Title: Parameters influencing queen body mass and their importance as determined by machine learning in honey bees (*Apis mellifera carnica*)

Authors: Janez PREŠERN, Maja Ivana SMODIŠ ŠKERL

Affiliation: Agricultural Institute of Slovenia, Dept. of Animal Production,
Hacquetova ulica 17, SI-1000 Ljubljana, Slovenia

Corresponding author: Janez PREŠERN, janez.presern@kis.si

Short title: Parameters influencing queen body mass

Running title: Parameters influencing queen body mass

Abstract

Most parameters describing queen bee quality are reflected in the queen's body mass, which is in turn considered a robust measure and the best indicator of queen quality. State-of-the-art machine learning was used for the first time to jointly evaluate both biological and rearing parameters influencing queen body mass. Three different models were developed using different combinations of parameters. Regardless of the model composition, we achieved high precision of classification. The parameters "Ovary mass" and "Breeder" were the most important factors for model predictions. Differences in rearing practices and vegetation were masked by "Breeder", demonstrating the pitfall of this method. Separate analysis confirmed the importance of the time spent in the hive after mating and the phytogeographical region as an indirect indication of food sources. Rearing practices together with phytogeographical information are not enough to explain variation in queen body mass, yet they can contribute to the prediction of queen body mass if "Breeder" is excluded from the model.

Keywords

Queen body mass, parameter importance, machine learning

Introduction

There is an ongoing debate as to what defines a good queen bee and which parameters should be taken into account at the time of purchase. However, the beekeeper who wishes to purchase queen bees has no technical means to assess most of these parameters. On the other hand, the majority of these parameters play a role to some degree in queen body mass (for review, see Hatjina et al. 2014, Amiri et al. 2017). Incidentally, queen body mass is also a parameter that seems easy to measure as it requires only a scale in the milligram range.

Body mass varies throughout the life of the queen: it decreases with time after hatching and increases again after mating (Skowronek et al. 2004). The initial decrease in body mass is understandable in light of the mating flight, which affects mating success (Hayworth et al. 2009). Greater body mass improved queens' acceptance into another colony in *Apis mellifera anatoliaca* (Akyol et al. 2009). However, bioassays did not relate queens' body mass to their attractiveness to the worker bees (Nelson and Gary 1983; subspecies not given). Different practices used in queen rearing play a role in defining body mass. For example, larval age at the time of grafting has an important role in the development of reproductive organs such as ovaria (e.g., Gilley et al. 2003). Ovaria represent a significant part of the queen's abdomen and up to 40% of body mass in fertilized queens (calculated from data in Hatjina et al. 2014). Some authors report different numbers of ovarioles for queens grafted immediately after eclosion in comparison with queens grafted two or three days after eclosion, which again is reflected in the queen's body mass (Gilley et al. 2003, Woyke 1971), though opinions are divided on the topic (Hatch et al. 1999, Jackson et al. 2011). The mass of the ovaria is not stable even after the onset of oviposition (Kahya et al. 2008), and it is also dependent on season: winter "break" is reflected in the developmental stage of eggs and their number in the ovarioles (Shehata et al. 1981). Parameters that are often mentioned in connection with mating ability and offspring viability are sperm count and spermatheca volume, which are again reflected in body mass (Bieńkowska et al. 2009, Woyke 1987).

The effects of food sources on queen bees are difficult to study since they are fed indirectly by workers' retinue. Increased pollen flow is related to the production of worker bees (Mattila and Otis 2006), and winter pollen storage is correlated with the size of the spring population (Farrar 1936). The composition of royal jelly also depends on the available food sources (Echigo et al. 1986). It was observed, however, that the availability of pollen in the diet of workers influenced egg laying (Fine et al. 2018). One could also assume that diet directly influences the mass of ovaria.

The Slovenian breeding program for *Apis mellifera carnica* (SBP; Kozmus et al. 2018) binds commercial queen breeders with research institutions. A database formed through SBP activities and side projects contains various data regarding rearing, pedigree and performance testing. In this paper, we used machine learning (ML) procedures to delve into the relationship of body mass and several anatomical, (patho)physiological and rearing parameters, which are considered "queen quality parameters". ML is an approach for mining big data sets and using this "experience" for the prediction of new results. ML has been extensively used in bioinformatics, medicine, security and, recently, in animal behavior (Valletta et al. 2017), including modeling of the honeybee dance (Saghafi and Tsokos, 2017), and its use is still gaining momentum.

Using data collected over three years and ML procedures, we investigated the joint effect of the abovementioned parameters on the body mass of the queen bee and elucidated the most important among them. We discuss the results from the point of usefulness to the beekeeper.

