scientific reports

Check for updates

OPEN

Establishing standards for genome wide association studies in eusocial insects through a case study in honey bees

Mulusew Fikere[1]✉, Dylan K. Ryals[1], Jana Obšteter[2,3], Luiz F. Brito[4], Gregor Gorjanc[3] & Brock A. Harpur[1]

As eusocial insects enter the post-genomic era, the availability of reference genomes and genomic datasets have facilitated Genome-wide Association Studies (GWAS) for traits measured on whole colonies. Several GWAS have been completed for traits like foraging and colony nesting behavior. However, no studies have addressed the specific challenges of applying GWAS to eusocial species. Eusocial species, such as honey bees, present several hurdles for traditional GWAS designs: they have high recombination rates, their social traits are generally polygenic, and the existence of multiple castes introduces additional layers of complexity in identifying genotype–phenotype associations. Moreover, many eusocial species are haplodiploid and polyandrous. To address these challenges, we used the honey bee (*Apis mellifera*) as a model to simulate GWAS scenarios for honey yield. We explored the influence of factors such as the number of quantitative trait loci (QTL), heritability level, number of SNP markers covering the genome, and sample size on QTL discovery. We also examined the impact of genotyping queens versus workers. Our results suggest that a p-value threshold of 1×10^{-6} and sample sizes of at least 1,000 are effective for identifying QTL for key traits. Including both queen and worker genotypes improved QTL discovery power, emphasizing the need to optimize GWAS methods for eusocial species. Our results should improve future GWAS in honey bees and highlights the need for additional simulations and standard-setting work in other eusocial species.

Keywords Genome wide association, eusocial, honey bee, social traits

A Genome-wide Association Study (GWAS) involves statistically testing genetic markers across the genomes of many samples of the same species to identify genotype–phenotype associations. Over the past decade, GWAS has revolutionized complex trait genetics, providing numerous associations for key traits across a diverse set of species[1–3]. Despite its success in identifying novel genes, quantitative trait loci (QTL), and biological pathways, and translating these findings into practical applications, GWAS in social insects remains underexplored, even though these species are of significant economic value[4,5], ecological importance[6–8], and provide products with recognized health benefits, including honey, propolis, and royal jelly[9–11]. Advances in genotyping and sequencing technologies, along with the increasing availability of high-quality reference genomes and population genomics data, provide an unprecedented opportunity to explore complex traits in non-model organisms[12–16]. Nevertheless, GWAS in eusocial insects remains challenging due to their complex biology.

In honey bees and other eusocial species, the organizational level that produces phenotypes differs from the level at which genotyping is typically conducted. Colony-level traits are expressed as group phenotypes, whereas genotyping is usually performed on individual queens, workers, or drones, or sometimes on pooled workers. This mismatch raises the issue of which individuals should be genotyped, since each strategy captures different aspects of colony genetic makeup and influences the identification and interpretation of causal genes[17]. Queens of highly eusocial species (like honey bees) typically mate with multiple males, producing patrilines of workers with distinct fathers[18–20]. Honey bee queens and workers are diploid, while drones are haploid and can be used to infer the queen's genotype. In honey bees, colony-level traits emerge from the coordinated

[1]Department of Entomology, Purdue University, West Lafayette, IN, USA. [2]Agricultural Institute of Slovenia, Ljubljana, Slovenia. [3]The Roslin Institute and Royal (Dick) School of Veterinary Medicine, The University of Edinburgh, Edinburgh, UK. [4]Department of Animal Sciences, Purdue University, West Lafayette, IN, USA. ✉email: mfikere@purdue.edu

behavior of thousands of individuals that function as a highly integrated social unit. Colony traits result from the collective activity of multiple interacting individuals. The queen influences colony traits through her through reproduction, producing the worker population and determining the colony's genetic composition through mating with multiple drones and through pheromonal regulation of colony organization, while the majority of tasks such as foraging, brood care, and nest maintenance are performed by the workers[20–24]. Social traits such as honey production[25], nest defense[15,26], and nest hygiene[16] are frequently the target of GWAS (Table 1). Modeling these traits requires careful partitioning of direct (worker) and indirect (queen) genetic effects on the phenotype. Eusocial Hymenoptera, and honey bees in particular, also have high recombination rates, which rapidly break down linkage disequilibrium between loci[27–29]. Combined with highly polygenic traits controlled by many loci with small effects, this necessitates dense marker coverage, large sample sizes, and sometimes multi-locus statistical methods that jointly model the effects of multiple genetic variants (e.g., mixed linear models or Bayesian approaches) to reliably detect associations[30–32].

A GWAS can be a powerful framework to identify genomic regions associated with important traits, enabling the discovery of markers for marker-assisted selection and improving our understanding of the genetic basis of complex social traits in honey bees[33–36]. Previous GWAS in honey bees have shown wide variability in study design with sample sizes ranging from nine to 1,441 colonies, DNA being extracted from individual workers, haploid drones, or pooled samples, and genotyping strategies including SNP chip arrays, whole-genome sequencing, or genotyping-by-sequencing at varying depths (Table S1). Few studies have applied stringent multiple testing correction and significance thresholds (e.g., $-\log_{10}(p) < 8$) for detection. A systematic evaluation of these design choices is lacking. In this study, we addressed this gap by evaluating how quantitative genetics parameters, source of information, and model design parameters affect statistical power, false discovery rate (FDR), and locus detection in honey bees. Using empirical genomic data from 1,625 worker bees (representing 980 colonies) and over 300,000 SNP markers after quality control, we simulated colony phenotypes under varying heritability levels, numbers of QTLs, SNP densities. By comparing models that accounted for queen effects, worker effects, or both, we provide an evidence-based framework for optimizing GWAS design in this complex eusocial system.

Materials and methods

Sample genotyping and quality controls

Empirical genomic data from 1,625 whole-genome sequences was collected from honey bees sourced from across the United States (accession: PRJNA1110198 and can be accessed DataDryad at https://doi.org/10.5061/dryad.j3tx95xpg) (Supplementary Table S1). DNA for genotyping was extracted from thoraces. The samples were placed in up to 200 µL of lysis buffer master mix containing proteinase K. The mixture was incubated at 65 °C for one hour. Following incubation, a subset of 30–50 µL was taken for bead-based extraction. After DNA extraction from individual workers, libraries were generated by Gencove Inc. (Lincoln, NE, USA) using the Illumina Nextera DNA Flex Library Preparation kit. Libraries were quantified using fluorometry, pooled, and assessed for quality with TapeStation (Agilent). Low-pass sequencing (~ 5x depth) was conducted on an Illumina NovaSeq 6000, and reads were assembled and mapped to the honey bee reference genome (Amel_HAv3.1[37]). The data were imputed to full genome coverage using 1,041 publicly available honey bee sequences data[38,39].

Genotype imputation was performed using Gencove's pipeline with an adapted version of GLIMPSE2[40], which applies the Li and Stephens hidden Markov model[41]. Posterior genotype probabilities were obtained through forward–backward computations on this HMM, where reference haplotypes define the hidden states, thus quantifying uncertainty in the imputed calls. After quality control of 15 million variants, 3.6 million biallelic SNPs were retained. To assess genotype imputation accuracy, 5,000 SNPs randomly distributed across the genome were masked and re-imputed using BEAGLE v5.4 [42]. This procedure was repeated 10 times, and Pearson's correlation coefficients calculated between observed and imputed genotype. Subsequently, SNPs with more than two alleles, < 99% genotype posterior probability, and minor allele frequencies lower than 0.05 were filtered out, resulting in a final dataset of 1,625 samples and 342,491 SNP markers (Figure S3). These SNPs were distributed across all chromosomes, providing representative genome-wide coverage for association testing.

Simulation parameters and GWAS scenarios

We simulated a total of 432 combinatorial scenarios (Fig. 1) to fully quantify the impacts of heritability (low, $h^2 = 0.02$; intermediate, $h^2 = 0.20$; and, high $h^2 = 0.50$), the number of QTL per chromosome (nQTL = 1, nQTL = 10, and nQTL = 100), the number of genotyped SNPs (nSNPs = 100, nSNPs = 1,000, and nSNPs = 10,000) per chromosome, the number of genotyped samples (nSamples = 100; 500; 1,000; 1,600), and the caste selected for genotyping (workers, queens, or both). Our strategy followed approaches previously described by Kanai et al. [43]; Saber and Shapiro[44] and Sesia et al., [45] and was designed to systematically evaluate how key GWAS design parameters such as sample size, heritability level, SNP density, number of QTLs, and the caste selected for genotyping affect statistical power, FDR, and QTL detection. The simulation scenarios were chosen to reflect the range of experimental designs used in honey bee GWAS to date (Supplementary Table S1) as well as the expected genetic architecture of social traits[46,47].

