625. Assessment of uncertainty in direct and maternal genetic parameters estimates for honeybee colony phenotypes

T. Kistler[1,2*], E.W. Brascamp[3], B. Basso[4,2], P. Bijma[2] and F. Phocas[1,2]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]UMT PrADE, France; [3]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [4]INRAE, UR 406 Abeille et Environnement, 84914 Avignon, France; tristan.kistler@inrae.fr

Abstract

Genetic parameters in honeybees are commonly estimated using REML methodologies applied to animal models with maternal effects. These methodologies were adapted to the species' peculiarities, including: phenotypes measured on colonies, haplo-diploïdy and polyandrous mating. However, estimations' reliability is hindered by the small size of the breeding nuclei commonly used. We assessed the uncertainty of variance component estimates in small simulated honeybee populations by evaluating the impact of the breeding nucleus size, the mating strategy and the direct-maternal genetic correlation. The convergence of estimations was strongly hindered when considering a small breeding nucleus with less controlled mating and negatively correlated genetic values. Furthermore, biases could be observed in this scenario. Individual estimates deviated strongly from the real values in about 40% of the populations. When considering real breeding populations, these results highlight the caution to be taken with estimates from small populations with complex pedigree structures.

Introduction

Most of economically important traits in honeybees are not recorded on an individual scale, but at the scale of the colony. A colony consists of a single diploid queen and thousands of sterile female workers that descended from the queen and from 10 to 20 haploid drones that have mated with the queen. The performance of a colony is affected by the joint genetic effects of all workers (called worker or direct effect) and of the queen (called queen or maternal effect). Polyandry and haplo-diploid sex determination are biological specificities that require an appropriate relation matrix for honeybee genetic evaluation. Efficient methodologies based on BLUP and REML applied to animal models are widely used to estimate breeding values in farm animals and have been adapted for honeybees (Bienefeld *et al.*, 2007; Brascamp and Bijma, 2014).

The phenotypic variance of colony performance can be partitioned into an environmental variance, an additive direct genetic variance, a maternal genetic variance and a potential covariance. A reliable estimation of these genetic variances as well as of the covariance between direct and maternal effects requires to get data and pedigree records in a large population constituted from genetically well-connected apiaries. Unfortunately, most of honeybee breeding programs are applied to small nucleus populations, usually ten to a few tens of breeding queens. This is different from livestock species for which seedstock includes hundreds or thousands of animals. Therefore, the main aim of our study was to assess the uncertainty in estimates of genetic parameters for polygenic traits measured in small simulated honeybee populations. In addition, we studied the impact of two mating designs on these estimates. In the first design, the drones fertilizing a virgin queen came from a single drone-producing queen (DPQ) that can be considered as a 'single diploid sire' (SS) of the workers. In the second design, the drones came from a group of sister DPQs, and this group of DPQs can be treated as a 'dummy diploid sire', referred to as a pseudo-sire (PS).

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

DOI: 10.3920/978-90-8686-940-4_625, © T. Kistler *et al.* 2022

Materials & methods

Population structure and stochastic simulations. Performance and breeding values of honeybee populations were simulated at the scale of individual queens, drones and worker groups. In the base population, queens' breeding values were drawn from a normal distribution $N(0, \Sigma^2_{BV})$ with Σ^2_{BV} the genetic variance-covariance matrix with both direct and maternal variances equal to 10 and a covariance assumed either to be 0 or -5. The base drones' breeding values were drawn from $N(0, \frac{1}{2}\Sigma^2_{BV})$. The only identified fixed effect was a year effect drawn from $N(0, 20)$ and residual effects on the performances were drawn from $N(0, 30)$.

Ten generations of within-line (one new breeding queen per maternal sister group) phenotypic selection were run, starting from non-inbred and unrelated breeding queens (either 12 or 24 BQs for a small and larger nucleus size scenario) and unrelated drones in the base population. Two-third of the same pool of maternal lines were selected each year to form DPQ families. Eight DPQs were then randomly chosen per selected family.

Two mating strategies were used: either sister DPQs formed a PS, jointly producing the drones mating a single potential dam queen with a random participation of each sister to the drone pool, or only one SS was randomly chosen to produce all the drones mating a single potential dam queen (controlled mating). All DPQs were mated to drones coming from a wild population (open-mating) with the same expected breeding values than those of the base population. Each queen (BQ or DPQ) was mated to 8 drones. Each year, wintering mortality was modelled by randomly eliminating 25% of all queens. Phenotypes were obtained after wintering, within the first year of birth and before reproduction of the dams. Both potential BQs and DPQs had phenotypes. Generation interval was 1.5 years (1 year on the maternal path and 2 years on the paternal path). Each BQ produced 24 potential BQs and 20 potential DPQs. Further details including genetic inheritance modelling can be found in Kistler *et al.* (2021). Each simulation scenario was repeated 100 times to assess the sampling standard deviations and the uncertainty of the estimates of genetic parameters.