Materials and methods

Queens

Queens (*Apis mellifera carnica*) used in morphological and physiological investigations were obtained from eighteen Slovenian commercial rearing operation stations in mid-June in 2006, 2008 and 2010. A total of 162 queens were collected every year, including nine queens per breeder; each of the queens was attended by 6 to 12 attendants. Additionally, 20 sister queens were measured in 2016 and added to newly formed nucs. Nucs were kept at the same location and expanded into full-size colonies in the next season. Brood surface was evaluated with a 5 cm x 5 cm mesh (Delaplane et al. 2013) in mid-May and mid-August 2017.

Anatomical and histological investigations and *Nosema* spp. spore quantification

Queens were anesthetized with CO_2 and weighed. The head, legs and wings were removed. The body of the queen was then pinned down with entomological needles, submerged in Hayes solution and dissected. The midgut, ovaries and spermatheca were carefully extracted.

Ovaries were weighed individually and then fixed in 4% formaldehyde. Dehydration was achieved with an ethanol queue (50%, 70%, 90%, 96%, 96%; 24 h for each step) and xylene (Sigma-Aldrich). Samples were then embedded in wax, and cross-sections were made at the ovary's midpoint with a microtome. Slices were dried on an object glass, deparaffinized in xylene, rehydrated in ethanol and stained in hematoxylin/eosin

(Sigma-Aldrich). Stained slices were investigated under the microscope, and ovarioles were counted for each ovarium (10 sections/ovary; Fig 1A).

To determine spermathecal volume, we first removed the spermathecal tracheal sheet and measured several spermathecal diameters under the microscope using the AxioVision program (Zeiss, Germany). Next, we calculated the spermathecal volume as the volume of a sphere using the average diameter as the entry parameter (Fig 1B). Moist spermatheca was punctured, and the sperm were transferred into a microcentrifuge tube containing 50 µl of Hayes solution. After 5 min, 950 µl of deionized water was added and kept for 10 min, followed by addition of 4 ml of fixative mixture (2 ml of a 4% solution of formaldehyde, 0.6 g 1 M $NaHCO_3$, and distilled water), according to Harizanis (1983). Spermatozoa counts were performed on a hemocytometer plate (Bürker-Türk); 80 fields were counted at 400x magnification. The number of spermatozoa in the spermatheca was calculated with the assumption that the sample volume inside a square of the hemocytometer is 0.004 mm^3 (1/250 mm^3):

Eq 1: $$N_{sperm} = \frac{mean\ sperm\ count}{square} \cdot dilution\ (5000) \cdot N_{fields}$$

Nosema spore presence was evaluated in the midgut of each queen. One milliliter of PBS (phosphate-buffered saline) was added to the sample and homogenized. A drop of homogenate was placed on a Bürker hemocytometer, and spores were counted. Attendant bee samples were pooled, and spore counts were obtained as described above. The spore count was then averaged over all attendants.

Queen rearing practices

Rearing parameters were collected with a questionnaire from each participating breeder. “Age at time of grafting” was either “egg”, larvae less than 12 h old, larvae between 12 and 24 h old and larvae older than 24 h. The parameter “Mating hive time” describes the time point at which the breeder removed the queens from the mating hive for shipment. The comb surface of the mating hive, contained in the parameter “Mating hive size”, was divided into three categories according to the summed surface of the comb(s). Types of grafting were described by the parameter “Grafting method” (Table 1).

Phytogeographical regions

Slovenia is divided into six phytogeographical regions: alpine, prealpine, submediterranean, dinaric, predinaric and subpannonic regions (roughly from west → east; Wraber 1969). These regions offer different forages to bees as a consequence of different abiotic parameters (e.g., altitude, soil, climate) that determine vegetation types and periods of nectar or dew flow. Every queen breeder was ascribed a region he/she belongs to, represented by the parameter “Phytogeog region”.

Data analysis

All the analyses were performed with custom-written Python 3 scripts using Scikit-Learn, Seaborn, Numpy and Scipy packages for analysis and graphical presentation. Models were built with the open-source machine learning software H_2O (H_2O.ai Inc., USA) via its Python API. Code is available at Zenodo (DOI: 10.5281/zenodo.3229393).

Relationships between parameters

Relationships between various parameters and between them and queen body mass were investigated. We used a simple linear regression and expressed the goodness of fit with R^2.

