621. Statistics for an accurate genome wide association study on *Varroa* resistance trait in a French honeybee

S.E. Eynard[1,2*], A. Vignal[2,3], Y. Agez[1], B. Basso[3,4], O. Bouchez[5], T. Bulach[5], Y. Le Conte[3,4], B. Dainat[6], A. Decourtye[3,7], L. Genestout[1], M. Guichard[6,8], F. Guillaume[9], E. Labarthe[2], R. Mahla[1], F. Mondet[3,4], M. Neudischko[8], F. Phocas[3,10], Y. Poquet[1], C. Sann[1], R.F. Serre[5], K. Tabet[2] and B. Servin[2,3]

[1]Labogena DNA, 78353 Jouy-en-Josas, France; [2]GenPhySe, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France; [3]UMT PrADE, France; [4]INRAE, Abeilles et Environnement, 84914 Avignon, France; [5]INRAE, GeT-PlaGe, Genotoul INRAE, 31326 Castanet-Tolosan, France; [6]Agroscope, 3003 Bern, Switzerland; [7]ITSAP, 84914 Avignon, France; [8]Agroscope, 1725 Posieux, Switzerland; [9]Evolution XY, 35538 Noyal-Sur-Vilaine, France; [10]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78352 Jouy-en-Josas, France; sonia.eynard@labogena.fr

Abstract

The honeybee *Apis mellifera*, major pollinator worldwide, is facing multiple threats to survival with the infestation by *Varroa destructor* being a major concern. Resistance to such infestation is a crucial mechanism for the maintenance of the honeybee population. However, to date, genetic components explaining this resistance are poorly known making selection for resistant honeybees a challenge. In this study we perform a genome wide association study on about 300 *A.m. mellifera* colonies for one trait measuring varroa resistance, the recapping of varroa infested brood cells. On about 300 *A.m. mellifera* colonies we identified multiple chromosome regions showing significant effects for this polygenic trait. We reported a heritability of 0.57 (±0.21). Recapping of varroa infested brood cells could be a trait of interest for the selection of resistant honeybees.

Introduction

The honeybee, *Apis mellifera*, is the main domestic pollinator worldwide. Despite its critical value for pollination services, production of honey, royal jelly and pollen, it is currently facing considerable challenges due to land anthropication (loss of flower resources and suitable habitats, pesticide use), climate change and the spread of biological threats. The ectoparasite *Varroa destructor* is one of these threats, as it is the principal cause of honeybee colony mortality in the Western world (Le Conte *et al.* 2010). For about a decade, resistance to varroa has become a key trait of interest for beekeepers. Colonies naturally resistant to varroa infestation have been identified by natural selection when untreated colonies were left to die or survive in isolated environments (Mondet *et al.* 2020). However, the genetic mechanisms to explain the variation in varroa infestation resistance observed across honeybee colonies is poorly known. Only a few genetic markers of interest have been identified in small honeybee populations (Mondet *et al.* 2020). Moreover, it is difficult to define a relevant measure to score resistance to a parasite and estimate its genetic parameters. Multiple traits have been described, for example mite none reproduction (MNR), varroa sensitive hygiene (VSH), hygienic behaviour, recapping, and very variable genetic parameters were estimated (Guichard *et al.* 2020). Identifying genetic markers linked to resistance to varroa infestation is key to understanding the genetics behind this resistance and in the long run, to perform genomic selection for this trait. In this study, about 300 French colonies of the *A.m. mellifera* genetic background, endemic to Western Europe, were sampled for genetic material and traits expected to be linked to resistance to varroa infestation and we present results from recapping of varroa infested brood cells by adult bees. A pool of workers from each colony was sequenced, providing sequencing depth and allele counts giving allele frequencies within the colony rather than traditional genotype information. In the context of eusocial insects, a method was developed to reconstruct queen genotype from such pool data (Eynard *et al.* 2021). We applied this method to reconstruct honeybee queen genotypes for each of the colonies which were

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_621, © S.E. Eynard *et al.* 2022

used in a Genome Wide Association Studies (GWAS) for the trait of interest. Its heritability was estimated and markers and genome regions with significant effects were identified. These results contribute to the knowledge on honeybee resistance to varroa infestation and will be valuable for future implementation of genomic selection.

Materials & methods

Genetic material and sequencing. A total of 306 honeybee colonies, with a major *A.m. mellifera* genetic background, coming from beekeepers in France were sampled within the framework of the FranceAgriMer, Investissement d'Avenir BeeStrong project. Each genetic sample consisted of about 500 honeybee workers. An initial DNA extraction, using tailor made protocols, was performed. Thereafter, pool sequencing was done in order to obtain 30X raw sequencing depth. These reads were aligned on the honeybee reference genome Amel HAV3.1 (Wallberg *et al.* 2019), using BWA-MEM (Li 2013). Pool sequences were analysed using Samtools mpileup (Li *et al.* 2009) and Popoolation2 (Kofler *et al.* 2011) on 6,914,704 single nucleotide polymorphisms identified in Wragg *et al.* (2021). Markers were filtered for minor allele frequency above 0.01 and missing rate below 5% for further analysis.

