegrin and metalloproteinase domain-containing protein 10, and to *Abscam* gene, coding for a Dscam family member Abscam protein, respectively. The last gene has been reported to be involved in the domestication of *A. mellifera*. The third QTL is located 13 kb upstream of *LOC102655631*, coding for a trehalose transporter. For gentleness, two QTL were identified on chromosomes 4 and 3. They are located within gene *LOC413669*, coding for a lap4 protein, and gene *LOC413416*, coding for a bicaudal C homolog 1-B protein, respectively. The identified positional candidate genes of both traits mainly affect the olfaction and nervous system of honey bees. Further research is needed to confirm the results and to better understand the genetic and phenotypic basis of calmness and gentleness.

Keywords Abscam, Apis mellifera, GWAS, lap4 protein, pool sequences

Introduction

Honey bee behaviour is a routinely applied selected criterion in beekeeping (Uzunov *et al.*, 2017; Guichard *et al.*, 2020). To date, two commonly applied traits to assess the behaviour of honey bees are calmness, i.e. the extent to which bees remain quiet on the brood comb during colony inspection, and gentleness, which is an estimator for colony defensive behaviour (Büchler *et al.*, 2013). The advantages of calm and gentle honey bee colonies are manifold as they are easier to inspect, prevent beekeepers from getting stung, support the rapid location of the queen on the brood comb, and limit the risk of harming the queen during brood frame manipulations. The latter characteristics are especially important for small and endangered honey bee populations in order to maintain genetic diversity (Guichard *et al.*, 2019). Beekeepers and experts use different protocols to assess the behaviour of honey bees (Guzmán-Novoa *et al.*, 2003; Büchler *et al.*, 2013). Population genetic analyses of different honey bee populations have shown that the heritability estimates of calmness and gentleness can vary between low (0.02) and moderate (0.51) values (Collins *et al.*, 1984; Moritz *et al.*, 1987; Brascamp *et al.*, 2016; Andonov *et al.*, 2019; Guichard *et al.*, 2020), whilst the two traits can be highly correlated (e.g. a genetic correlation of 0.91 in an Austrian Carnica population) (Brascamp *et al.*,

Address for correspondence

Matthieu Guichard, Agroscope, Swiss Bee Research Centre, Schwarzenburgstrasse 161, 3003 Bern, Switzerland.
E-mail: matthieu.guichard@agroscope.admin.ch

Accepted for publication 11 April 2021

2016). Despite the release of the honey bee reference genome in 2006 (Honeybee Genome Sequencing Consortium, 2006), little knowledge on the genetic background of calmness and gentleness is currently available. A recent study based on whole-genome sequences of 37 *Apis mellifera carnica* drones identified 1035 genome-wide single nucleotide polymorphisms (SNPs) with a putative effect on gentleness (Jones *et al*., 2020). However, to the best of our knowledge, no quantitative trait loci (QTL) were reported for calmness and gentleness, whilst several QTL were reported for stinging behaviour (Arechavaleta-Velasco *et al*., 2003).

Stinging behaviour, which is also part of the gentleness assessment, is triggered by an alarm pheromone, an odour compound emitted by workers when the colony is disturbed (Breed *et al*., 2004). The response to this pheromone leads to an elevation of the worker's metabolism, which is especially strong in very aggressive colonies (Southwick & Moritz, 1985; Andere *et al*., 2002). This defensive behaviour is expressed by a fraction of the workers (Moore *et al*., 1987), originating from specialised patrilines (Robinson & Page, 1988; Breed *et al*., 1990). Stinging behaviour has been extensively studied in North American raised and managed Africanised and European honey bees. To date, three QTL—*sting-1*, *sting-2*, and *sting-3*—were reported on chromosomes 2, 3, and 7 (Elsik *et al*., 2016), respectively, including genes affecting the activity of the nervous system and sensory signalling (Hunt *et al*., 1998; Hunt *et al*., 2007). For the defensive behaviour of honey bees (including stinging and guarding), several candidate genes were identified (Guzman-Novoa *et al*., 2002; Arechavaleta-Velasco *et al*., 2003; Galindo-Cardona *et al*., 2013). Gene expression analyses of candidate genes showed that some of these genes are up- and down-regulated by the alarm pheromone (Alaux *et al*., 2009), whilst simultaneously the alarm pheromone production is controlled by other genes (Hunt *et al*., 1999). Therefore, these results demonstrate that the defensive behaviour of honey bees involves a complex interplay of genes affecting alarm pheromone production and the nervous system.

This research study aimed to identify QTL associated with calmness and gentleness to enhance the selection of the native *Apis mellifera mellifera* in Switzerland (Guichard *et al*., 2020) and to better understand the genetic architecture of currently applied selection traits in honey bees.

Materials and methods

Data sampling

Between 2016 and 2019, we sampled the phenotype and genotype information of 244 honey bee colonies, the majority of the samples belonging to *A. m. mellifera* (MEL) from Switzerland. Our data collection included 146 MEL colonies belonging to a selection programme (SL_CH), sampled at different test apiaries in Switzerland, as well as 70 MEL colonies from two conservation areas located in Switzerland (CS_CH, 45 colonies) and France (CS_FR, 25 colonies). We also included 28 *A. m. carnica* colonies (CAR) to identify putative outliers within our MEL sample and to verify the identified QTL in another honey bee subpopulation. For each colony, we collected approximately 500 workers using a standardised sampling method: the volume corresponding to 500 workers had previously been determined and was used to select a jar of corresponding capacity for sampling. This sampling aimed to statistically include all existing patrilines among workers in the colony as rare patrilines can affect the phenotype (Robinson & Page, 1988), as well as to limit the effects of drift and robbing on sampling. The sampling size of 500 workers was also chosen in order to minimize the proportion of individual contributions: some slightly larger bees could produce more DNA, for instance. Worker samples were stored at −20°C until DNA extraction.