Data preparation, machine learning procedures and model evaluation

Data from eighteen breeders participating for all three years were used to build the datasets. We built three models with different combinations of parameters: 1) only data collected in all three years (model "2006 & 2008 & 2010"; $N_{\text{of queens}} = 486$; $N_{\text{of training}} = 413$; $N_{\text{of validation}} = 73$), e.g., without data on the number of ovarioles, data on the volume of spermatheca or *Nosema* presence in attendant bees; 2) only measurements collected in years 2006 and 2008 (model "2006 & 2008"; $N_{\text{of queens}} = 324$; $N_{\text{of training}} = 275$; $N_{\text{of validation}} = 49$), including the number of ovarioles but without spermatheca volume and *Nosema* in attendant bees; and 3) data collected in years 2008 and 2010, including spermatheca volume and *Nosema* in attendant bees but excluding the number of ovarioles (model "2008 & 2010", see Table 2; $N_{\text{of queens}} = 324$; $N_{\text{of training}} = 275$; $N_{\text{of validation}} = 49$). All three models included both phytogeographical data and data about rearing practices (Table 1). Rearing practices (Table 1), anatomical, physiological and health data (Table 2) were combined with phytogeographical region for each participating breeder. Body mass measurements were classified into quartiles: the 1st quartile represented high end body mass values, and the 4th quartile represented low-end body mass values. Quartiles were in turn used as target values in the model runs.

Modeling and machine learning procedures

We created several models based on the ML procedures to disentangle the complex relationships between several queen quality parameters. The “Gradient Boosting Machine” (GBM) algorithm from the open-source machine learning software H_2O was used in model creation, validation and determination of the importance of measured parameters, and we interfaced our analysis scripts via the Python API of the software. GBM was set to multiclass classification, predicting one of the four quartiles. The measured queen bee input parameters were treated as features in the model.

Briefly, the GBM algorithm in H_2O creates decision trees, which are constructed via an algorithmic approach that identifies ways to split the dataset at a node. Which feature to split on, and the split criteria are selected for each node, finding the greatest reduction in the residual sum of squares in the subtree at that point. We limited the number of trees to 50 for each run and the tree depth to 5 per tree; the number of bins per feature was set to 20, and the loss function was set to multinomial. No hyperparameters were set. Categorical features were encoded using the *enum* strategy. The model outputs a confusion matrix of correct vs. incorrect classifications and the relative predictive strength of each feature in the prediction task. This parameter importance score is normally expressed as the percent of contribution (Hastie et al., 2009). For the correct setup of the GBM algorithm, we followed the guidelines for use of ML in ecology (Elith et al. 2008).

In each iteration, the data were randomly split into a training set, consisting of 85% of the data and a validation set consisting of the remaining 15% of the data. The GBM learner was trained on the training set. The quality of the prediction was obtained by computing the precision ratio between correct classifications and total classifications and the error rate, which is the ratio between incorrect classifications and total classifications in the validation set. For each dataset, ten iterations were performed, and the results were pooled together and presented as the mean ± SD. Parameter importance was collected for each run, pooled with those from the other runs and presented for each dataset as the mean ± SD.

Results

Individual parameters and their impact on queen body mass

Prior to designing the model, we investigated the relationships between queen body mass and individual parameters. Most of the parameters did not have direct bearing on queen body mass, with the exception of ovary mass (Fig S1A), volume of spermatheca in 2010 (Fig S1D) and *Nosema* count in the gut of the queen (Fig S1E). It should be noted, however, that queen body mass in the infected subsample did not stand out of the sampled population. We performed a simple statistical test and confirmed no significant differences between infected and noninfected subsamples (N.S., unpaired t-test: $p = 0.92$; $t = -0.1$). Furthermore, we found no or a very weak relationship between the number of ovarioles and ovary mass (Fig S2A), between sperm count and spermatheca volume (Fig S2B), between ovary mass and sperm count (Fig S2C) and between ovary mass and volume of spermatheca (Fig S2D).

Impact of measured parameters on queen body mass

As mentioned above, the majority of measured parameters were collected every year, yet the datasets differ by the inclusion of one or another parameter depending on the year in the analysis (see Table 2). Building three different models allowed us to utilize all available data for each year and to compare the importance of the missing data.

Classifications were very good when no available parameter was withheld: the lowest mean precision of prediction was 0.84 (model “2006 & 2008 & 2010”; 2nd and 3rd quartile) and the highest was 0.97 (model “2006 & 2008”, 4th quartile). The mean misclassified fractions shown in the off-diagonal were between 0 and 0.07 (Fig 2B). The parameters “Ovary mass”, “Breeder” and “Sperm count” were constantly ranked as the most important parameters, with mean importance from 32-36%, 30-36% and 11-19%, respectively. Model “2006 & 2008” used the parameter “Ovarioles” (mean importance 10%), which improved the lowest average precision to 0.86 from 0.84 and the highest average precision to 0.97 from 0.95 (Fig 2A, B). Very good precision was achieved also by model “2008 & 2010” with a range of mean precisions between 0.88 and 0.95, which can be attributed to the extensive use of the parameters “Volume of spermatheca” (15.0 ± 2.0) and “*Nosema* sp. attendants” (6.0 ± 1.0%). The importance of both “Ovary mass” and “Breeder” was decreased to mean values of 32 and 30%, respectively, as a consequence. The importance of parameters related to rearing practices and phytogeographical region was valued below 0.5% regardless of the model (Fig 2A).