Genetic evaluation model and estimation of genetic parameters. The vector of phenotypes y was described using a linear animal model accounting for maternal genetic effects: $y = Xb + Z_d a_d + Z_m a_m + e$, where b is the vector of fixed year effects with a corresponding incidence matrix X, a_d the vector of direct effects with incidence matrix Z_d, a_m the vector of maternal effects with incidence matrix Z_m, and e the vector of residual effects. Pedigree and performances of all simulated colonies were used for a single REML evaluation after 10 years of queen phenotypic selection. The total number of colony records was either 4,176 or 8,352, with 132 or 264 dam families for small or larger breeding nuclei, respectively.

BLUPf90 (Misztal *et al.*, 2002) with its AIREMLF90 program was used to estimate genetic parameters based on the inverse of a relationship matrix, following Brascamp and Bijma (2014, 2019). In this, it accounts for the DPQs being either a pseudo-sire, a single sire or a wild set of bees (open mating).

Results

Table 1 presents the 8 study scenarios. Around 90% of all estimations converged. However, PS mating hindered convergence, as well as a small nucleus size and a negative direct-maternal correlation (r_{dm}) to a lesser extent. Thus, only 2/3 of the 100 simulations lead to converged estimates of variance components in scenario PS12$^-$.

Among converged cases, scenario PS12$^-$ had a significantly biased σ^2_d estimate (P-value<5% using a t-test). Other scenarios showed tendencies to over- (or under-) estimate one or several (co)variances but without clear statistical significance.

Even among scenarios with mean estimations centred on the true values, individual estimates varied strongly around the real (simulated) values, with the nucleus size having the greatest impact (Table 1). Smaller nucleus scenarios had (co)-variances estimates deviating strongly in 10 to 20% more repetitions than bigger nucleus scenarios.

Strongly deviating estimates occurred more often for the direct variance (in 45% of all repetitions) than for the maternal (37%) or for their covariance (32%). To a lesser extent, PS mating compared to SS mating also increased the number of strongly deviating estimates, mostly for the maternal variance and the direct-maternal covariance.

For the large nucleus size scenarios with PS mating, Figure 1 shows the relation between direct and maternal variance deviations, coloured by the deviation of the direct-maternal covariance (strong negative deviations in lavender and strong positive ones in red). In both scenarios, a clear structure appears: when

Table 1. Summary statistics of differences between estimated and true genetic (co)variances in 100 replicates of 8 simulated scenarios. True variances σ^2_d and σ^2_m are equal to 10.

ID	r_{dm}	% converged estimates	Difference between estimated and true σ^2_d		Difference between estimated and true σ^2_m		Difference between estimated and true σ_{dm}	
			Mean (se)	% strong deviates[1]	Mean (se)	% strong deviates[1]	Mean (se)	% strong deviates[1]
PS12	0	78	0.19 (0.35)	51	-0.07 (0.33)	45	-0.20 (0.28)	40
PS24	0	89	0.12 (0.24)	38	-0.12 (0.26)	42	-0.31 (0.16)	21
SS12	0	94	-0.58 (0.30)	54	-0.05 (0.26)	37	0.17 (0.22)	39
SS24	0	98	-0.41 (0.21)	34	0.05 (0.19)	26	-0.15 (0.16)	18
PS12$^-$	-0.5	66	1.48 (0.42)	58	0.54 (0.32)	44	-0.65 (0.33)	48
PS24$^-$	-0.5	78	0.49 (0.29)	38	-0.13 (0.24)	38	-0.15 (0.24)	31
SS12$^-$	-0.5	97	-0.05 (0.31)	57	-0.22 (0.27)	42	0.20 (0.25)	40
SS24$^-$	-0.5	100	-0.36 (0.21)	30	-0.30 (0.15)	24	0.14 (0.15)	22

[1] Proportion of estimated variances deviating by more than 20% of the true values.