Phenotyping

The sampled colonies were evaluated once for calmness and gentleness. Experts used a standardised scoring system (Büchler *et al*., 2013) to correctly assess the behaviour of colonies. Briefly, the scoring system utilises numerical scores between 1 and 4; the minimum and maximum scores are used to characterise extremely nervous/defensive (1) and calm/gentle (4) colonies, respectively. Colonies expressing a behaviour between the two extremes are accordingly rated with a score of 2 or 3. To discriminate between colonies expressing a very similar level of behaviour, it was also possible to use half- and quarter-points. In total, eight experts assessed the calmness and gentleness of the 244 colonies at 22 apiaries (18 in Switzerland and 4 in France) using different hive types (Dadant Blatt, Swiss hive, and others), whilst at each apiary, all colonies were kept in a single hive type and evaluated by a single expert. At some apiaries, calmness and gentleness could be measured several times; for these colonies, the mean value of the multiple observations, supposed to better fit the genetic value of the colony (Brascamp *et al*., 2016), was used for the analyses. Significant subpopulation effects on observations were identified by ANOVA followed by a Tukey multiple comparison of means with a 95% confidence interval. We used a simple linear regression model including year, expert, hive type, and apiary to determine by ANOVA the significant effects on calmness and gentleness. Furthermore, we calculated a phenotypic correlation between the two traits corrected for the apiary effect.

DNA extraction

The approximately 500 workers per colony were shredded in a DNA extraction solution containing 4 M urea, 10 mM

Tris-HCl pH 8, 300 mM NaCl, 10 mM EDTA, using a blender. The resulting mix was centrifuged for 15 min at 3500 ***g***, and 200 µl of supernatant was used for DNA extraction. 0.5 mg proteinase K and 15 µl of DTT 1 M were added before incubation overnight at 56 °C. The lysate is then used for DNA extraction with a QIAsymphony instrument and DNA Mini Kit (Qiagen). The final elution volume is 100 µl. Pair-end sequencing was performed on a IlluminaTM HiSeq 3000 or a NovaSeq 6000 platform. The number of samples per lane was optimised to target approximately 30× raw sequencing data per sample.

Sequence alignment

Raw reads from pool sequencing of the 244 colonies were aligned to the honey bee reference genome Amel_HAV3.1, Genebank assembly accession GCA_003254395.2 (Wallberg *et al.*, 2019), using BWA-MEM (v0.7.15; Li, 2013), and duplicates were marked with Picard (v2.18.2; https://broadinstitute.github.io/picard/, last accessed 5 March 2021). Samples sequenced in two runs were merged with SAMTOOLS (v1.8; Li *et al.*, 2009).

Pool sequence analysis

The subsequent analyses were restricted to approximately 7 million informative SNPs. This pre-selection of genome-wide SNPs was validated using 870 honey bee drones from France and other western European countries (A. Vignal, personal communication) as low-frequency SNPs can be identified more easily in haploid sequence data. With the application of the reduced SNP panel, it was possible to significantly increase computing time without ignoring low-frequency SNPs.

After the alignment, the resulting BAM files were converted into pileup files using the samtools mpileup utility (Li *et al.*, 2009) using the following parameters: coefficient of 50 for downgrading mapping quality for reads with excessive mismatches (-C 50), minimum mapping quality of 20 for an alignment (-q 20), and minimum base quality of 20 (-Q 20), following standard protocols. Files produced by mpileup were interpreted by the PoPoolation2 utility mpileup2sync (Kofler *et al.*, 2011) for the Sanger Fastq format, with a minimum quality of 20. Finally, sync files were converted to a depth file containing a sequencing depth value for each SNP and count files summarising reference and alternative allele counts for each SNP, using a custom-made Python script (van Rossum & Drake, 2009).

Allele frequencies and quality filtering

Based on the aforementioned count files including 7 023 977 SNPs, we removed 99 555 SNPs with multiple alternative alleles and 207 904 SNPs with an excessively high and low sequencing depth. After this quality control, we calculated the frequencies of the reference and alternative alleles for 6 716 518 SNPs and additionally excluded 820 141 homozygous loci. The remaining 4 554 843 SNPs were summarised in PLINK dosage and map files and further edited for minor allelic frequency above 5% and extreme deviation from the Hardy–Weinberg equilibrium using the INFO metric as implemented in PLINK (Purcell *et al.*, 2007). The final 1 389 840 genome-wide SNPs were used for population structure and genome-wide association analyses.

Population structure analysis

To illustrate the population structure of the colonies and to identify putative outliers within MEL samples, we carried out a principal component analysis (PCA) based on the frequency of the reference allele with R (R-Core-Team, 2018). The first and second principal components (PCs) were used for the graphical presentation of the population structure and included as covariates in the GWAS to account for population stratification.

Genome-wide association studies and candidate gene mapping

Genome-wide association studies were performed on SNP dosage data using a linear regression model as implemented in PLINK (Purcell *et al.*, 2007). We adjusted the model for covariates capturing population stratification (PCA) and significant effects (e.g. apiary) on the traits. These effects were identified by ANOVA on a linear regression model containing apiary, hive type, evaluator, and year. Significant associated SNPs were determined based on a 1% genome-wide Bonferroni-corrected threshold. The GWAS results were visualised using Manhattan and quantile–quantile plots with the R package *qqman* (Turner, 2014). Furthermore, the effect and the allele frequency of the best-associated SNP for each trait and subpopulation (including CAR) were calculated and visualised. Significant subpopulation effects on allele frequencies were identified by ANOVA followed by a Tukey multiple comparison of means with a 95% confidence interval. Genes within the identified QTL regions were determined using the NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_003254395.2) and the reference genome assembly Amel_HAv3.1 (Wallberg *et al.*, 2019).