Importance of “Breeder” for model predictions

Rearing practices did not stand out in the model runs, and their importance was usually rated below 0.5%. We investigated the possibility that most of their informational value is already included in some other parameter, namely, “Breeder”. For that reason, we excluded the parameter “Breeder” and reran the model in the same manner as above. We noted an increase in the importance of these parameters to between 1.0 ± 0% and 6.0 ± 1.0% (Fig 3A). Despite mobilization of “neglected” parameters, there was also a marked drop in the precision of classification: for example, the mean precision of models “2006 & 2008 & 2010” and “2006 & 2008” dropped by 9% and 7%, respectively, in the prediction of the 2nd quartile (c.f. Fig 2B and Fig 3B). The presence of two additional parameters in the model “2008 & 2010” seem to compensate for the absence of the “Breeder”.

Importance of rearing practices and location for model predictions

To evaluate the importance of rearing practices and the location of the breeding operation, we excluded them from the model as well (besides “Breeder”). The present parameters increased in their importance as expected; for example, the importance of “Ovary mass” increased up to 11% (Fig 4A, top half), yet the precision of prediction decreased for first two models (by 19% max.). Exclusion of the rearing and location parameters had the least impact on the “2008 & 2010” model, which had two more parameters to start with (Fig 4B, c.f. Fig 3B).

We also performed classifications with rearing and phytogeographical parameters only. For all three models, the highest precisions of classification were for the 1st and 4th quartiles, which were between 0.48 and 0.71, both above randomness (0.25) but below the desired precision. Precision in the prediction of the other two quartiles was mostly below random for all three models. In fact, the 2nd and 3rd quartiles were incorrectly assigned into the 1st or 4th quartile at a rate greater than that by chance (Fig 4C), showing that the dataset used is not balanced. Phytogeographical region carried the highest importance in all three models (39 ± 4.0% - 41.0 ± 5.0%), followed by the time that a newly mated queen spends in her mating nuc (“Mating time hive”) and age at grafting (Fig 4A, bottom half). Despite their noted importance, the rearing parameters together with the phytogeographical data are not enough to explain the variation in body mass of the queen.

Discussion

The term “queen quality” can encompass several queen characteristics, which include genetic merit, developmental conditions, success in mating and, later, the environment in a (new) colony (Oldroyd et al. 1990; Dodologlu & Gene 2003). Queen body mass is one of these characteristics and is often regarded as a tool for the prediction of queen quality and, as such, is held in great esteem among beekeepers. In this paper, we turned the analysis around: instead of focusing on the body mass’ relationship with several descriptors of queen bee quality, which were empirically linked to brood production and overall colony health in the past, we investigated the contributions of these parameters to queen body mass. We show for the first time how these biological parameters and rearing practices influence the queen’s body mass, which often serves as the beekeeper’s tool for prediction of the queen’s performance before purchase or when selecting among queens.

Value of the parameters

In the past, parameters influencing body mass were often studied individually (for review, see Hatjina et al. 2014) or jointly via methods such as PCA to determine the anatomical and physiological parameters that best explained queen body mass (e.g., Tarpy et al. 2012). The combination of numeric features such as measured values of biological parameters and categorical features such as types of grafting required a new approach to evaluate the features’ joint importance.

Our data showed that a single parameter does not possess enough explanatory power to predict the body mass of the queen (Figs S1, S2). Dominating among “biological” parameters that steered classification was “Ovary mass”. Ovaria of the mated queen are approximately eight times larger than those of the virgin queen (Shehata 1981) and represent a large fraction of a queen's body mass and abdominal volume (Winston 1987). In our case, the median mass of ovaria differed between the studied years. It should be noted that the median body mass of the queens in our study also differed between years. However, the index between ovary mass and body mass also differed between years (Fig S1A), showing that ovarian growth does not entirely depend on the same parameters as body mass.