We simulated a honey bee population using the SIMplyBee R package[48], an extension of AlphaSimR v2.1.0 [49]. The genotype dataset was phased using BEAGLE v5.4. Haplotypes and genetic map data for 1,625 individuals were imported into AlphaSimR v2.1.0, and a random subset of 200 individuals was sampled to serve as founder genomes. Colony simulations were initialized in SIMplyBee with a colony size of 100 workers, the number of drones a queen mates with drawn from a Poisson distribution with mean 10, and 100 colonies per scenario (see Data availability section). Unmated queens and drones were generated from the founder population, and colonies were established through controlled mating. Each colony was then built up to contain 100 workers and 30 drones, after which colony yield phenotypes were calculated using the *calcColonyPheno*() function, integrating contributions from both the queen and workers. These colony-level phenotypes were used for

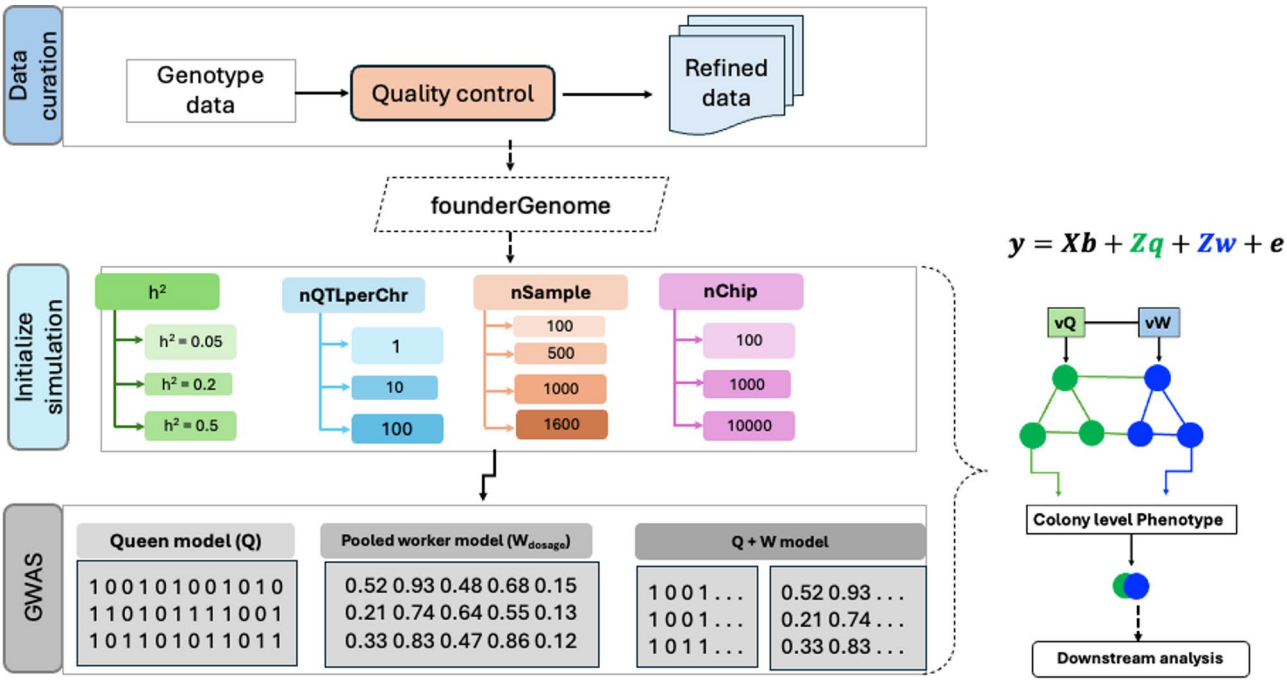

Fig. 1. Overview of the simulation parameters and genome-wide association study (GWAS) scenarios workflow. h^2 = trait heritability; nQTLPerChr= number of quantitative trait loci per chromosomes; nSample = number of samples ran in each GWAS scenario; nChip (per chromosome) = number of SNPs subsetted from whole genotyped data used to perform GWAS. The last row of the figure illustrates the marker data format, including hardcoded genotype (0,1,2) calling for the queen model (Q) where queen genotypes were fitted; and dosage-based genotype probabilities for the worker model (W) where workers genotypes are fitted and a combined queen-worker model was fitted to the simulated phenotype and marker data using a linear mixed model. Subsequently, the breeding value was back solved to get SNP effects, and GWAS summary statistics were generated following the methodology outlined in the model parameter section. In addition, to assess the impact of minor allele frequency on statistical power, analyses were conducted using three filtering thresholds (< 1%, 5%, and 10%) across all sample size scenarios under the queen-worker model.

downstream GWAS analyses (Figure S20. To balance biological realism and computational efficiency, we simulated 100 workers per colony, which was assumed sufficient to capture within-colony (patriline) genetic diversity in the model (Figure S3). Model parameters, including phenotypic variance, were scaled to colony size (e.g., proportional to 10/nWorker for workers), ensuring consistent genetic expectations. Additional analyses across a wider range of worker densities (10-1000) are provided in the Supplementary Material to assess the robustness of this assumption (Figure S3). These analyses demonstrate that sampling 100 workers per colony provides estimates of colony-level genetic values that are comparable in accuracy and precision to those obtained with larger sample sizes (e.g., 1000 workers), with only marginal gains beyond this point, indicating that the chosen sample size is adequate.

We simulated a single trait that represented honey production, expressed at the colony level[15,50,51]. Internally, we simulated two attributes representing queen and worker effects, assuming a moderate negative genetic correlation (r = -0.40) consistent with previously reported antagonistic relationships between queen and worker traits in the literature[50,52,53]. QTLs for each trait were randomly sampled and assigned normally distributed effects. A colony's genetic value for honey yield was calculated as the sum of the additive effects of the 100 simulated workers and their queen.

Lastly, the colony phenotype was calculated using the *calcColonyPheno*() function in the SIMplyBee package[48]. This function maps caste-specific genetic values to colony-level phenotype by retaining the queen's genetic value and summing the contributions of all workers in the colony. The queen component represents the maternal genetic influence on colony organization and reproduction, whereas the summed worker values represent the collective contribution of workers to colony performance. These components are then combined to generate a single colony-level phenotype that reflects the joint effects of queen and worker genetics. Honey production was assumed to follow a normal distribution with a phenotypic mean of 20 kg and phenotypic variance of 4 kg², resulting in most colonies producing between 14 kg and 26 kg, closely matching real-world values[54]. Colony level trait heritability for honey yield trait was calculated as $h^2 = \frac{\sigma^2_g}{\sigma^2_P}$; where $\sigma^2_g = \sigma^2_W + \sigma^2_Q + 2\mathrm{Cov}(W, Q)$ is the total genetic variance at the colony level (comprising worker variance, queen variance, and their covariance), and $\sigma^2_P = \mathrm{Var}(y)$ is the phenotypic variance of observed colony phenotypes. To achieve the desired trait heritability specified in our simulation scenarios (0.05, 0.20, and 0.50), the environmental variance associated

with the effects of the queen and worker bees was iteratively adjusted (Fig. 1 and Figure S2). All calculations were implemented using R code (see code availability section).

Genomic relationship matrix (GRM) calculation

A genomic relationship matrix (GRM) was calculated for all samples using a standard method as outlined by *VanRaden (2008)*[55] and implemented in the SIMplyBee R package[48]. For queen samples, hardcoded genotypes were used, with genotypes encoded as 0, 1, or 2 to represent the different genotype states. For worker samples, the GRM was constructed using dosage-format data, representing allele frequencies within the colony (i.e., pooled sequencing). Following VanRaden (2008)[55], the standardized genotype matrix was computed as:

$$Z = M - P$$

where M is an $n \times m$ matrix of genotypes (coded as 0, 1, or 2, or in dosage format), and P is an $n \times m$ matrix with each column equal to $2p_i$, the expected allele count at locus i given the allele frequency p_i of the reference allele. This centering step ensures that Z represents deviations of individual genotypes from the population mean. The general estimation of the GRM for both queens and workers followed this same methodology.

Model and genetic parameters

Our model accounted for direct (worker) and indirect (queen) genetic effects[14,50,56,57]. Queen genetic effects were considered indirect because the queen does not directly perform colony tasks but influences colony phenotypes through the genetic composition of the worker population she produces. Each simulation incorporated SNP chip data of varying densities (100, 1,000, and 10,000 markers). We used ASReml-R v4.2[58] to estimate additive genetic effects and partition the phenotypic variance of colony traits into contributions from queens and/or workers. As described in Fig. 1, the GWAS scenarios performed were: a queen-only model to estimate indirect genetic effects, a worker-only model to estimate genetic effects, and a combined queen–worker model to jointly estimate both sources of genetic variation. Genomic estimated breeding values (GEBV), representing the aggregate genetic merit captured by genome-wide SNP markers through the genomic relationship matrix (G), were first estimated. SNP effects were subsequently back-solved from the genomic estimated breeding values (GEBV) using the relationship between GEBV and marker effects defined by the genomic relationship matrix, allowing marker-level inference.

Queen model (Q)

The queen-only GWAS model estimates the genetic contribution of queens to colony traits:

$$y = 1\mu + Z_Q u_Q + e$$

Where y is the vector of phenotypic observations, μ is the overall mean, $Z_Q$is the incidence matrix linking records to queens, $u_Q \sim N(0, G_Q \sigma^2_Q)$ is the vector of queen additive genetic effects with G_Q being the queen GRM calculated using VanRaden (2008)[55] method, and $e \sim N(0, I^2_{\sigma e})$, e is the residual errors. SNP effects were back-solved from the estimated GEBVs $(\hat{u}_Q)$ using:

$$\hat{a}_Q = M_Q^T G_Q^{-1} \hat{u}_Q$$

Where M_Q is the centered genotype matrix. The variance–covariance matrix of SNP effects was computed as:

$$Var\left(\hat{a}_Q\right) = M_Q^T G_Q^{-1} Var\left(\hat{u}_Q\right) G_Q^{-1} M_Q$$

Standard errors were obtained from the square root of the diagonal, and two-sided p-values were calculated from t-statistics:

$$t_j = \frac{\hat{a}_j}{SE\left(\hat{a}_j\right)}, p_j = 2xpt(-|t_j|, df)$$

Worker model (W)

The worker-only GWAS model estimates the genetic contribution of workers to colony traits:

$$y = 1\mu + Z_W u_W + e$$

Where $Z_W$is the incidence matrix for workers, $u_W \sim N(0, G_W \sigma^2_W)$ is the vector of worker additive genetic effects with G_W being the worker GRM calculated from dosage genotype calls, and e is the residual error. SNP effects were back-solved using the same procedure as for queens.