Results

Calmness and gentleness assessments of the sampled colonies

For calmness, CS_FR colonies had the highest mean uncorrected score, followed by CAR, SL_CH, and CS_CH

(Fig. 1a). The difference between the three MEL subpopulations was statistically significant ($P < 0.05$, ANOVA), whilst CAR was only significantly different from CS_CH ($P < 0.05$, Tukey multiple comparison of means). Detailed results of the statistical tests are provided as Supplementary Material (File S1, SM 1A). For gentleness, almost the opposite result was observed as CS_CH colonies showed the highest and CS_FR the lowest uncorrected scores, respectively (Fig. 1b). Compared to calmness, only the difference between CS_CH and the other subpopulations was statistically significant ($P < 0.05$, Tukey multiple comparison of means; File S1, SM 1B). According to the contradictory observations, the phenotypic correlation between the two traits was quite low ($r = 0.39$, standard error = 0.06).

Population stratification of the sampled colonies

The first principal component (PC1), accounting for 98% of the total variance, clearly separated the CAR from MEL colonies and simultaneously highlighted the existence of five outliers within the MEL colonies (one SL_CH and four CS_FR; Fig. 2). The second principal component (PC2), accounting for 1% of the total variance, illustrated that CS_FR gradually overlapped with SL_CH colonies, whilst CS_CH colonies tightly clustered. After this result, the five MEL outliers were removed from the dataset, and the PCA was re-calculated without CAR colonies to account for the population stratification of the 211 purebred MEL colonies in the GWAS analyses (not shown). The GWAS analyses were performed on 1 326 634 genome-wide SNPs.

Calmness association

The GWAS analysis on calmness was adjusted for population stratification using the aforementioned first two PCs, accounting for 75% of the total variance (PC1 = 65%, PC2 = 10%), and four covariates showing a significant effect on the trait (year, apiary, hive type, and expert; File S1, SM 2A). After this adjustment, calmness was significantly ($P < 0.01$) associated with three QTL on chromosomes 6, 8, and 12 (Fig. 3a). The detailed list of the 10 best associated SNPs can be found in Table S1, SM 4. The best-associated QTL on chromosome 8 at 750 884 base pairs (bp) refers to an intron of the *LOC409692* gene. The second QTL on chromosome 6 at 1 584 897 bp is located in the intron of the *Abscam* gene, whilst the third QTL on chromosome 12 at 604 143 bp is not embedded in a gene region. The nearest gene, *LOC102655631*, is located 13 kb upstream of the QTL. Dividing the colonies into two groups according to their uncorrected phenotypes, according to the allelic frequency of the best-associated SNP on chromosome 8, showed that colonies segregating the T allele at high frequency (>50%) were calmer, whilst the majority of colonies carried the T allele at low frequency (<50%) and were less calm (Fig. 3b). Furthermore, it can be seen that some colonies with a low T allele frequency were assessed with a maximum score of 4 (very calm). In general, the associated T allele segregated at low frequency in all subpopulations, whereas on average CS_CH showed the highest and CAR the lowest frequencies, respectively (Fig. 3c, File S1, SM 3A).

Gentleness association

For the GWAS analysis on gentleness, the same number of covariates were included in the model whilst the hive type was removed (File S1, SM 2B). Gentleness was significantly ($P < 0.01$) associated with two QTL on chromosomes 3 and 4 (Fig. 4a). The detailed list of the 10 best associated SNPs can be found in Table S1, SM 4. The best-associated QTL, on chromosome 4 at 3 378 436 bp, corresponds to an intron of the gene *LOC413669*, whilst the second QTL, on

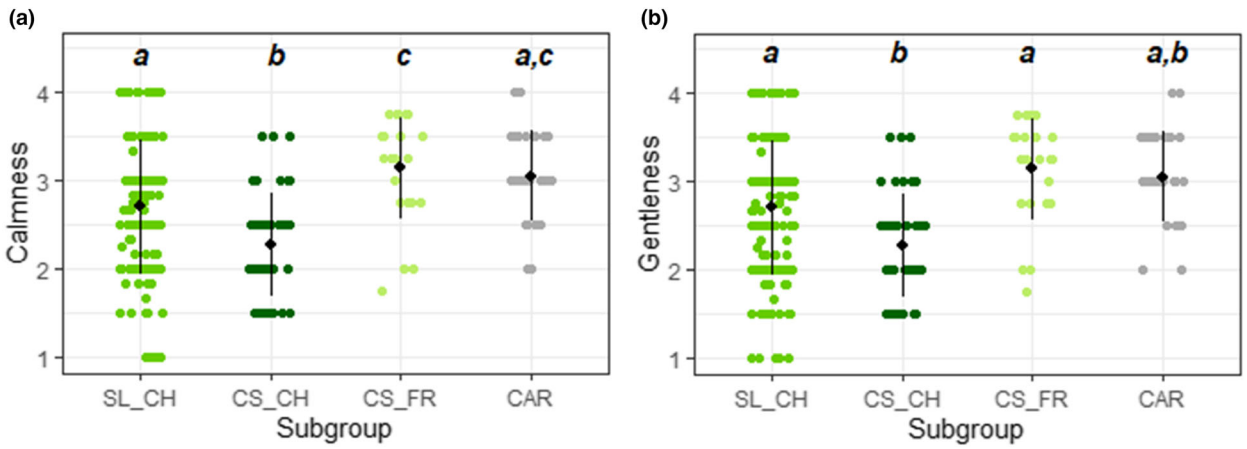

Figure 1 Jitterplots of uncorrected observations for calmness (a) and gentleness (b) according to subgroup. Black rhombus corresponds to mean value, bars correspond to standard deviation. Different letters indicate significant ($P < 0.05$) differences between groups following a Tukey multiple comparison of means with a 95% confidence interval