The parameters "N of Ovarioles", "Volume of spermatheca" and "*Nosema* sp. attendants" individually have a weak relationship with body mass. However, adding any of these three to the model significantly improved the models' performance, giving these parameters biological value. Mating triggers the growth of ovarioles (Tanaka and Hartfelder 2004) as a consequence of the expression of certain genes in both the ovaries and the brain, thereby inducing physiological changes (Kocher et al. 2008). We confirmed the absence of a correlation between the number of ovarioles and queen body mass (Fig S1B), as established by Hatch and colleagues (1999); the literature links the count of ovarioles to grafting age instead (Dedej et al. 1998, Tarpy et al. 2000). Both queens and workers are susceptible to infection with *Nosema* spp. The possible methods of infection are both horizontal (Higes et al. 2009) and vertical (Peng et al. 2016) with sperm. It was shown that in colonies with an infected queen, there is a greater proportion of infected workers (Czekońska 2000). The desire of beekeepers to obtain uninfected queens is therefore understandable. The regression plot shows that the severity of infection influences the queen's body mass, yet our sample is not great enough to confirm whether an infection would make infected queens stand out from the rest of the population (Fig S1E). Current statistical tests do not support such conclusions. It also seems that infected attendant bees are not the cause of infection in the queens; infected attendants were far more numerous than infected queens. However, in cases when attendants were infected, the spore count in the queen was higher (Fig S2E). According to Alaux et al. (2011), infection of queens with *Nosema ceranae* increased the level of vitellogenin, queen mandibular pheromone and antioxidant capacity. Atrophy of hypopharyngeal glands is one of the effects of *Nosema* infection in worker bees and supposedly the main reason the queen escapes infection (Wang and Mofller 1970).

Seasonal differences (“Year”) observed both in body and ovary mass were ranked as important but were overshadowed by “Breeder” in all three models. During model construction, we attempted to strip the rearing practices from the parameter “Breeder” and use them as separate model parameters. As mentioned above, none of them contributed significantly to the body mass in initial model runs. We found it curious that Tarpy and coworkers (2011) experimentally created high- and low-quality queens by grafting at different ages. Additionally, ontogenetically, body mass decreases following emergence and is at its lowest a day after the last mating flight, after which it increases back to its approximate value at the beginning of oviposition and gains an additional 10% over the next three days. After the onset of oviposition, the body mass decreases to somewhere between 5 and 10% more than the mass at emergence (Kahya et al. 2008). For both reasons, we expected a significant impact by “Mating hive time” and “Age at grafting” or at least a significant contribution by them.

The initial misleading results were the consequence of a caveat of the ML method used: only the parameters that contribute to the explanation of target values were considered, and all the information provided by the technical data was already included in the "Breeder" parameter. After "Breeder" was removed as a separate parameter, parameters covering rearing practices and phytogeography were mobilized to explain queen body mass. It seems that there is more to "Breeder" than just the rearing practices of the breeder and the vegetation at the breeding location; however, the classifications were still good but not as good as before. Two qualities that could remain entwined in the parameter "Breeder" are microlocation of the mating hives and nucs and the genetic lines with which the breeders work.

Regional information, which defines the time frame of various forages, contained under the "Phytogeographical region" was important in all cases after the exclusion of "Breeder", and the most important parameter when parameters covering breeding practices and phytogeographical information were tested separately. This highlights the importance of forage sources. Mao and coworkers (2015) showed that certain plant compounds such as p-coumaric acid, often found in beebread and honey, seem to inhibit the development of ovaria in worker bees. Similarly, plant miRNAs seem to play a role as well (Zhu et al. 2017). Due to the possibility of different dietary preferences of colonies at the same location (c.f. Waddington et al. 1994), it is probably impossible to tackle this issue with field observation and without the manipulation of colony feed stocks.

Conclusions

As a measure of queen quality, queen body mass is directly useful for the prediction of brood production, taking into account the large safety margin, shown as the range of the confidence interval, at the desired brood surface (Fig S2F). Our machine learning approach showed that body mass highly reflects both rearing parameters and production potential. We acknowledge that models do not reflect real biology, yet when their predictions have high precision and R^2 values, they support ideas about the synergistic effects of multiple factors. The parameters marked as important by the model could be masking other important parameters, which is probably the greatest weakness of the approach used. Our models show that higher body mass means favorable connection with at least one of the production-related parameters. However, the independence of parameters (other than “Ovary mass”) from the queen’s mass means they contribute to “body mass” on an individual basis, and there is no guarantee that a queen with a high body mass has a large number of ovarioles or that the sperm count in its spermatheca is high.

Selecting queens by body mass, however, should also be performed cautiously. It seems that considering absolute mass value as a threshold for queen quality is not a correct approach because measured masses varied between seasons, as shown in Fig S3A. Tarpy and coworkers (2012) found that variability within a rearing operation is higher than interoperation variability. Consequently, it was suggested that general queen quality could be improved by culling low-end queens before going to market. Beekeepers who wish to purchase queens are normally in no position to determine the average annual queen body mass and which breeder currently produces the heaviest queens; at best, he or she can make comparisons within the rearing operation. However, knowledge about the phytogeographical region of the operation and time spent in mating nucs might help. In some cases, it is possible to make use of breeders’ past production. In Slovenia, for example, queen quality is assessed yearly by taking samples from the breeders involved with the Slovenian Breeding Program to assist potential customers.