Combined queen + worker model (Q + W)

The combined model jointly estimates indirect queen and direct worker genetic effects:

$$y = 1\mu + Z_Q u_Q + Z_W u_W + e$$

Where $u_Q \sim N(0, G_Q \sigma^2_Q)$ and $u_W \sim N(0, G_W \sigma^2_W)$ are independent. SNP effects for both queens and workers were back-solved from their respective GEBVs, with the block-diagonal GRM used in the variance–covariance computation. For the combined model, GEBV was back solved to obtain SNP effects for both queens and workers using a block-diagonal GRM:

$$G = \begin{bmatrix} G_Q & 0 \\ 0 & G_W \end{bmatrix}, \; \acute{a} = M^T G^{-1} u, \; Var\left(\acute{a}\right) = M^T G^{-1} Var\left(\acute{u}\right) G^{-1} M$$

We followed this approach to ensure consistency across all models and GWAS scenarios, providing SNP-level estimates, standard errors, and two-sided p-values derived from the t-statistics of back-solved effects.

Estimation of study-wide significance thresholds

We extracted summary statistics and information from each run that included variant identifiers, genomic positions, effect sizes, standard errors, and p-values across every simulated scenario (Fig. 1). To determine a genome-wide significance threshold ($-log_{10}(p_{sig})$), we estimated the distribution of the minimum p-values observed across all tested variants, denoted as p_{min}, following *Kanai et al. (2016)*[43]. The minimum p-value for each scenario corresponds to the single variant with the strongest association signal. We then transformed P_{min} into $-log_{10}(p_{min})$ and used the Harrell–Davis distribution-free quantile estimator (*hdquantile* function from Hmisc R package[59]; to determine the 95th percentile at a significance level of $\alpha = 0.05$. The Harrell–Davis estimator was used because it provides a more stable and efficient estimate of population quantiles by using a weighted combination of all order statistics, which is particularly advantageous when estimating extreme quantiles from finite samples. The estimated threshold was subsequently converted back to the original p-value scale.

A bootstrap resampling approach (100 iterations) was applied to estimate 95% confidence intervals (CI) for $-log_{10}(p_{sig})$. In each bootstrap iteration, the set of genome-wide SNPs p-values was resampled with replacement while maintaining the same number of markers as in original dataset. For each resampled dataset, the minimum p-value (p_{min}) was recalculated and transformed to $-\log_{10}(pP_{sig})$, generating an empirical distribution of the significance threshold across iterations. The 95% CI for $-\log_{10}(p_{sig})$ was then derived using the percentile-based confidence interval estimation approach, where the 2.5th and 97.5th percentiles of the bootstrap distribution were used as the lower and upper confidence limits[43]. One hundred bootstrap iterations were used to obtain a stable approximation of the confidence interval while maintaining computational efficiency given the large number of genome-wide SNP markers analyzed. To further validate the study-wide significance threshold derived from the p_{min} distribution, we applied the Benjamini–Hochberg (BH) procedure to control the FDR at a threshold of 0.05[60,61]. The genome-wide significance threshold was then applied to the SNP-level p-values obtained from back-solving the GEBV for each model, ensuring that the identified significant variants reflect both the estimated genetic effects and study-wide error control.

False discovery rate (FDR) across GWAS scenarios

We evaluated the impact of trait heritability, sample size, queen–worker model assumptions, and the number of QTLs in the genome (henceforth "study variables") on the False Discovery Rate (FDR). True Positives (TP) were defined as the loci corresponding to the true simulated QTLs, whereas False Positives (FP) referred to additional loci identified that were not among the true QTLs. FDR was estimated using a 50 kb window around each true site at a significance threshold of $p < 5 \times 10^{-6}$. The 50 kb window was selected to account for local linkage disequilibrium around simulated QTLs and reflects the relatively short LD extent observed in the honey bee genome, ensuring that nearby markers capturing the same association signal are considered true positives. For practical implementation, the G2P tool[62] was used to compare detected significant markers, with the simulated causal variants, enabling the classification of true positive (TP) and false positive (FP) signals. FDR was then calculated as the proportion of FP windows among all detected TP signals[61].

Estimating statistical power

To assess the effect of each study variable on statistical power, we followed established methods[63,64]. First, we calculated the non-collinearity parameter, reflecting the proportion of trait variance explained by a genetic marker[65,66]. Next, we determined the critical significance threshold by calculating the Z-critical value corresponding to the genome-wide significance level, using the inverse cumulative distribution function of the standard normal distribution:

$$Z_{critical} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

In this study, we used $\alpha = -log10\,(p) = 6$, which corresponds to the genome-wide threshold, ensuring appropriate control of false positives.

The power of each SNP was then calculated as:

$$Power = \phi\,(\lambda - Z_{critical}) + (1 - \phi\,(\lambda + Z_{critical}))$$

Where ϕ is the cumulative distribution function of the standard normal distribution, and λ is the non-centrality parameter associated with the SNP effect. We further assessed the impact of MAF filtering thresholds on statistical power. This approach allowed us to quantify the probability of detecting true genetic effects under

each study variable, with average power estimated across 10 replications, which provided stable estimates while maintaining computational efficiency given the large-scale GWAS simulations performed.

Results

Establishing study-wide p-value thresholds for GWAS in honey bees

We first sought to evaluate and establish a p-value threshold for a honey bee GWAS. We followed *Kanai et al., (2016)*[43] and examined the $-\log_{10}(p_{min})$ distributions of all models (i.e., queen, worker, and a combined queen-worker model) and study conditions (Fig. 2A). The distribution of p-values varied across models, leading to different significance thresholds for capturing GWAS peaks (Fig. 2A). An average threshold of 10^{-4} was identified for the queen model, 10^{-5} for the worker model, and 10^{-7} for the combined queen-worker model. Given this variation, a conservative study-wide p-value threshold of $p_{sig} = 1.2 \times 10^{-6}$ was selected, with the 95%

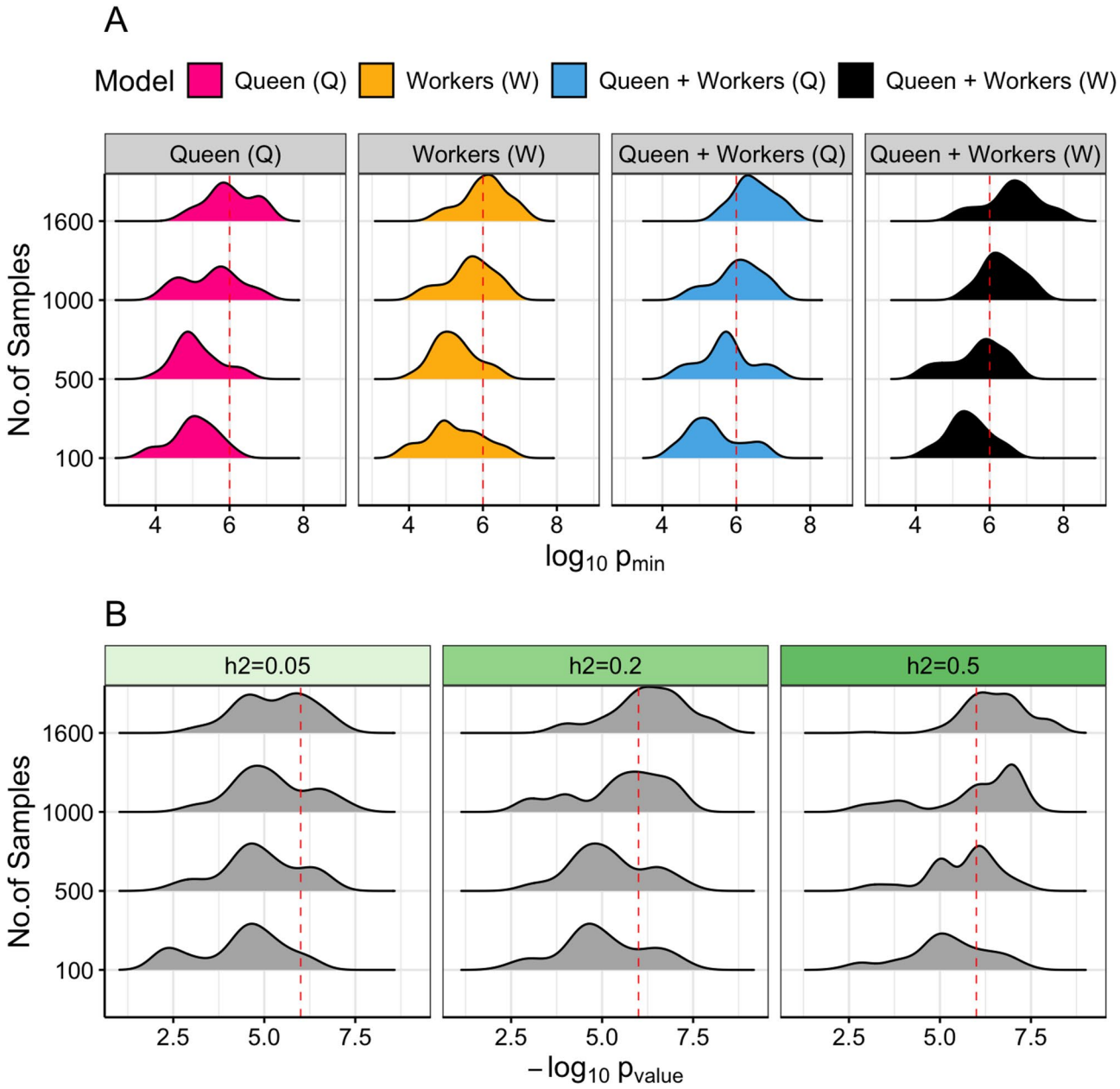

Fig. 2. The distribution of $-\log10(p_{min})$ values is presented for four honey bee GWAS models (**A**). Distribution of -log10(p-value) across samples for three heritability levels (0.05, 0.2, 0.5) (**B**). The minimum p-value (p_{min}) for each model assumption (detailed in the Methods section) was estimated within specific window size. This assumption was evaluated across different sample sizes and the queen-worker model while maintaining a constant moderate heritability ($h^2 = 0.2$), a fixed number of 10 predefined QTLs, and 10,000 markers (chip density) per chromosomes. The dotted vertical line indicates the study-wide significance threshold of 1.2×10^{-6}.

confidence interval derived from the bootstrap distribution of $-\log_{10}(p_{min})$, to ensure stringent control of FP across all analyses (Fig. 2A and B). To validate this threshold, we assessed the $-\log_{10}(p_{min})$ distribution pattern for a scenario with lowest FDR (FDR < 10%; see section below) and we found a similar p-value distribution pattern between a scenario that exhibited a lower FDR and the $-\log_{10}(p_{min})$ approach, supporting the robustness of the selected threshold. The study-wide p-value threshold of $p_{sig} = 1.2 \times 10^{-6}$ offers a balance between stringent significance control and the ability to detect true genetic associations across different model assumptions.