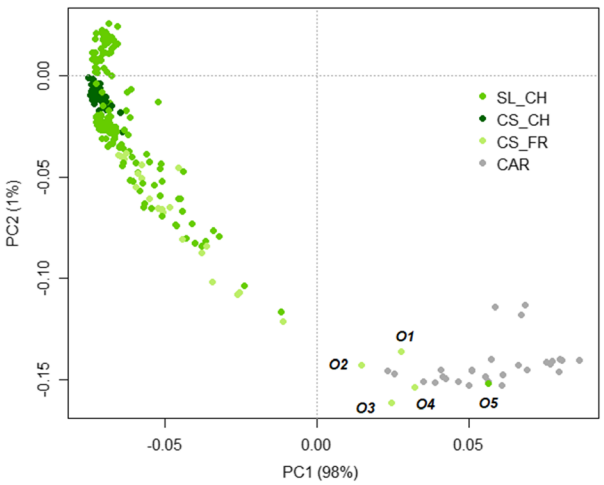

Figure 2 Principal component (PC) analysis of the sampled colonies, according to subgroup. *Apis mellifera mellifera* (MEL) outliers (probably admixed colonies) clustering with the *Apis mellifera carnica* (CAR) colonies (PC1 > 0) are identified (O1–O5) and removed from GWAS

chromosome 3 at 1 483 128 bp, is located in an intron of the gene *LOC413416*. The allele frequency of the best associated SNP demonstrated that colonies with a high G allele frequency (>50%) were gentler compared to those carrying the G allele at low frequency (<50%; Fig. 4b). Compared to calmness, the colonies were better distributed according to their uncorrected phenotypes over the two frequency groups, whereas a few colonies with a low G allele frequency were assessed to be very gentle (maximum score of 4). The G allele frequency of the different subpopulations revealed that the occurrence of this allele is significantly different between MEL and CAR, as well as between CS_CH and the remaining MEL subpopulations (SEL_CH and CS_FR), whilst the associated G allele is highly segregated within CS_CH (Fig. 4c, File S1, SM 3B).

Discussion

We conducted GWAS analyses for calmness and gentleness using pool sequences of 211 *A. m. mellifera* colonies. These analyses led to the identification of three QTL associated with calmness and two QTL associated with gentleness, respectively. Due to the small sample collection and the subjective scoring of the colonies, we cannot exclude the possibility that our current GWAS results are strongly biased. It would need to be confirmed with more colonies, also in different populations.

For calmness, the best-associated QTL on chromosome 8 is located in a gene (*LOC409692*) coding for the disintegrin and metalloproteinase domain-containing protein 10. Homologues of this protein are involved in many biological pathways, including neurogenesis in vertebrates and *Drosophila melanogaster*, where e.g. Kuzbanian plays a role in axonal extension (Fambrough *et al.*, 1996; Rooke *et al.*, 1996; Pan & Rubin, 1997; Chen *et al.*, 2007). Furthermore, this family of proteins affect different cellular functions through shedding (i.e. the release of transmembrane molecules from the cell surface (Reiss & Saftig, 2009)). According to the described functions in other species, this protein might also affect the development of the nervous system in the honey bee.

The second identified QTL, on chromosome 6, corresponds to the *Abscam* gene, coding a protein of the Dscam family. Dscam proteins are well known to be associated with neuronal wiring in *Drosophila melanogaster* and vertebrates (Schmucker *et al.*, 2000; Hattori *et al.*, 2008), whilst in *Drosophila* these proteins are also involved in axonal targeting of olfactory receptors (Hummel *et al.*, 2003; Zhu *et al.*, 2006). In the honey bee, the *Abscam* gene is expressed in the optic and antennal lobes, which integrate visual and olfactory information, especially at the level of olfactory neuron axons (Funada *et al.*, 2007). Furthermore, it has been demonstrated that this gene participates in neuronal connections during honey bee development (Funada *et al.*, 2007).

The third QTL, on chromosome 12, is located 13 kb upstream from a gene (*LOC102655631*) coding for a trehalose transporter (facilitated trehalose transporter Tret1-like). Trehalose is known to be a key carbohydrate synthesized during honey bee larval development, especially during larval stages (Farjan *et al.*, 2015; Łopieńska-Biernat *et al.*, 2018). In *D. melanogaster*, trehalose is synthesized by neurosecretory neurons (Miyamoto & Amrein, 2019) and is particularly known to confer developmental robustness (Matsushita & Nishimura, 2020). In *D. melanogaster*, *Tret* genes are particularly expressed in the central nervous system, and mutant flies unable to produce trehalose die due to local failure in the central nervous system (Matsuda *et al.*, 2015).

For gentleness, the best-associated QTL on chromosome 4 refers to *LOC413669*, a gene coding a protein called lap4. This protein is involved in many cellular growth and differentiation processes, whilst it also affects olfactory behaviour in *D. melanogaster* (Anholt *et al.*, 1996; Ganguly *et al.*, 2003). The second QTL on chromosome 3 is located in a gene (*LOC413416)* coding a protein called bicaudal C homologue 1-B. In *Megachile rotundata*, a bicaudal C coding gene was associated with diapause development (Yocum *et al.*, 2018), whilst in Asian honey bees (*Apis cerana cerana*) this protein affects cold resistance (Xu *et al.*, 2017).