Queen genotype reconstruction. Honeybee queen genotypes for all markers and each colony were reconstructed from sequencing depth and allele counts data using the method of Eynard *et al.* (2021). In short, the method is based on the likelihood of the queen genotype that can be written as:

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

where g_l^c is the unknown honeybee queen genotype, f_l^c is the unknown reference allele frequency in the males founder of the colony and d_l^c and x_l^c the sequencing depth and allele counts obtained from pool sequencing experiments for locus l and colony c. Assuming all colonies come from an homogeneous population, Equation 1 can be expressed as a function of a common allele frequency f_l which can be estimated by maximum likelihood and used to compute posterior probabilities of the queen genotype.

Measure of resistance to varroa infestation. One way for honeybee colonies to control *Varroa* infestation is by recapping infested brood cells as this behaviour breaks the reproductive cycle of the mite. Recapping status was thus proposed as a measure of varroa resistance (Büchler *et al.* 2017) and therefore the proportion of recapped brood cells that are infested by varroa was used as a resistance trait in this study.

Genome wide association study. We performed a GWAS using a multivariate linear model as proposed in GEMMA (Zhou and Stephens 2012). The relationship matrix between colonies was estimated on allele frequencies for 2,387,893 markers from the pool sequencing experiment, taking SNP linkage disequilibrium into account using LDAK Speed *et al.* (2012). The association of the reconstructed queen genotypes with the phenotype was tested for 2,694,182 markers. Markers' *P*-values were obtained by GEMMA and q-values and s-values were estimated using the R package 'ashr' (Stephens 2017). SNPs with a s-value less than 0.1 were considered significant.

Results

The proportion of phenotypic variance explained (pve) by the markers, a proxy for heritability, for recapping of varroa infested brood cells was 0.57 (95% confidence interval [0.16, 0.98]). The GWAS inflation factor was 1, indicating an accurate correction for population stratification (Figure 1).

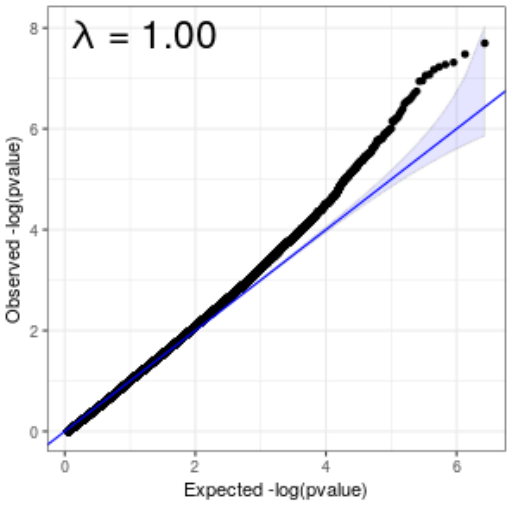

Figure 1. Quantile-quantile (QQ) plot of observed versus expected *P*-values for the GWAS with inflation factor, lambda.

We identified 18 chromosome regions significantly supported by at least one marker for both standard *P*-values and after adaptive shrinkage (Figure 2). These regions were found on 13 of the 16 chromosomes.

Discussion

Resistance to varroa infestation in honeybees is a complex trait to measure and thus far there is no consensus way to accurately measure it. One way to infer varroa resistance in a honeybee colony is to estimate the ratio of varroa infested brood cells that have been recapped. This is under the assumption that recapping of an infested brood cell by worker bees will cause a disturbance in the *Varroa destructor* reproductive cycle within the sealed brood cells and thus reducing its pressure on the colony by impairing its development (Oddie *et al.* 2018, Oddie *et al.* 2021). This trait relies highly on the initial infestation status, meaning that a colony with a low infestation rate will have a less accurate estimate of recapping than a