Trait heritability significantly impacted the optimal p-value threshold used to detect true genetic associations (Fig. 2). No significant peaks were detected when applying very stringent p-value thresholds for traits with low heritability ($h^2 = 0.05$). However, significant peaks emerged when using less stringent thresholds ($p < 10^{-4}$ to 10^{-5}). Under these relaxed criteria, FDR was notably impacted (~60% more FPs). In contrast, traits with moderate ($h^2 = 0.20$) and high ($h^2 = 0.50$) heritabilities allowed for more stringent thresholds ($p < 10^{-4}$ to 10^{-8}), reflecting stronger genetic signals (Fig. 2A). Similarly, fewer significant associations and a more widely spaced p-value distribution were observed when the sample size was reduced from 1,600 to 100, reflecting weaker genome-wide signals (Fig. 2). We further compared the p-value distributions between models and genotyping format (dosage vs. hardcoded). This difference likely arises because genotype dosage retains the probabilistic information from genotype calling or imputation, representing the expected allele count as a continuous value (e.g., 0.01–1.99), whereas hard-called genotypes discretize these probabilities into integer classes (0, 1, or 2). Hard-calling therefore ignores genotype uncertainty and may introduce rounding errors, particularly for markers with ambiguous genotype probabilities. By preserving the underlying probability information, dosage-based genotypes capture subtle variation in allele dosage and reduce information loss. This improved representation of genotype variation can increase the precision of association tests and produce a more accurate null p-value distribution, which in turn leads to a more stringent genome-wide significance threshold. Interestingly, using dosage (worker model) instead of hard-coded genotypes resulted in a lower p-value threshold (from $p < 10^{-4}$ to $p < 10^{-5}$), even under moderate trait heritability conditions (Fig. 2B). We proceeded with a $p_{sig} = 1.2 \times 10^{-6}$ for all following examinations, unless otherwise specified. This is also a true reflection of the colony genotype, as worker genotypes include paternal contributions and therefore provide a more complete representation of the colony's genetic makeup.

Effects of sample size and marker density on power and FDR

Power was impacted by all study variables, with sample size, trait heritability, and the genotyping strategy (queen vs. worker) having the largest effects. For low-heritability traits ($h^2 = 0.05$), larger sample sizes (> 1,500) were required to achieve sufficient statistical power (e.g., ≥ 80%) to reliably detect true associations, whereas for higher heritability traits ($h^2 = 0.50$) smaller sample sizes (< 100) were adequate to reach comparable power (Fig. 3). When trait heritability increased from moderate ($h^2 = 0.20$) to high ($h^2 = 0.50$), power improved by up to 20% compared to low heritability ($h^2 = 0.05$) at the same sample size. In scenarios where the sample size was less than

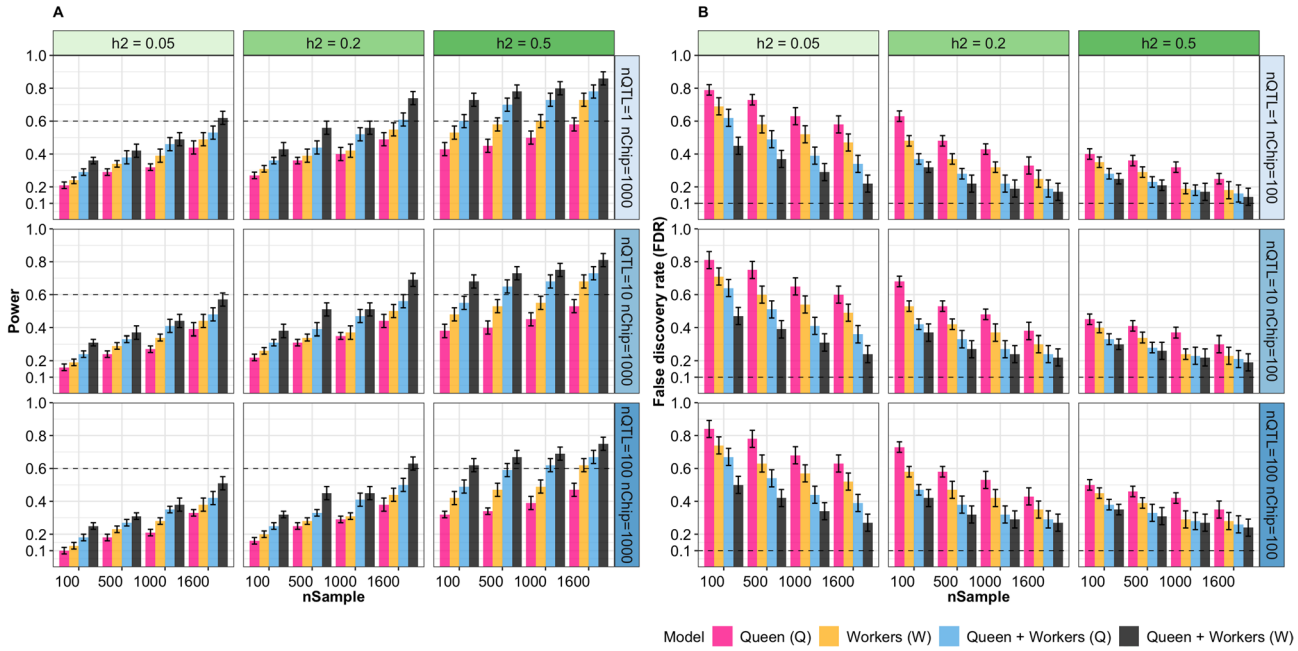

Fig. 3. Impact of sample size, trait heritability (h^2), number of QTL per chromosome (nQTL), and model type on honey bee GWAS outcomes. (**A**) shows the statistical power, and (**B**) shows the false discovery rate (FDR) across different scenarios. Each panel compares four models: Queen-only (Q), Workers-only (W), Queen + Workers (Q), and Queen + Workers (W). Results are presented for varying sample sizes (100, 500, 1000, 1600) and three levels of heritability ($h^2 = 0.05, 0.2, 0.5$). The analysis focuses on the nChip = 100 per chromosome scenario, as illustrated in Fig. 1.

500, trait heritability played a crucial role in lowering FDR (FDR < 20%) and increasing power (> 60%). Overall, increasing the sample size from 100 to 1,600 enhanced power (rising from approximately 10% to 90%; Fig. 3 and Supplemental Figures S4 and S5) and reduced FDR (lowering from 60% to 1%). The number of QTL also impacted Power and FDR. In addition, the number of predefined QTL (nQTL = 1) exhibited increased Power while simultaneously reducing the FDR (Fig. 3, Supplementary Figure S4&S5). In contrast, a larger number of QTL led to a decline in power (Figs. 3, Supplementary Figure S5). Generally, our results suggest that over 1,000 samples are required to achieve a 5% FDR and > 60% Power for most traits. These findings are consistent with results observed in standard diploid animal models[67–69], highlighting parallels while also extending such evidence to eusocial insects.

Effect of queen and worker components on power and FDR

We found that the selection of the genotyped individual (queen vs. worker) and the inclusion of queen or worker effects in the GWAS model have dramatic influence on FDR and Power. A worker model generally outperformed a queen model (Figs. 3 and 4, Supplementary Figure S4); however, the combined model incorporating worker dosages and queen genotype consistently yielded higher Power (> 80%) and lower FDR (< 1%) compared to either the queen or worker models alone (Figs. 3 and 4, S3). Notably, using the queen model alone resulted in a higher FDR, even at moderate trait heritability ($h^2 = 0.20$), across a sample size range of 500–1,600. When both the queen and worker components were combined, power increased by over 40%, demonstrating the advantage of integrating castes in honey bee GWAS analyses. In addition, comparing the number of significant SNPs across models revealed that incorporating the queen-worker model (direct genetic effect) in the GWAS resulted in the identification of more significant SNPs than using either the queen or worker model alone (Fig. 4A-D).