None of the previously reported candidate genes for stinging behaviour were located in the reported QTL regions (Hunt *et al.*, 1998; Hunt *et al.*, 2007; Elsik *et al.*, 2016). Thus, the detected QTL suggests that other major genes affect the genetics of gentleness, even though these traits also assess the defensive behaviour of honey bees. Compared to previous studies, calmness and gentleness did not show a high phenotypic correlation. Nevertheless, we identified a common genetic background associated with olfaction in

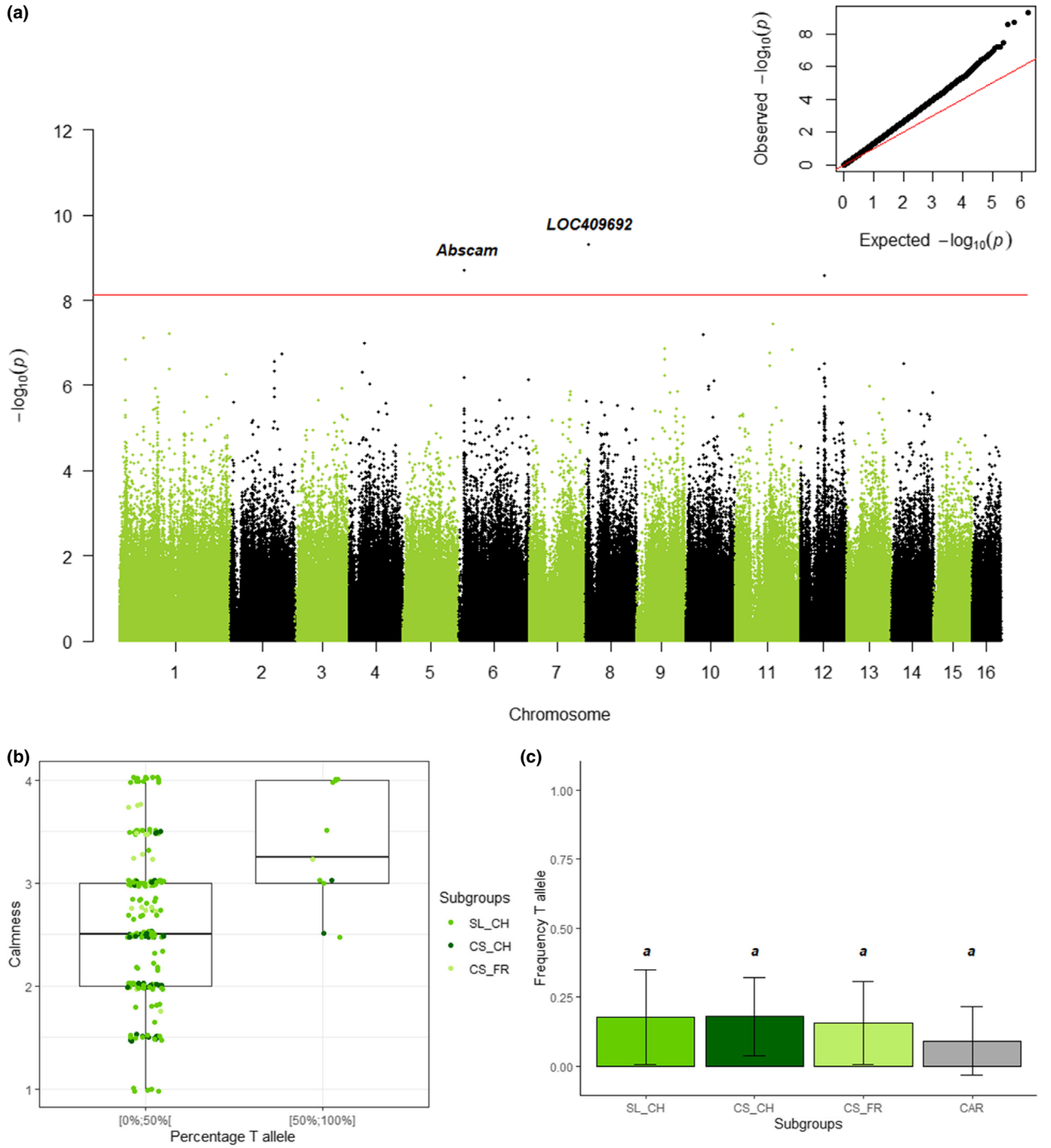

Figure 3 GWAS for calmness. (a) Manhattan plot and quantile–quantile plots for calmness ($N = 211$ *Apis mellifera mellifera* [MEL] colonies, outliers removed). The red line is the threshold for SNPs having a significant ($P < 0.01$) effect on phenotype. Three SNPs have a highly significant effect. The best SNP located at 750 884 base pairs on chromosome 8 corresponds to the *LOC409692* gene. One other significant SNP located at 1 584 897 base pairs on chromosome 6, is situated in the *Abscam* gene. A third significant SNP exists at 6 041 432 base pairs on chromosome 12 but does not correspond to a gene. (b) Calmness (uncorrected phenotype) according to percentage of T allele of the best SNP (chromosome 8). (c) Mean percentage of T allele for best SNP and associated standard deviation in each subgroup. Different letters indicate significant ($P < 0.05$) differences between groups following a Tukey multiple comparison of means with a 95% confidence interval

honey bees. Therefore, our results suggest that for both traits, olfactory signals play an important role and are useful in recognising alarm and brood pheromones in the colony as well as external odours (e.g. intruders).

Around 7000 years ago, humans started to keep honey bees in artificial hives. Since that time, the honey bee genome has been exposed to human-mediated selection. Population genomic analyses of the honey bee have