Queen bees' body mass and other "queenly" qualities have often been discussed in the literature, sometimes with opposing results. Our investigation is one of the few that also indirectly covers the rarely discussed impact of diet on the queens' body mass and production potential, which should be the focus of future research in this area.

Acknowledgments

We are indebted to the late Marjan Kokalj and to Mitja Nakrst, Andrej Blejec, Jernej Bubnič and Vesna Lokar. This work was supported by the Slovenian Research Agency (Research program P4–133) by the “Decree on implementing the apiculture programme in Slovenia” for periods covering the years 2006, 2008, 2010 and 2016 and by the “Slovenian Breeding program for *Apis mellifera carnica*”. Last but not least, we are grateful to the two anonymous reviewers.

Authors contribution

MISŠ performed the experimental work, and JP analyzed the results. Both authors wrote the paper.

Literature

1. Alaux C, Folschweiller M, McDonnell C, Beslay D and others. 2011. Pathological effects of the microsporidium *Nosema ceranae* on honey bee queen physiology (*Apis mellifera*). J Invertebr. Pathol. 106(3): 380 – 385.

2. Amiri E, Strand M, Rueppel O, Tarpy RD. 2017. Queen quality and the impact of honey bee diseases on queen health: potential for interactions between two major threats to colony health. Insects, 8(2): 48.

3. Akyol E, Yeninar H, Kaftanoglu O. 2009. Live weight of Queen Honey Bees (*Apis mellifera* L.) Predicts Reproductive Characteristics. J Kansas Entomo Soc 81: 92—100.

4. Bieńkowska M, Panasiuk B, Węgrzynowicz P, Gerula D. 2009. Weight of honey bee queens and ist effect on the quality of instrumentally inseminated queens. In *Proc of 41st International Apicultural Congress*, *Montpellier,* France.

5. Czekońska K. 2000. The influence of Nosema apis on young honeybee queens and transmission of the disease from queens to workers. Apidologie, 31(6): 701—706.

6. Dedej S, Hartfelder K, Aumeier P, Rosenkranz P, Engels W. 1998. Caste determination is a sequential process: Effect of larval age at grafting on ovariole number, hind leg size and cephalic volatiles in the honey bee (*Apis mellifera carnica*). J Api Res, 37(3): 183–190.

7. Delaplane KS, Van der Steen J, Guzman-Novoa E. 2013. Standard methods for estimating strength parameters of *Apis mellifera* colonies. J Api Res 52(1): 1—12.

8. Dodologlu A, Gene F. 2003. Comparison of some features of queens reared from different honeybee (*Apis mellifera* L.) genotypes. J App Anim Res, 24: 105–109.

9. Echigo T, Takenaka T, Yatsunami K. 1986. Comparative studies on chemical composition of honey, royal jelly and pollen loads. Bull Fac Agr Tamagawa Univ 26: 1–12.

10. Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. J Animal Eco 77(4): 802 – 813.

11. Farrar CL. 1936. Influence of pollen reserves on the surviving populations of over-wintered colonies. Am Bee J, 76: 452 – 454.

12. Fine JD, Shpigler HY, Ray AM, Beach NJ, Sankey AL and others. 2018. Quantifying the effects of pollen nutrition on honey bee queen egg laying with a new laboratory system. PloS One 13(9): e0203444.

13. Gilley DC, Tarpy DR, Land BB. 2003. Effect of queen quality on interactions between workers and dueling queens in honey bee (*Apis mellifera* L.) colonies. Beh Eco Socbio, 55(2), 190 – 196.

14. Harizanis PC. 1983. Biology and activity of honey bee sperm: migration, concentration and ATP analysis. PhD University of California, Davis, USA.

15. Hastie T, Tihshirani R, Friedman J. 2009. The elements of statistical learning. Data mining, inference, and prediction. Springer NY.

16. Hatch S, Tarpy DR, Fletcher DJC. 1999. Worker regulation of emergency queen rearing in honey bee colonies and the resultant variation in queen quality. Insectes Soc. 46: 372 – 377

17. Hatjina F, Bienkowska M, Charistos L, Chlebo R and others. 2014. A review of methods used in some European countries for assessing the quality of honey bee queens through their physical characters and the performance of their colonies. J Api Res 53(3): 337 – 363.

18. Hayworth MK, Johnson NG, Wilhelm ME, Gove RP, Metheny JD and Rueppel O. 2009. Added weigths lead to reduced flight behavior and mating success in polyandrous honey bee queens (*Apis mellifera*). Ethology 115(7): 698—706.