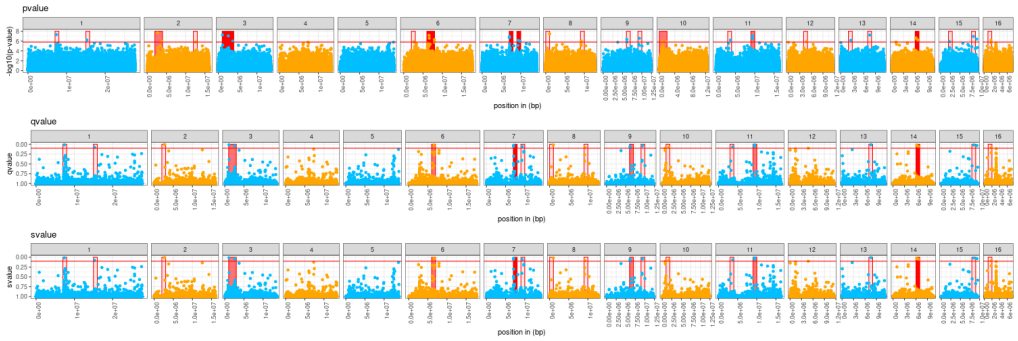

Figure 2. Manhattan plot of the GWAS results for -log10(*P*-value), q-value and s-value, estimated by adaptive shrinkage, along each of the chromosomes. Regions containing significant markers are highlighted in red, brightness of red being linked to the number of significant markers in the region. The red line marks the significance threshold.

highly infested colony, as the ratio will be based on a smaller number of observed brood cells. This causes a high variability between measures that need to be accounted for, which is done in our case by using an Empirical Bayes transformation giving weights to the phenotypes that are proportional to their estimated accuracy. Additionally, a logit transformation was needed to linearise the measured ratio and allow the use of this trait for GWAS. In our study, we identified 18 chromosome regions that appear to have a significant effect on the trait. Most chromosomes carry a region of interest meaning that the recapping of infested brood cells is a highly polygenic trait. Our results contrast with other studies in which a small number of significant markers have been observed (Guichard *et al.* 2021). We could estimate the heritability, as the phenotypic variability explained by our markers, for the trait to be 0.57 (±0.42). Although still imprecise, this estimate of heritability shows the potential there is to use this trait in a breeding program. It would now be interesting to identify genes underlying regions that appear to have a significant effect on the recapping of varroa infested brood cells in order to unravel the biological mechanisms explaining this behaviour in worker honeybees.

References

Büchler R., Costa C., Mondet F., Kezic N. and Kovacic M. (2017) Screening for low Varroa mite reproduction (SMR) and recapping in European honey bees.

Eynard S.E., Vignal A., Basso B., Le Conte Y., Decourtye A. *et al.* (2021) BioRxiv https://doi.org/10.1101/2021.11.08.467442

Guichard M., Dietemann V., Neuditschko M. and Dainat B. (2020) Genet Sel Evol 52:71 https://doi.org/10.1186/s12711-020-00591-1

Guichard M., Dainat B., Eynard S., Vignal A., Servin B. *et al.* (2021) Anim Genet https://doi.org/10.1111/age.13150

Kofler R., Pandey R.V. and Schlötterer C. (2011) Bioinformatics 27(24):3435-3436 https://doi.org/10.1093/bioinformatics/btr589

Le Conte Y., Ellis M. and Ritter W. (2010) Apidologie 41(3):353-363 https://doi.org/10.1051/apido/2010017

Li H., Handsaker B. Wysoker A., Fennell T., Ruan J. et al (2009) Bioinformatics 25:2078-2079 https://doi.org/10.1093/bioinformatics/btp352

Li H. (2013) ArXiv http://arxiv.org/abs/1303.3997

Mondet F., Beaurepaire A., McAfee A., Locke B., Alaux C. et al (2020) Int J for Parasit 50(6-7):433-437 https://doi.org/10.1016/j.ijpara.2020.03.005

Oddie M.A.Y., Büchler R., Dahle B., Le Conte Y., Locke B. et al (2018) Sci Rep 8:7704 https://doi.org/10.1038/s41598-018-26001-7

Oddie M.A.Y., Burke A., Dahle B., Le Conte Y., Mondet F. et al (2021) Sci Rep 11:9133 https://doi.org/10.1038/s41598-021-88592-y

Speed D., Hemani G., Johnson M.R. and Balding D. (2012) AJHG 91(6):1011-1021 https://doi.org/10.1016/j.ajhg.2012.10.010

R Core Team, R: A language and environment for statistical computing 2021. Available at: https://www.R-project.org/

Stephens M (2017) Biostat 18(2):275-294. https://doi.org/10.1093/biostatistics/kxw041

Wallberg A., Bunikis I., Pettersson O., Mosbech M., Childers A. et al (2019) BMC Genomics 20:275 https://doi.org/10.1186/s12864-019-5642-0

Wragg D., Eynard S., Basso B., Canale-Tabet K., Labarthe E. *et al.* (2021) BioRxiv https://doi.org/10.1101/2021.09.20.460798

Zhou X. and Stephens M. (2012) Nat Genet 44:821-824. https://doi.org/10.1038/ng.2310