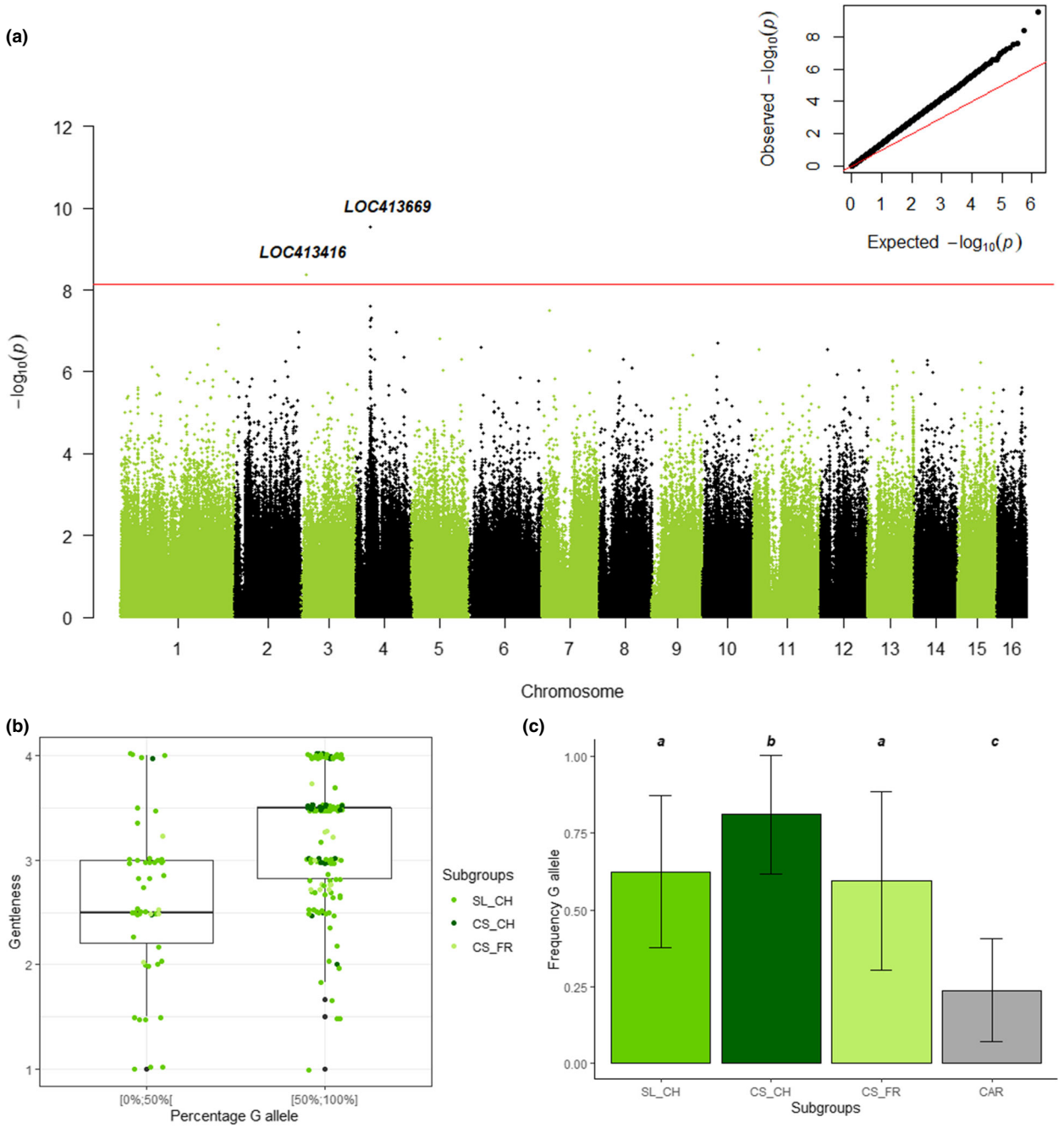

Figure 4 GWAS for gentleness. (a) Manhattan plot and quantile–quantile plots for gentleness ($N = 211$ *Apis mellifera mellifera* [MEL] colonies, outliers removed). The red line is the threshold for SNPs having a significant ($P < 0.01$) effect on phenotype. The best SNP, located at 3 378 436 base pairs on chromosome 4, corresponds to gene *LOC413669*. Another significant SNP, located at 1,483,128 base pairs on chromosome 3, corresponds to gene *LOC413416*. (b) Gentleness (uncorrected phenotype) according to percentage of G allele of the best SNP (chromosome 4). (c) Mean percentage of G allele for best SNP and associated standard deviation in each subgroup. Different letters indicate significant ($P < 0.05$) differences between groups following a Tukey multiple comparison of means with a 95% confidence interval

already revealed strong signatures of positive selection on currently applied selection traits (e.g. royal jelly production; Rizwan *et al.*, 2020). In this study, we associated the *Abscam* gene with calmness. Recently, this gene was also identified to be under selection by comparing modern and historic *A. m. mellifera* samples from Switzerland (Parejo *et al.*, 2020). Therefore, we assume that behavioural-associated genes in particular are good candidates to investigate the domestication of *A. mellifera*. However, due to the low frequency of the associated allele in all sampled

subpopulations, we were not able to show a meaningful selection in *A. mellifera*.

Conclusions

We identified five QTL associated with calmness and gentleness using whole-genome sequences. The positional candidate genes mainly affect the olfaction and nervous system of honey bees. Candidate genes for behaviour might be useful in investigating the domestication of *A. mellifera*. Further research is needed to confirm the results and to better understand the molecular and phenotypic basis of the two traits. We believe that our results might be used in the future to implement marker-assisted selection in *A. m. mellifera* to select calm and gentle honey bees.

Acknowledgements

Financial support for this study was provided by Bundesamt für Landwirtschaft BLW (Swiss Federal Office for Agriculture FOAG), grant No. 627000708, and by Labogena and FranceAgriMer (Programme d'Investissements d'Avenir). We thank Lucie Genestout (Labogena) for coordinating the Beestrong project, Christina Sann and Yannick Poquet (Labogena) for organising the sampling in Savoie, and Alice Revel (Agroscope) for her help in organising the sampling in Switzerland. We thank Rachid Mahla (Labogena) for DNA extraction and Olivier Bouchez (GeT-PlaGe) for sequencing. Sequencing was performed in collaboration with the GeT platform, Toulouse (France), a partner of the National Infrastructure France Genomique, thanks to support by the Commissariat aux Grands Investissements (ANR-10-INBS-0009). Sequence alignments and allele frequencies estimations were performed on the computer cluster of the Genotoul bioinformatics platform Toulouse Midi-Pyrénées (www.bioinfo.genotoul.fr). We are grateful to all of the experts for colony evaluation and to all of the beekeepers for supporting the sample collection.

This study was performed in collaboration with the Beestrong Consortium.