19. Higes M, Martin-Hernandez R, Garcia-Palencia P, Marin P, Meana A. 2009. Horizontal transmission of Nosema ceranae (Microsporidia) from worker honeybees to queens (*Apis mellifera*). Environ.Microbiol Rep 1: 495–498.

20. Jackson JT, Tarpy DR, Fahrbach SE. 2011. Histological estimates of ovariole number in honey bee queens, *Apis mellifera* reveal lack of correlation with other queen quality measures. J Ins Sci, 11(1): 82.

21. Kahya Y, Gençer HV and Woyke J. 2008. Weight at emergence of honey bee (*Apis mellifera caucasica*) queens and its effect on live weights at the pre and post mating periods. J Api Res, 47(2): 118 – 125.

22. Kocher SD, Richard FJ, Tarpy DR, Grozinger CM. 2008. Genomic analysis of post-mating changes in the honey bee queen (*Apis mellifera*). BMC Genomics, 9(1): 232.

23. Kozmus P, Podgoršek P, Smodiš Škerl MI, Prešern J, Metelko M and Hrastelj M. 2018. Rejski program za kranjsko čebelo (2018-2023): (*Apis mellifera carnica*, Pollmann 1879). Lukovica: Čebelarska zveza Slovenije; Ljubljana: Kmetijski inštitut Slovenije

24. Mao W, Schuler MA, Berenbaum MR. 2015. A dietary phytochemical alters caste-associated gene expression in honey bees. Sci Adv, 1: e1500795.

25. Mattila HR and Otis GW. 2006. The effects of pollen availability during larval development on the behavior and physiology of spring-reared honey bee workers. Apidologie, 37(5): 533 – 546.

26. Nelson DL and Gary NE. 1983. Honey productivity of honeybee colonies in relation to body weight, attractiveness and fecundity of the queen. J Api Res 22(4): 209—213.

27. Oldroyd BP, Goodman RD, Allaway MA. 1990. On the relative importance of queens and workers to honey production. Apidologie, 21: 153–159.

28. Peng Y, Grassl J, Millar AH, Baer B. 2016. Seminal fluid of honeybees contains multiple mechanisms to combat infections of the sexually transmitted pathogen *Nosema apis*. Proc R Soc B, 283(1823): 20151785

29. Saghafi A, Tsokos CP. 2017. Honey Bee Dance Modeling in Real-time using Machine Learning. arXiv:1705.07362v1 [stat.ML].

30. Shehata SM, Townsend GF, Shuel RW. 1981. Seasonal physiological changes in queen and worker honeybees. J Apic Res, 20: 69–78.

31. Skowronek W, Bienkowska M, Kruk C. Changes in body weight of honey bee queens during their maturation, J Api Sci, 48(2): 61–68.

32. Tarpy DR, Hatch S, Fletcher DJC. 2000. The influence of queen age and quality during queen replacement in honeybee colonies. Anim Behav, 59: 97–101.

33. Tarpy DR, Keller JJ, Caren, JR, Delaney DA. 2011. Experimentally induced variation in the physical reproductive potential and mating success in honey bee queens. Insectes Soc, 58: 569–574

34. Tarpy DR, Keller JJ, Caren JR, Delaney DA. 2012. Assessing the mating 'health' of commercial honey bee queens. J Econ Entomol, 105: 20–25.

35. Tanaka E, Hartfelder K. 2004. The initial stages of oogenesis and their relation to differential fertility in the honey bee (*Apis mellifera*) castes. Arthropod Struct Dev, 33(4): 431-42.

36. Valletta JJ, Torney C, Kings M, Thornton A, Madden J. 2017. Applications of machine learning in animal behaviour studies. Anim Beh 124: 203—220.

37. Waddington KD, Herbert TJ, Visscher PK, Richter MR. 1994. Comparisons of forager distributions from matched honey bee colonies in suburban environments. Beh Eco Socbio, 35(6): 423—429.

38. Wang D, Moeller FE. 1970. Comparison of the free amino acid composition in the hemolymph of healthy and *Nosema* -infected female honey bees. J Invert Patho, 15: 202–206.

39. Winston ML. 1987. The Biology of the Honey Bee; Harvard University Press: London, UK.

40. Woyke J. 1971. Correlations between the age at which honeybee brood was grafted, characteristics of the resultant queens, and results of insemination. J Api Res, 10(1): 45 – 55.

41. Woyke J. 1987. Can the number of ovarioles be estimated by external characters of living queens? Proc XXXIst Int Apicult Congress, Warsaw, Poland.