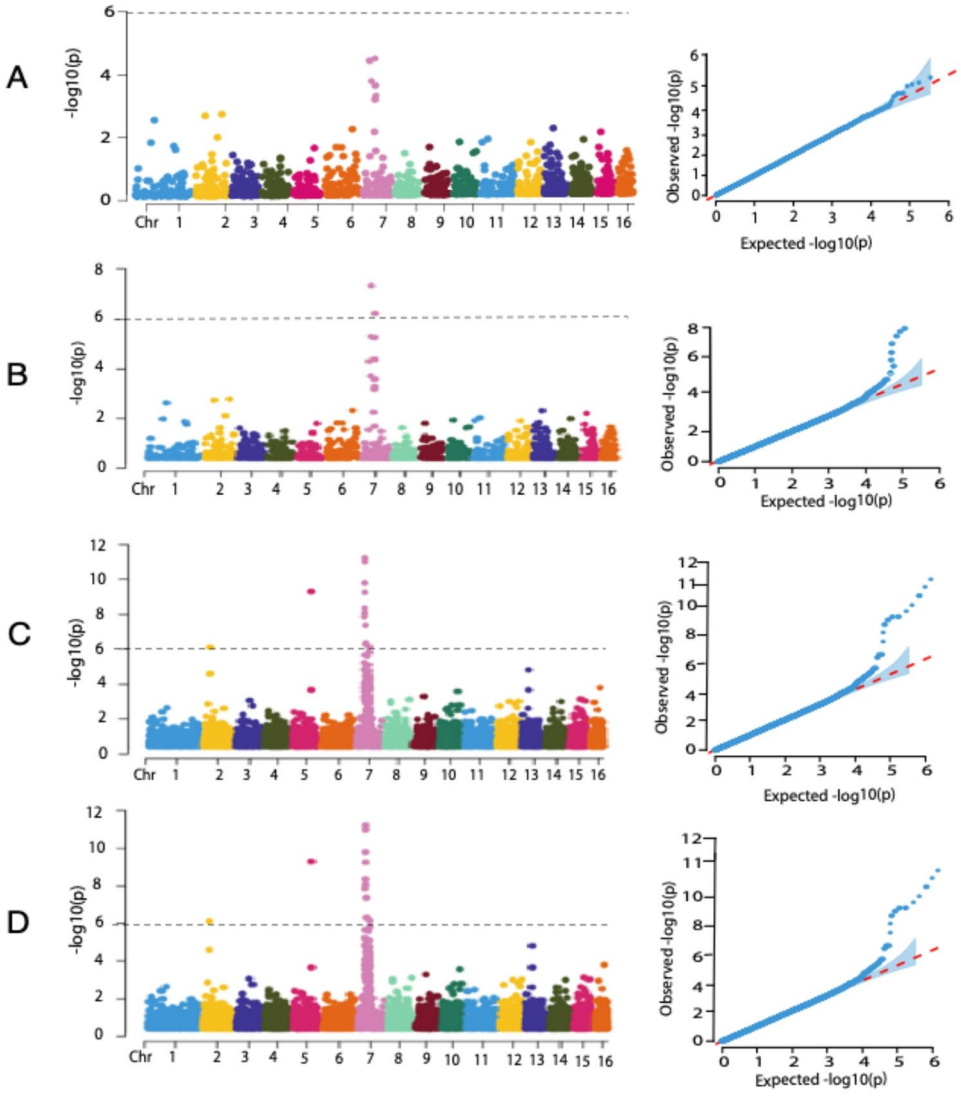

Fig. 4. GWAS Summary Statistics for Different Models (Example Scenario). (**A**) Manhattan and QQ plots for the Queen model (lambda = 1.02); (**B**) Worker model (lambda = 0.98); (**C**) Queen + Worker model (Queen effect) (lambda = 1.13); (**D**) Queen + Worker model (Worker effect) (lambda = 1.01).

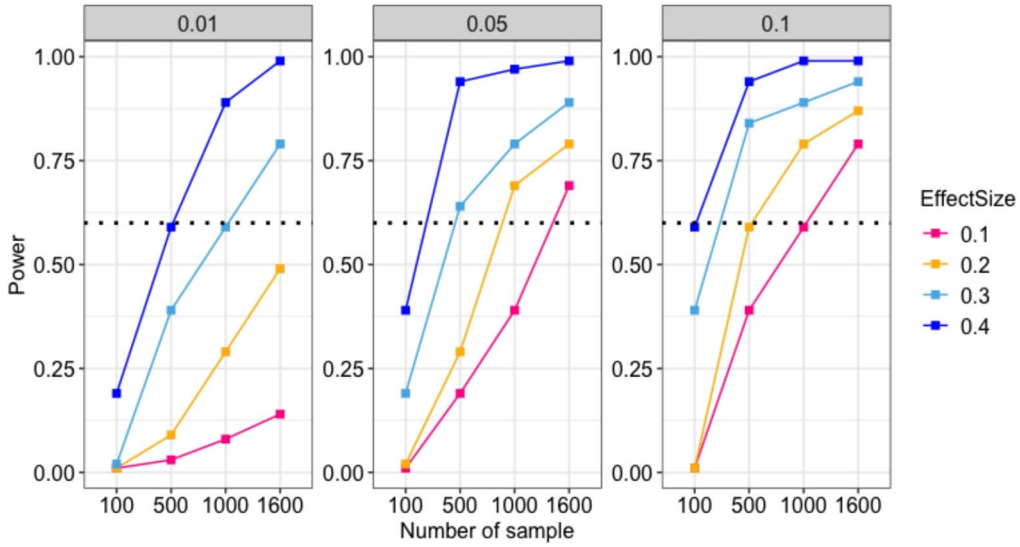

Fig. 5. Estimated statistical power across different minor allele frequency (MAF) thresholds, sample sizes, and SNP effect sizes. Power to detect true genetic associations in the GWAS simulations under varying study conditions. Panels represent different MAF filtering thresholds (0.01, 0.05, and 0.10). The x-axis shows the number of sampled colonies (100, 500, 1,000, and 1,600), while the y-axis indicates the estimated statistical power. Colored lines correspond to different SNP effect sizes (β = 0.1, 0.2, 0.3, and 0.4). The horizontal dashed line represents the commonly used power benchmark of 0.6 for reliable detection of true associations.

Allele frequency and power of detection

We further investigated the effect of MAF thresholds (1%, 5%, and 10%) across varying sample sizes (100, 500, 1,000, and 1,600) and stratified marker effect sizes (β) on power. Our analyses revealed that both increasing sample size (from 500 to 1600) and higher MAF threshold (1% to 5%) substantially improved the power to detect true associations (Fig. 5). Notably, the power gains were more pronounced at higher β values, indicating that QTLs with large effect are more reliably detected even at lower allele frequencies when sample sizes are large (> 1,000). These findings emphasize the importance of optimizing both marker selection (e.g., through quality control and selecting variants that represent genomic regions such as high linkage disequilibrium (LD) blocks) and sample size to enhance power.

A similar power trend was observed across MAF classes, with common variants (MAF > 10%; 52,495 SNPs) yielding the highest power (> 70%), followed by moderate-frequency variants (1% < MAF ≤ 5%; 109,483 SNPs), and rare variants (MAF < 1%; 180,513 SNPs) showing the lowest power (Fig. 5). This suggests that the detection of genetic associations is more efficient for common variants, likely due to their higher allele frequency and better representation in the population, which enhances statistical power. In contrast, rare variants (MAF ≤ 1%) exhibited considerably lower power, even at larger sample sizes, underscoring the inherent challenges in capturing the effects of such variants. This limitation may arise from low allele counts and increased variability in effect estimation, which together reduce the likelihood of detecting statistically significant associations by ~ 20% unless the effect size is substantial. Therefore, future GWAS aiming to detect rare variant associations in honey bees may benefit from integrating alternative strategies, including tailored modeling approaches (queen-only, worker-only, or combined) and diverse sequencing designs (queen, worker, drone, or pooled samples), and larger samples sizes to complement traditional GWAS methods.

Discussion

In this study, we evaluated the impact of sample size, trait heritability, and statistical models on power and FDR. Our findings show how model choice influences power and FDR in a social phenotype context, while also highlighting that both trait heritability and sample size are critical for increasing power and reducing FDR. Furthermore, we identified optimal p-value thresholds for different GWAS scenarios, providing foundational insights into optimizing GWAS methodologies for honey bee trait analysis and offering a framework for improving the accuracy and reliability of genetic association studies in populations with complex social structures.

Specifically, our findings indicate that selecting an appropriate sample size requires balancing the need for sufficient power to detect true associations while maintaining statistical rigor. For low-heritability traits, sample sizes above 1,000 colonies are recommended, whereas moderate-to-high heritability traits can achieve reliable associations with smaller sample sizes (> 500), provided marker density is adequately high. To date, only a few honey bee GWAS studies have exceeded 1,000 samples (Supplemental Table S1). Additionally, incorporating both queen and worker effects into the GWAS model has the potential to reduce false positives, even for traits with low to moderate heritability, where many honey bee traits typically exhibit heritability below 0.50.

Trait architecture and complexity in honey bees: Implications for GWAS and genomic selection (GS)

Genetic improvement programs exist for honey bees and have historically relied on artificial selection for traits of human interest. However, the implementation of advanced genomic breeding approaches in social insects remains limited despite the availability of reference genomes for species such as honey bee[37], wasps[70], ants[71], and others. Addressing key research priorities and methodological gaps is important given the critical economic, health, and ecological roles these insects serve. Beyond the relatively late initiation of large-scale sequencing efforts in insect species, unique biological and statistical challenges remain that must be addressed to effectively implement advanced genomic-based improvement schemes in honey bees.

A recent GWAS study by *Eynard et al. (2025)*[14] demonstrated that the polygenic nature of honey bee traits, such as Varroa infestation, recapping behavior, and mite non-reproduction (MNR), together with the haplodiploid genetic system of the species, demands specialized GWAS models that account for these unique breeding features. In their study, queen genotypes were reconstructed from worker genotypes across 1,500 samples, and both single-trait and meta-GWAS were performed using a queen-based model. Despite the large sample size, the study detected few or no significant associations for disease-related traits, which may reflect the highly polygenic architecture of these traits, as suggested in previous studies, as well as potential limitations associated with relying solely on reconstructed queen genotypes. Our findings suggest that a combined worker–queen modeling approach would capture both direct and indirect genetic effects, thereby increasing statistical power and leading to the detection of more significant SNPs associated with these traits. These insights may support advanced breeding strategies such as genomic selection (GS), which relies on genome-wide marker information and well-defined reference populations of genotyped and phenotyped individuals, but also incorporate previously identified QTL or other biological information (e.g., through Bayesian models) to improve prediction accuracy and enhance future colony improvement programs[52 72–76].