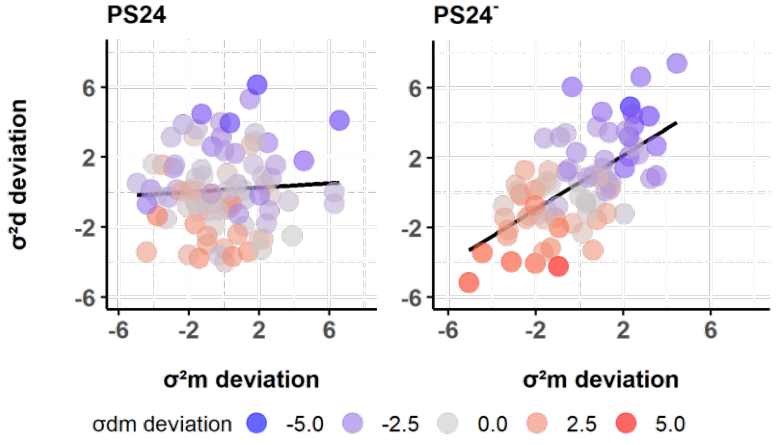

Figure 1. Estimated genetic parameters' deviations from true (simulated) values for two scenarios with a null (PS24) or negative (PS24$^-$) direct-maternal genetic correlation.

both direct and maternal variances were underestimated, the covariance was overestimated, and *vice-versa* for overestimated variances. In addition, in particular for the negative correlation scenario, when the direct variance was under- (over-) estimated, the maternal variance tended to be under- (over-) estimated as well.

Discussion

Very unreliable estimates of variance components were observed for scenarios combining a small size of the breeding nucleus, PS mating and a negative direct-maternal genetic correlation. In PS mating, true relationships between DPQs and descendants are unknown and have to be probabilistically accounted for, making estimates less precise. Additionally, direct and maternal effects are more difficult to disentangle when they are correlated, which can be seen in Figure 1. A last explanation that may explain some bias in the estimates of variance components lies in the different hypotheses considered for the contributions of various DPQs and drones to any queen's offspring in the simulation of the population and in the derivation of the relationship matrix. In our simulations, these contributions were balanced while they were drawn from a Poisson distribution in Brascamp and Bijma (2014)' relationship matrix.

Beyond the biased estimates in one scenario, the number of estimates deviating strongly from their true values was always very high. Using real datasets with very small populations sizes, Guichard *et al.* (2020) and Basso *et al.* (2022) could not estimate jointly direct and maternal effects for various honeybee traits. While 24 breeding queens in a nucleus population is currently considered as a decent size in honeybee breeding, it remains a small population for genetic parameters' estimation purposes, even with 10 years of records. Furthermore, we simulated a simplistic situation in which the whole population was composed of a unique apiary only submitted to a year effect. In real life, Clément *et al.* (2001) showed that low genetic connectedness between herds (apiaries) would induce severe biases in the estimated genetic parameters. A strategy to avoid these issues in genetic evaluation for small honeybee populations could be to use estimates of genetic parameters based on large datasets, such as in Hoppe *et al.* (2020), and to assume that they are very similar in the small population of interest.

References

Basso B., Kistler T., Gerez T. and Phocas F. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands, in press

Bienefeld K., Ehrhardt K., and Reinhardt F. (2007) Apidologie 38:77–85. http://dx.doi.org/10.1051/apido:2006050

Brascamp E.W., and Bijma P. (2014) Genet. Sel. Evol. 46:53. http://dx.doi.org/10.1186/s12711-014-0053-9

Brascamp E.W., and Bijma P. (2019) Proc. of the 46th Apimondia, Montreal, Canada Clément V., Bibé B., Verrier É., Elsen J.-M., Manfredi E. *et al.* (2001) Genet. Sel. Evol. 15;33(4):369. http://dx.doi.org/10.1186/1297-9686-33-4-369

Guichard M., Neuditschko M., Soland G., Fried P., Grandjean M. *et al.* (2020) Apidologie 51(5):876-891. http://dx.doi.org/10.1007/s13592-020-00768-z

Hoppe A., Du M., Bernstein R., Tiesler F. K., Kärcher M., and Bienefeld K. (2020) Insects 11(11):768. http://dx.doi.org/10.3390/insects11110768

Kistler T., Basso B., and Phocas F. (2021) Genet. Sel. Evol. 53(1):1-16. http://dx.doi.org/10.1186/s12711-021-00665-8

Misztal I., Tsuruta S., Strabel T., Auvray B., and Lee D. (2002). Proc. of the 7th WCGALP, Montpellier, France.