Beestrong Consortium: Sonia Eynard, Alain Vignal, Bertrand Servin (INRAE, GenPhyse, 31326 Castanet-Tolosan, France), Lucie Genestout, Rachid Mahla, Christina Sann, Yannick Poquet, Yves Agez (LABOGENA DNA, Laboratoire d'Analyses Génétiques pour les Espèces Animales, Domaine de Vilvert, 78353 Jouy-en-Josas, France), François Guillaume (Evolution XY, rue Eric Tabarly, 35538 Noyal-Sur-Vilaine, France), Benjamin Basso, Yves Le Conte, Fanny Mondet (INRAE, Abeilles et Environnement, Avignon, France & UMT PrADE, Avignon, France), Benjamin Basso, Axel Decourtye (ITSAP, Avignon, France), Florence Phocas, Jean-Pierre Bidanel (Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France), Olivier Bouchez (GeT-PlaGe, Genotoul, INRAE Auzeville, F-31320, Castanet-Tolosan, France).

Conflict of interest statement

The authors declare no conflicts of interest.

Data availability statement

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

References

Alaux C., Sinha S., Hasadsri L., Hunt G.J., Guzman-Novoa E., Degrandi-Hoffman G., Uribe-Rubio J.L., Southey B.R., Rodriguez-Zas S. & Robinson G.E. (2009) Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15400–5.

Andere C., Palacio M.A., Rodriguez E.M., Figini E., Dominguez M.T. & Bedascarrasbure E. (2002) Evaluation of the defensive behavior of two honeybee ecotypes using a laboratory test. *Genetics and Molecular Biology* **25**, 57–60.

Andonov S., Costa C., Uzunov A., Bergomi P., Lourenco D. & Misztal I. (2019) Modeling honey yield, defensive and swarming behaviors of Italian honey bees (*Apis mellifera ligustica*) using linear-threshold approaches. *BMC Genetics* **20**, 78.

Anholt RRH, Lyman R.F. & Mackay TFC (1996) Effects of single P-element insertions on olfactory behavior in *Drosophila melanogaster*. *Genetics* **143**, 293–301.

Arechavaleta-Velasco M.E., Hunt G.J. & Emore C. (2003) Quantitative trait loci that influence the expression of guarding and stinging behaviors of individual honey bees. *Behavior Genetics* **33**, 357–64.

Brascamp E.W., Willam A., Boigenzahn C., Bijma P. & Veerkamp R.F. (2016) Heritabilities and genetic correlations for honey yield, gentleness, calmness and swarming behaviour in Austrian honey bees. *Apidologie* **47**, 739–48.

Breed M.D., Guzman-Novoa E. & Hunt G.J. (2004) Defensive behavior of honey bees: organization, genetics, and comparisons with other bees. *Annual Review of Entomology* **49**, 271–98.

Breed M.D., Robinson G.E. & Page R.E. (1990) Division of labor during honey bee colony defense. *Behavioral Ecology and Sociobiology* **27**, 395–401.

Büchler R., Andonov S., Bienefeld K., Costa C., Hatjina F., Kezic N., Kryger P., Spivak M., Uzunov A. & Wilde J. (2013) Standard methods for rearing and selection of *Apis mellifera* queens. *Journal of Apicultural Research* **52**.

Chen Y.Y., Hehr C.L., Atkinson-Leadbeater K., Hocking J.C. & McFarlane S. (2007) Targeting of retinal axons requires the metalloproteinase ADAM10. *Journal of Neuroscience* **27**, 8448–56.

Collins A.M., Rinderer T.E., Harbo J.R. & Brown M.A. (1984) Heritabilities and correlations for several characters in the honey bee. *Journal of Heredity* **75**, 135–40.

Elsik C., Tayal A., Diesh M.C., Unni D., Emery M., Nguyen H. & Hagen D.E. (2016) Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Research* **44**, D793–D800.

Fambrough D., Pan D., Rubin G.M. & Goodman C.S. (1996) The cell surface metalloprotease/disintegrin Kuzbanian is required for axonal extension in *Drosophila*. *Proceedings of the National Academy of Sciences* **93**, 13233–8.

Farjan M., Żółtowska K., Lipiński Z., Łopieńska-Biernat E. & Dmitryjuk M. (2015) The effect of dietary vitamin c on carbohydrate concentrations and hydrolase activity, during the development of honey bee worker brood. *Journal of Apicultural Science* **59**, 5–16.

Funada M., Hara H., Sasagawa H., Kitagawa Y. & Kadowaki T. (2007) A honey bee Dscam family member, AbsCAM, is a brain-specific cell adhesion molecule with the neurite outgrowth activity which influences neuronal wiring during development. *European Journal of Neuroscience* **25**, 168–80.

Galindo-Cardona A., Acevedo-Gonzalez J.P., Rivera-Marchand B. & Giray T. (2013) Genetic structure of the gentle Africanized honey bee population (gAHB) in Puerto Rico. *BMC Genetics* **14**, 65.

Ganguly I., Mackay TFC & Anholt RRH (2003) *Scribble* is essential for olfactory behavior in *Drosophila melanogaster*. *Genetics* **164**, 1447–57.

Guichard M., Neuditschko M., Fried P., Soland G. & Dainat B. (2019) A future resistance breeding strategy against *Varroa destructor* in a small population of the dark honey bee. *Journal of Apicultural Research* **58**, 814–23.

Guichard M., Neuditschko M., Soland G., Fried P., Grandjean M., Gerster S., Dainat B., Bijma P. & Brascamp E.W. (2020) Estimates of genetic parameters for production, behaviour, and health traits in two Swiss honey bee populations. *Apidologie* **51**, 876–891.

Guzman-Novoa E., Hunt G.J., Uribe J.L., Smith C. & Arechavaleta-Velasco M.E. (2002) Confirmation of QTL effects and evidence of genetic dominance of honeybee defensive behavior: Results of colony and individual behavioral assays. *Behavior Genetics* **32**, 95–102.