Several approaches have been used to investigate trait-marker associations in honey bees (Table S1). Some GWAS studies consider only the queen's genotype[15,35], others use pooled worker samples[16,77,78], while others reconstructed the queen's genotype from pooled worker genotypes[14,14],. These studies apply varying GWAS significance thresholds to identify significant SNPs. In our study, we evaluated multiple study variables and model combinations, resulting in over 432 GWAS scenarios to help establish standardized practices for conducting GWAS in honey bees (Fig. 1). Our results suggest that a p-value threshold of 1×10^{-6} provides an effective balance between detection power and control of false discovery rate under our simulation settings. Specifically, this threshold minimized false positives while maintaining sufficient sensitivity to detect QTLs, particularly at sample sizes of at least 1,000. Furthermore, incorporating both queen and worker genotypes improved discovery power, reinforcing the importance of optimizing GWAS strategies for eusocial species.

To ensure the statistical validity of genotype–phenotype associations, GWAS must employ appropriate p-value thresholds, FDR controls, and power calculations to account for the multiple-testing burden[79,80]. Although a genome-wide significance threshold of $p = 5.0 \times 10^{-8}$, >60% Power, and a 20% FDR are commonly used in diploid organisms[60,69,81,82], their suitability for honey bees and other eusocial populations remains uncertain and not standardized. What sample size is required to achieve optimal Power while maintaining a low FDR? What significance thresholds should be applied, and how should they be determined? Which statistical models are best suited for accurately disentangling the genetic effects of queens and workers? These are critically important questions that need to be addressed before pursuing functional genomics or marker-assisted selection[79,83–86].

Our study also demonstrated how sample size and trait heritability are critical factors in the design of genomic studies. In honey bees, heritability ranges from as low as ~0.08 ± 0.01 for traits[51,87] like gentleness and Varroa infestation to as high as 0.65 for hygienic and defensive behaviors[88–90]. For instance, hygienic behavior is highly heritable ($h^2 = 0.65$). However, some of these estimates are derived from relatively small sample sizes and are associated with large standard errors, which should be considered when interpreting their magnitude. *Harbo and Harris, (1999)*[89] and *Spivak, (1996)*[91] reported a relationship between hygienic behavior and the removal of mites from brood cells; with colonies selected from hygienic behavior showing reduced mite levels compared to controlled colonies. However, the strength of this relationship has not always been consistent across years or studies, and evidence for the long-term repeatability and maintenance of these traits under selection remains limited. Therefore, while hygienic behavior is considered a promising trait for improving resistance to Varroa destructor, its effectiveness may depend on population, environmental conditions, and breeding strategies. Interestingly, most of these polygenic traits exhibit genetic correlations[92], presenting opportunities to combine genetically related traits in GWAS-meta analyses. This approach enhances the power and effectiveness of GWAS meta-analyses, making them valuable for novel gene discovery and functional analysis by exploring pleiotropic effects across traits[69,72,73,93].

Statistical challenges in genomic studies of social insects

Another major hurdle in the genomic study of social insects is the lack of statistical tools capable of comprehensively addressing the biological structure and genetic makeup of hymenopteran social insects. For instance, constructing a GRM for haplodiploid species requires accounting for fundamental differences in genetic makeup. Unlike diploid species, where individuals possess two sets of chromosomes, in haplodiploid species, females develop from fertilized (diploid) eggs, while males develop from unfertilized (haploid) eggs[94]. This reproductive system results in unique relatedness patterns within colonies; for example, full-sister workers can share up to ~75% of their genes due to haplodiploidy, although this relatedness is often reduced in honey bees because queens typically mate with multiple drones. Consequently, colonies are composed of multiple patrilines, and many colony-level phenotypes represent the aggregated contributions of individual workers carrying different genotypes. This ploidy difference fundamentally alters the calculation of genetic relatedness: while diploid relatedness is determined based on the proportion of alleles shared identical by descent (IBD), the

presence of haploid males necessitates adjustments, typically halving the relatedness between diploid individuals and haploid males[95–99]. In this study, we utilized *SIMplyBee*, the first simulation package specifically designed for honey bees[48], to simulate phenotypes and construct a GRM using empirical genotype data.

In addition to these challenges, population stratification represents an important source of bias in genome-wide association studies (GWAS) of social insects. Differences in allele frequencies among genetically structured populations can lead to spurious associations if underlying structure is not adequately accounted for[100]. Evidence from human, plant, and animal GWAS has consistently shown that properly accounting for population structure improves the accuracy of association signals, reduces false positives, and increases the power to detect true genotype–phenotype relationships[45,100,101]. In honey bees, such structure can arise from differences between feral and managed populations, the coexistence of multiple evolutionary lineages (e.g., African, Western and Eastern European), and varying levels of admixture and ancestry across breeding populations[102]. These factors can lead to substantial heterogeneity in allele frequencies and linkage disequilibrium patterns, thereby influencing GWAS outcomes. Although the genomic inflation factor observed in our dataset was low, suggesting limited confounding, the complex and multi-layered population structure of honey bees, combined with the presence of multiple patrilines within colonies, may still affect association signals. Accounting for such structure through approaches such as principal component analysis or mixed linear models is therefore essential to improve the robustness and interpretability of GWAS results in honey bees. Building on these considerations, another key challenge in GWAS of social insects is the quality and reliability of phenotypic data, particularly for behavioral and hygienic traits, which are often influenced by environmental variation and colony-level interactions. Measurement error and phenotypic plasticity can substantially reduce the power to detect true genetic associations, emphasizing the need for standardized and replicated phenotyping protocols[103]. In this context, incorporating pedigree information can provide additional resolution by capturing relatedness and improving the partitioning of genetic and environmental variance. Such integrative approaches are likely to enhance the accuracy of marker–trait associations and strengthen downstream applications in genomic selection.

Simulation in social insects: Insights into colony-level phenotype mapping

Simulation testing is a valuable tool for evaluating study designs, especially where empirical data are limited. In honey bee research, simulations have been particularly useful for assessing the efficiency of breeding strategies[104–106]. For instance, *Plate et al. (2020)*[106] investigated alternative breeding schemes for endangered honey bee species, and *Kistler et al. (2021)*[105] showed that polyandrous mating led to equal or greater genetic gains compared to monoandrous mating, while maintaining significantly lower inbreeding rates.

Many honey bee traits are social in nature, arising not from an individual's genotype alone but from interactions among colony members[107,108]. These traits involve both direct genetic effects expressed through workers reflecting both maternal and paternal genetic contributions and indirect genetic effects from the queen, which influence colony phenotypes through reproduction and regulation of colony organization. Together, these are referred to as colony-level or social phenotypes. Traits such as honey production, hygienic behavior, temperament (e.g., calmness, gentleness, aggression), and disease resistance exemplify this complexity. Accurate mapping and modeling of these traits at the colony level are crucial[53,109,110]. Simulation studies have helped clarify the limitations of GWAS, such as the effects of sample size, population structure, and genetic complexity, while providing insights into mitigating biases and reducing false positives[111–114].

Implications of the study

Our simulated phenotypes explicitly included both queen and worker effects, the GWAS model incorporating these effects performed best. This suggests that the advantage of such models would likely be even greater when applied to large and well-structured datasets.

However, since this model is likely closer to biological reality, our results provide important estimates of the loss in statistical power when fitting only queen or worker effects singly. These results are especially useful when designing experiments where sequencing strategies may only accurately yield queen or worker genotypes, but not both. In practice, obtaining queen genotypes can be challenging because queens are often difficult to sample without disturbing colony functioning and are usually represented by a single individual; moreover, destructive sampling would result in the loss of the queen. In contrast, worker sampling is typically easier and allows pooling individuals from multiple patrilines, providing a broader representation of the colony's genetic composition. Given that worker effects outperformed queen effects when modeled singly, sequencing approaches that capture worker genetic variation (e.g., sequencing worker pools) may therefore be advantageous. However, in honey bees, worker pool sequencing cannot fully resolve paternal contributions due to the queen's polyandrous mating system, and accurate estimation of allele frequencies from pooled samples requires sufficient sequencing depth and careful consideration of potential dosage estimation challenges and alignment biases.

Conclusion

In summary, our study demonstrates that both sample size and trait heritability are critical determinants of power and false discovery rate in honey bee GWAS. We show that reliable detection of rare variant associations requires large sample sizes (> 1,000 colonies), while moderate-to-high heritability traits can be studied with fewer samples provided marker density is sufficient. Importantly, models that incorporate both queen and worker effects yielded the best performance in terms of statistical power, reduced false discovery rates, and improved detection of true QTLs, underscoring the need to account for the unique biology of eusocial insects. By systematically evaluating various GWAS scenarios, we provide practical guidelines for optimal p-value thresholds, model choice, and sequencing strategies, offering a framework to improve the design, reliability, and impact of future genetic studies in honey bees and other social insects.

Data availability

The datasets generated and/or analysed during the current study are available in the European Variation Archive (EVA) repository under project accession PRJEB112074. The associated analysis accession is ERZ29311177. The data are archived and will be publicly available from 1 May 2026.All analysis scripts and code used in this study are available on GitHub at: https://github.com/DrMulusewFikere/BeeGWASThese resources provide sufficient detail to reproduce the analyses presented in this study.

Received: 31 October 2025; Accepted: 28 April 2026
Published online: 01 May 2026

References

1. Liu, N. et al. Unraveling genetic mysteries: A comprehensive review of GWAS and DNA insights in animal and plant pathosystems. *Int. J. Biol. Macromol.* **285**, 138216 (2025).
2. Sharma, A. et al. Stories and Challenges of Genome Wide Association Studies in Livestock - A Review. *Asian-Australas J. Anim. Sci.* **28**, 1371–1379 (2015).
3. Visscher, P. M. et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
4. Aukema, J. E. et al. Economic Impacts of Non-Native Forest Insects in the Continental United States. *PLoS ONE.* **6**, e24587 (2011).
5. Losey, J. E. & Vaughan, M. The Economic Value of Ecological Services Provided by Insects. *BioScience* **56**, 311–323 (2006).
6. Saunders, M. E. & Rader, R. Ecosystem Services of Social Insects. in Encyclopedia of Social Insects (ed Starr, C. K.) 352–358 (Springer International Publishing, Cham, (2021).
7. Vidano, C. Social insects and the economy from an ecological point of view. *Ethol. Ecol. Evol.* **3**, 160–161 (1991).
8. Klein, A. M. et al. Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci* 274, 303 – 13 (2007).
9. Bratburd, J. R., Arango, R. A. & Horn, H. A. Defensive Symbioses in Social Insects Can Inform Human Health and Agriculture. *Front. Microbiol.* **11**, 76 (2020).
10. Cristancho, S. & Thompson, G. Building Resilient Healthcare Teams: Insights from Analogy to the Social Biology of Ants, Honey Bees and Other Social Insects. *Perspectives Med. Education* (2023).
11. Meunier, J. Social immunity and the evolution of group living in insects. *Philos Trans. R Soc. Lond. B Biol. Sci* **370** (2015).
12. Hotaling, S. et al. Long Reads Are Revolutionizing 20 Years of Insect Genome Sequencing. *Genome Biology Evolution* **13** (2021).
13. Li, F. et al. Insect genomes: progress and challenges. *Insect Mol. Biol.* **28**, 739–758 (2019).
14. Eynard, S. E. et al. Sequence-Based Multi Ancestry Association Study Reveals the Polygenic Architecture of Varroa destructor Resistance in the Honeybee Apis mellifera. *Mol. Ecol.* **34**, e17637 (2025).
15. Guichard, M. et al. Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences. *Anim. Genet.* **52**, 472–481 (2021).
16. Harpur, B. A. et al. Integrative Genomics Reveals the Genetics and Evolution of the Honey Bee's Social Immune System. *Genome Biol. Evol.* **11**, 937–948 (2019).
17. Eynard, S. E. et al. Reconstructing queen genotypes by pool sequencing colonies in eusocial insects: Statistical Methods and their application to honeybee. *Mol. Ecol. Resour.* **22**, 3035–3048 (2022).
18. Heidinger, I. M., Meixner, M. D., Berg, S. & Büchler, R. Observation of the Mating Behavior of Honey Bee (Apis mellifera L.) Queens Using Radio-Frequency Identification (RFID): Factors Influencing the Duration and Frequency of Nuptial Flights. *Insects* **5**, 513–527 (2014).
19. Tarpy, D. R., Delaney, D. A. & Seeley, T. D. Mating Frequencies of Honey Bee Queens (Apis mellifera L.) in a Population of Feral Colonies in the Northeastern United States. *PLOS ONE.* **10**, e0118734 (2015).
20. Yu, L., He, X., Shi, X., Yan, W. & Wu, X. Honey bee maternal effects improve worker performance and reproductive ability in offspring. *Front. Cell. Dev. Biol.* **11**, 1156923 (2023).
21. Holmes, L. A. et al. Queen quality, performance, and winter survival of imported and domestic honey bee queen stocks. *Sci. Rep.* **13**, 17273 (2023).
22. Saar, M., Eyer, P. A., Kilon-Kallner, T., Hefetz, A. & Scharf, I. Within-colony genetic diversity differentially affects foraging, nest maintenance, and aggression in two species of harvester ants. *Scientific Reports* **8** (2018).
23. Johnson, B. R. Division of labor in honeybees: form, function, and proximate mechanisms. *Behav. Ecol. Sociobiol.* **64**, 305–316 (2010).
24. Ramirez-Moreno, D., Galizia, C. G. & Nouvian, M. Division of labour during honeybee colony defence: poetic and scientific views. *Philosophical Trans. Royal Soc. B: Biol. Sciences* **380** (2025).
25. Maucourt, S., Fortin, F., Robert, C. & Giovenazzo, P. Genetic Parameters of Honey Bee Colonies Traits in a Canadian Selection Program. *Insects* 11 (2020).
26. Harpur, B. A., Kadri, S. M., Orsi, R. O., Whitfield, C. W. & Zayed, A. Defense Response in Brazilian Honey Bees (Apis mellifera scutellata × spp.) Is Underpinned by Complex Patterns of Admixture. *Genome Biol. Evol.* **12**, 1367–1377 (2020).
27. Beye, M. et al. Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* **16**, 1339–1344 (2006).
28. Jones, J. C., Wallberg, A., Christmas, M. J., Kapheim, K. M. & Webster, M. T. Extreme Differences in Recombination Rate between the Genomes of a Solitary and a Social Bee. *Mol. Biol. Evol.* **36**, 2277–2291 (2019).
29. Rueppell, O. et al. A New Metazoan Recombination Rate Record and Consistently High Recombination Rates in the Honey Bee Genus Apis Accompanied by Frequent Inversions but Not Translocations. *Genome Biol. Evol.* **8**, 3653–3660 (2016).
30. Chen, M. et al. Evidence of Polygenic Adaptation in Sardinia at Height-Associated Loci Ascertained from the Biobank Japan. *Am. J. Hum. Genet.* **107**, 60–71 (2020).
31. Maier, R. M., Visscher, P. M., Robinson, M. R. & Wray, N. R. Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychol. Med.* **48**, 1055–1067 (2018).
32. Uffelmann, E. et al. Genome-wide association studies. *Nat. Reviews Methods Primers.* **1**, 59 (2021).
33. Grozinger, C. M. & Zayed, A. Improving bee health through genomics. *Nat. Rev. Genet.* **21**, 277–291 (2020).
34. Adams, H. A., Southey, B. R., Robinson, G. E. & Rodriguez-Zas, S. L. Meta-analysis of genome-wide expression patterns associated with behavioral maturation in honey bees. *BMC Genom.* **9**, 503 (2008).
35. Avalos, A. et al. Genomic regions influencing aggressive behavior in honey bees are defined by colony allele frequencies. *Proc. Natl. Acad. Sci.* **117**, 17135–17141 (2020).
36. Saelao, P. et al. Genome-wide patterns of differentiation within and among US commercial honey bee stocks. *Bmc Genomics* **21** (2020).
37. Wallberg, A. et al. A hybrid de novo genome assembly of the honeybee, Apis mellifera, with chromosome-length scaffolds. *Bmc Genomics* **20** (2019).
38. Parejo, M. et al. AmelHap: Leveraging drone whole-genome sequence data to create a honey bee HapMap. *Sci. Data.* **10**, 198 (2023).
39. Wasik, K. et al. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genom.* **22**, 197 (2021).

40. Rubinacci, S., Hofmeister, R. J., Sousa da Mota, B. & Delaneau, O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat. Genet.* **55**, 1088–1090 (2023).
41. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
42. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
43. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).
44. Saber, M. M. & Shapiro, B. J. Benchmarking bacterial genome-wide association study methods using simulated genomes and phenotypes. *Microbial Genomics* **6** (2020).
45. Sesia, M., Bates, S., Candès, E., Marchini, J. & Sabatti, C. False discovery rate control in genome-wide association studies with population structure. *Proceedings of the National Academy of Sciences* 118, e2105841118 (2021).
46. Guzmán-Novoa, E., Hunt, G. J., Uribe, J. L., Smith, C. & Arechavaleta-Velasco, M. E. Confirmation of QTL effects and evidence of genetic dominance of honeybee defensive behavior:: Results of colony and individual behavioral assays. *Behav. Genet.* **32**, 95–102 (2002).
47. Tsuruda, J. M., Harris, J. W., Bourgeois, L., Danka, R. G. & Hunt, G. J. High-Resolution Linkage Analyses to Identify Genes That Influence Sensitive Hygiene Behavior in Honey Bees. *Plos One* **7** (2012).
48. Obsteter, J., Strachan, L. K., Bubnic, J., Presern, J. & Gorjanc, G. SIMplyBee: an R package to simulate honeybee populations and breeding programs. *Genetics Selection Evolution* **55** (2023).
49. Gaynor, R. C., Gorjanc, G. & Hickey, J. M. AlphaSimR: an R package for breeding program simulations. *G3-Genes Genomes Genetics* **11** (2021).
50. Brascamp, E. W., Willam, A., Boigenzahn, C., Bijma, P. & Veerkamp, R. F. Heritabilities and genetic correlations for honey yield, gentleness, calmness and swarming behaviour in Austrian honey bees. *Apidologie* **47**, 739–748 (2016).
51. Kistler, T. et al. Heritability and correlations for honey yield, handling ease, brood quantity, and traits related to resilience in a French honeybee population. *Apidologie* **55**, 52 (2024).
52. Bernstein, R. et al. First large-scale genomic prediction in the honey bee. *Heredity* **130**, 320–328 (2023).
53. Bienefeld, K., Ehrhardt, K. & Reinhardt, F. Genetic evaluation in the honey bee considering queen and worker effects — A BLUP-Animal Model approach. *Apidologie* **38**, 77–85 (2007).
54. Nearman, A. & vanEngelsdorp, D. Water provisioning increases caged worker bee lifespan and caged worker bees are living half as long as observed 50 years ago. *Sci. Rep.* **12**, 18660 (2022).
55. VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
56. Gilley, D. C., Tarpy, D. R. & Land, B. B. Effect of queen quality on interactions between workers and dueling queens in honeybee (Apis mellifera L.) colonies. *Behav. Ecol. Sociobiol.* **55**, 190–196 (2003).
57. Wei, H. et al. A Maternal Effect on Queen Production in Honeybees. *Curr. Biol.* **29**, 2208–2213e3 (2019).
58. Butler, D. G., Cullis, B. R., Gilmour, A. R. & Gogel, B. J. & R., T. ASReml estimates variance components under a general linear mixed model by residual maximum likelihood (REML) ASReml-R Version 4.2. (2023).
59. Harrell, F. E. & Davis, C. E. A New Distribution-Free Quantile Estimator. *Biometrika* **69**, 635–640 (1982).
60. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125**, 279–284 (2001).
61. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U S A*. **100**, 9440–9445 (2003).
62. Tang, Y. & Liu, X. L. G2P: a Genome-Wide-Association-Study simulation tool for genotype simulation, phenotype simulation and power evaluation. *Bioinformatics* **35**, 3852–3854 (2019).
63. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *Plos Genetics* **9** (2013).
64. Kazeem, G. R. & Farrall, M. Integrating case-control and TDT studies. *Ann. Hum. Genet.* **69**, 329–335 (2005).
65. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: a review. *Plant. Methods*. **9**, 29 (2013).
66. Yongtao, G. & Matthew, S. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Annals Appl. Stat.* **5**, 1780–1815 (2011).
67. Khanzadeh, H., Hossein-Zadeh, G., Ghovvati, S. & N. & Statistical power and heritability in whole-genome association studies for quantitative traits. *Meta Gene*. **28**, 100869 (2021).
68. Ma, C. A. & Li, J. K. Characterization of the mitochondrial genome of a high royal jelly-producing honeybee strain. *Mitochondrial DNA Part. B-Resources*. **6**, 2939–2940 (2021).
69. Xiang, R., van den Berg, I., MacLeod, I. M., Daetwyler, H. D. & Goddard, M. E. Effect direction meta-analysis of GWAS identifies extreme, prevalent and shared pleiotropy in a large mammal. *Commun. Biology*. **3**, 88 (2020).
70. Sheffer, M. M. et al. Chromosome-level reference genome of the European wasp spider Argiope bruennichi: a resource for studies on range expansion and evolutionary adaptation. *GigaScience* 10 (2021).
71. Ward, P. S. et al. Reference genome of the bicolored carpenter ant, Camponotus vicinus. *J. Hered.* **115**, 120–129 (2023).
72. Fikere, M. et al. Meta-analysis of GWAS in canola blackleg (Leptosphaeria maculans) disease traits demonstrates increased power from imputed whole-genome sequence. *Sci. Rep.* **10**, 14300 (2020).
73. Bolormaa, S. et al. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.* **10**, e1004198 (2014).
74. Zeggini, E. & Ioannidis, J. P. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191–201 (2009).
75. Adams, H. A., Southey, B. R., Robinson, G. E. & Rodriguez-Zas, S. L. Meta-analysis of genome-wide expression patterns associated with behavioral maturation in honey bees. *BMC Genom.* **9**, 503 (2008).
76. Traniello, I. M. et al. Meta-analysis of honey bee neurogenomic response links Deformed wing virus type A to precocious behavioral maturation. *Sci. Rep.* **10**, 3101 (2020).
77. Spötter, A., Gupta, P., Mayer, M., Reinsch, N. & Bienefeld, K. Genome-Wide Association Study of a Specific Defense Behavior in Honeybees. *J. Hered.* **107**, 220–227 (2016).
78. Traniello, I. M. et al. Molecular analyses of individual variation in honey bee sociability. *bioRxiv*, 12.22.629994 (2025). (2024).
79. Chen, W. et al. Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. *Nat. Commun.* **12**, 7117 (2021).
80. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
81. Abo-Ismail, M. K. et al. Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genet. Sel. Evol.* **49**, 82 (2017).
82. Jannot, A. S., Ehret, G. & Perneger, T. $P < 5 \times 10-8$ has emerged as a standard of statistical significance for genome-wide association studies. *J. Clin. Epidemiol.* **68**, 460–465 (2015).
83. Di Leo, G. & Sardanelli, F. Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *Eur. Radiol. Experimental*. **4**, 18 (2020).
84. Ottwell, R. et al. Assessing the Effects of Lowering the P Value Threshold From.05 to. 01 in Randomized Clinical Trials Published in Top Dermatology Journals: A Cross-Sectional Analysis. *J. Am. Acad. Dermatol.* **89**, Ab17–Ab17 (2023).
85. Wasserstein, R. L. & Lazar, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. *Am. Stat.* **70**, 129–131 (2016).