Guzmán-Novoa E., Prieto-Merlos D., Uribe-Rubio J.L. & Hunt G.J. (2003) Relative reliability of four field assays to test defensive behaviour of honey bees (*Apis mellifera*). *Journal of Apicultural Research* **42**, 42–6.

Hattori D., Millard S., Wojtowicz W. & Zipursky L. (2008) Dscam-mediated cell recognition regulates neural circuit formation. *Annual Review of Cell and Developmental Biology* **24**, 597–620.

Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–49.

Hummel T., Vasconcelos M.L., Clemens J.C., Fishilevich Y., Vosshall L.B. & Zipursky L. (2003) Axonal targeting of olfactory receptor neurons in *Drosophila* is controlled by Dscam. *Neuron* **37**, 221–31.

Hunt G.J., Amdam G.V., Schlipalius D. *et al.* (2007) Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften* **94**, 247–67.

Hunt G.J., Collins A.M., Rivera R., Page R.E. & Guzman-Novoa E. (1999) Quantitative trait loci influencing honeybee alarm pheromone levels. *Journal of Heredity* **90**, 585–9.

Hunt G.J., Guzman-Novoa E., Fondrk M.K. & Page R.E. (1998) Quantitative trait loci for honey bee stinging behavior and body size. *Genetics* **148**, 1203–13.

Jones J.C., Du Z.G., Bernstein R. *et al.* (2020) Tool for genomic selection and breeding to evolutionary adaptation: Development of a 100K single nucleotide polymorphism array for the honey bee. *Ecology and Evolution* **10**, 6246–56.

Kofler R., Pandey R.V. & Schlötterer C. (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–6.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. & Subgroup GPDP (2009) The Seqence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–9.

Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.

Łopieńska-Biernat E., Żółtowska K., Zaobidna E.A., Dmitryjuk M. & Bąk B. (2018) Developmental changes in gene expression and enzyme activities of anabolic and catabolic enzymes for storage carbohydrates in the honeybee, *Apis mellifera*. *Insectes Sociaux* **65**, 571–80.

Matsuda H., Yamada T., Yoshida M. & Nishimura T. (2015) Flies without trehalose. *Journal of Biological Chemistry* **290**, 1244–55.

Matsushita R. & Nishimura T. (2020) Trehalose metabolism confers developmental robustness and stability in *Drosophila* by regulating glucose homeostasis. *Communications Biology* **3**.

Miyamoto T. & Amrein H. (2019) Neuronal gluconeogenosis regulates systemic glucose homeostasis in *Drosophila melanogaster*. *Current Biology* **29**, 1263–72.

Moore A.J., Breed M.D. & Moor M.J. (1987) The guard honeybee: ontogeny and behavioural variability of workers performing a specialized task. *Animal Behaviour* **35**, 1159–67.

Moritz RFA, Southwick E.E. & Harbo J.R. (1987) Genetic analysis of defensive behaviour of honeybee colonies *Apis mellifera* L. in a field test. *Apidologie* **18**, 27–42.

Pan D. & Rubin G.M. (1997) Kuzbanian controls proteolytic processing of Notch and mediates lateral inhibition during *Drosophila* and vertebrate neurogenesis. *Cell* **90**, 271–80.

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

Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* **81**, 559–75.

R-CORE-TEAM (2018) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.R-project.org/

Reiss K. & Saftig P. (2009) The "a disintegrin and metalloprotease" (ADAM) family of sheddases: physiological and cellular functions. *Seminars in Cell & Developmental Biology* **20**, 126–37.

Rizwan M, Liang P, Ali H *et al.* (2020) Population genomics of honey bees reveals a selection signature indispensable for royal jelly production. *Molecular and Cellular Probes* **52**, 101542.

Robinson G.E. & Page R.E. (1988) Genetic determination of guarding and undertaking in honeybee colonies. *Nature* **333**, 356–8.

Rooke J., Pan D., Xu T. & Rubin G.M. (1996) KUZ, a conserved metalloprotease-disintegrin protein with two roles in *Drosophila* neurogenesis. *Science* **273**, 1227–31.

Schmucker D., Clemens J.C., Shu H., Worby C.A., Xiao J., Muda M., Dixon J.E. & Zipursky L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–84.

Southwick E.E. & Moritz RFA (1985) Metabolic response to alarm pheromone in honey bees. *Journal of Insect Physiology* **31**, 389–92.

Turner S.D. (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. BioRxiv.

Uzunov A., Brascamp E.W. & Büchler R. (2017) The basic concept of honey bee breeding programs. *Bee World* **94**, 84–7.

van Rossum G. & Drake F.L. (2009) *Python 3 Reference Manual*. CA, CreateSpace: Scotts Valley.

Wallberg A, Bunikis I, Pettersson OV, Mosbech MB, Childers AK, Evans JD, Mikheyev AS, Robertson HM, Robinson GE & Webster MT (2019). A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics* **20**, 275.

Xu K., Niu Q., Zhao H., Du Y. & Jiang Y. (2017) Transcriptomic analysis to uncover genes affecting cold resistance in the Chinese honey bee (*Apis cerana cerana*). *PLoS One* **12**, e0179922.

Yocum G.D., Childers A.K., Rinehart J.P., Rajamohan A., Pitts-Singer T.L., Greenlee K.J. & Bowsher J.H. (2018) Environmental history impacts gene expression during diapause development in the alfalfa leafcutting bee, *Megachile rotundata*. *Journal of Experimental Biology* **221**, jeb173443.

Zhu H., Hummel T., Clemens J.C., Berdnik D., Zipursky L. & Luo L. (2006) Dendritic patterning by Dscam and synaptic partner matching in the *Drosophila* antennal lobe. *Nature Neuroscience* **9**, 349–55.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.
Table S1 SNPs significant at 0.01 and 0.05 for calmness and gentleness association studies
File S1 Detailed results of statistical tests