86. Zhang, M. et al. Genetic variants underlying differences in facial morphology in East Asian and European populations. *Nat. Genet.* **54**, 403–411 (2022).
87. Gabel, M., Hoppe, A., Scheiner, R., Obergfell, J. & Büchler, R. Heritability of Apis mellifera recapping behavior and suppressed mite reproduction as resistance traits towards Varroa destructor. *Front. Insect Sci.* **3**, 1135187 (2023).
88. Gowda, K. B., Jerry, D. R. & Zenger, K. R. Genetic improvement of farmed insect species: programmes, progress, and prospects. *Journal Insects as Food Feed*, 1–28 (2025).
89. Harbo, J. R. & Harris, J. W. Heritability in Honey Bees (Hymenoptera: Apidae) of Characteristics Associated with Resistance to Varroa jacobsoni(Mesostigmata: Varroidae). *J. Econ. Entomol.* **92**, 261–265 (1999).
90. Guichard, M. et al. Estimates of genetic parameters for production, behaviour, and health traits in two Swiss honey bee populations. *Apidologie* **51**, 876–891 (2020).
91. Spivak, M. Honey bee hygienic behavior and defense against Varroa jacobsoni. *Apidologie* **27**, 245–260 (1996).
92. Gabel, M., Hoppe, A., Scheiner, R., Obergfell, J. & Büchler, R. Heritability of Apis mellifera recapping behavior and suppressed mite reproduction as resistance traits towards Varroa destructor. *Frontiers Insect Science* **3** (2023).
93. Silva, M. R. G. et al. A meta-analysis of genome-wide association studies to identify candidate genes associated with feed efficiency traits in pigs. *Journal Anim. Science* **103** (2025).
94. Heimpel, G. E. & de Boer, J. G. Sex determination in the hymenoptera. *Annu. Rev. Entomol.* **53**, 209–230 (2008).
95. D Hamilton, W. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156**, 477–488 (1967).
96. Slater, G. P., Dapper, A. L. & Harpur, B. A. Haploid and Sexual Selection Shape the Rate of Evolution of Genes across the Honey Bee (Apis mellifera L.) Genome. *Genome Biol. Evol* **14** (2022).
97. Huang, K. et al. Estimating pairwise relatedness between individuals with different levels of ploidy. *Mol Ecol Resour* 15, 772 – 84 (2015).
98. Soro, A., Quezada-Euan, J. J. G., Theodorou, P., Moritz, R. F. A. & Paxton, R. J. The population genetics of two orchid bees suggests high dispersal, low diploid male production and only an effect of island isolation in lowering genetic diversity. *Conserv. Genet.* **18**, 607–619 (2017).
99. Blackmon, H., Hardy, N. B. & Ross, L. The evolutionary dynamics of haplodiploidy: Genome architecture and haploid viability. *Evolution* **69**, 2971–2978 (2015).
100. Fikere, M. et al. Accounting for population structure in genomic prediction of strawberry sweetness at a global scale. *Sci. Rep.* **15**, 40547 (2025).
101. Naret, O. et al. Correcting for Population Stratification Reduces False Positive and False Negative Results in Joint Analyses of Host and Pathogen Genomes. *Frontiers Genetics* Volume 9–2018 (2018).
102. Carpenter, M. H. & Harpur, B. A. Genetic past, present, and future of the honey bee (*Apis mellifera*) in the United States of America. *Apidologie* **52**, 63–79 (2021).
103. Liao, J. et al. Impact of measurement error on testing genetic association with quantitative traits. *PLoS One.* **9**, e87044 (2014).
104. Bernstein, R., Du, M., Hoppe, A. & Bienefeld, K. Simulation studies to optimize genomic selection in honey bees. *Genet. Sel. Evol.* **53**, 64 (2021).
105. Kistler, T., Basso, B. & Phocas, F. A simulation study of a honeybee breeding scheme accounting for polyandry, direct and maternal effects on colony performance. *Genet. Selection Evol.* **53**, 71 (2021).
106. Plate, M., Bernstein, R., Hoppe, A. & Bienefeld, K. Long-Term Evaluation of Breeding Scheme Alternatives for Endangered Honeybee Subspecies. *Insects* 11 (2020).
107. Grüter, C. & Hayes, L. Sociality is a key driver of foraging ranges in bees. *Curr. Biol.* **32**, 5390–5397e3 (2022).
108. Wild, B. et al. Social networks predict the life and death of honey bees. *Nat. Commun.* **12**, 1110 (2021).
109. Brascamp, E. W. & Bijma, P. Methods to estimate breeding values in honey bees. *Genet. Selection Evol.* **46**, 53 (2014).
110. Plate, M., Bernstein, R., Hoppe, A. & Bienefeld, K. The importance of controlled mating in honeybee breeding. *Genet. Selection Evol.* **51**, 74 (2019).
111. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *Am. J. Hum. Genet.* **110**, 179–194 (2023).
112. Chen, H. S. et al. Genetic simulation tools for post-genome wide association studies of complex diseases. *Genet. Epidemiol.* **39**, 11–19 (2015).
113. Fortune, M. D. & Wallace, C. simGWAS: a fast method for simulation of large scale case-control GWAS summary statistics. *Bioinformatics* **35**, 1901–1906 (2019).
114. Peng, B. & Amos, C. I. Forward-time simulation of realistic samples for genome-wide association studies. *BMC Bioinform.* **11**, 442 (2010).

Acknowledgements

We thank Amos Buschkoetter, the lab technician at the Harpur Lab, for his assistance in organizing the samples for sequencing. We would like to acknowledge USDA AFRI (2023-67013-38997), FFAR New Innovator and PDA for funding the project.

Author contributions

BAH , **LFB** , and **GG** contributed to the conceptualization of the research, led the project, and **BAH** supervised the study. **GG** and **MF** designed the study and defined the modeling framework. **MF** wrote a code, performed all the data analyses, generated the visualizations, and drafted the manuscript. **BAH** , **JO** , **LFB** , and **DR** reviewed, edited, and contributed to the final version of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the USDA Agriculture and Food Research Initiative (AFRI; Grant No. 2023-67013-38997), the Foundation for Food & Agriculture Research (FFAR) New Innovator Award, and the Purdue Department of Agriculture (PDA).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-51476-0.

Correspondence and requests for materials should be addressed to M.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026