42. Wraber M. 1969. Pflanzengeographische Stellung und Gliederung Sloweniens. Vegetatio, 17(1-6): 176-199.

43. Zhu K, Liu M, Fu Z, Zhou Z and others. 2017. Plant microRNAs in larval food regulate honeybee caste development. PLoS Genet, 13(8): e1006946.

Figure captions

Figure 1. Cross-section of an ovary (A). Measuring volume of the spermatheca. Red lines show the lines of the diameter measurement (B).

Figure 2. A: Importance of individual parameters for classification of queens' body mass expressed in percent (mean ± SD). Empty fields indicate parameters not used during the model run. Body mass values were assigned to quartiles for all three years. B: Precision of classification for each model. Values on the diagonal of confusion show average precision of classification. Off-diagonal values show the fraction of misclassification. Red indicates values below or equal to chance (≤0.25), and green indicates values above chance (>0.25).

Figure 3. A: Importance of individual parameters, without the parameter "Breeder", for classification of queens' body mass, expressed as a percentage. Empty fields indicate parameters not used during the model run. Body mass values were assigned to quartiles for all three years. B: Precision of classification for each model without "Breeder". Values on the diagonal of confusion show precision of classification. Off-diagonal values show the fraction of misclassification. Red indicates values below or equal to chance (≤0.25), and green indicates values above chance (>0.25).

Figure 4. A: Importance of individual rearing parameters and vegetational parameters for precision of classification of queens' body mass, expressed as a percentage. White fields show parameters not used in the model's dataset. Body mass values were assigned to quartiles for all three years. The top half of the figure shows the importance

of individual parameters with rearing practices included and vegetational parameters excluded. The bottom half of the figure shows the importance of individual parameters with only rearing practices and phytogeographical parameters included. B: Precision of classification for individual models with rearing practices and vegetational parameters excluded. C: Precision of classification for individual models with only rearing practices and vegetational parameters included. The precision was not high enough to allow reliable predictions in any of the cases. In all confusion matrices, values on the diagonal show precision of classification. Off-diagonal values show the fraction of misclassification. The sum of the off-diagonal values shows the error rate. Red indicates values below or equal to chance (≤0.25), and green indicates values above chance (>0.25).

Tables

Table 1. Rearing practices used in the analysis with possible options. Green color labels selected options and gray color labels options never selected among the selected breeders.

Grafting method	Age at grafting	Mating hive time	Mating hive size (comb surface)
single	eggs	eggs	small (< 0.1 m^2)
double	larvae up to 12 h old	open brood	middle (0.1 $m^2 \leq$ 0.15 m^2)
Jenter/Nicot	larvae between 12 h and 24 h of age	covered brood	large (>0.15 m^2)
other	larvae more than 24 h old	hatching bees	

Table 2. Top: Overview of anatomical, physiological and health parameters measured in 2006, 2008 and 2010. Below: Inclusion of the same parameters in three different models.

“	Body mass	Breeder	Ovary mass	Ovarioles	Sperm count	Volume of spermatheca	*Nosema* sp. queen	*Nosema* sp. attendants	
2006	YES	YES	YES	YES	YES	NO	YES	NO	
2008	YES	YES	YES	YES	YES	YES	YES	YES	
2010	YES	YES	YES	NO	YES	YES	YES	YES	
MODEL NAME	PARAMETERS INCLUDED IN THE MODEL								**N of data**
2006 & 2008 & 2010	YES	YES	YES	NO	YES	NO	YES	NO	486
2006 & 2008	YES	YES	YES	YES	YES	NO	YES	NO	324
2008 & 2010	YES	YES	YES	NO	YES	YES	YES	YES	324

Figure 1

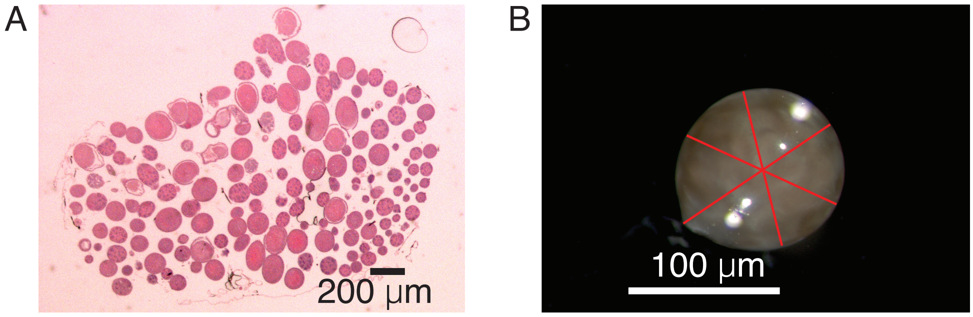


Figure 2

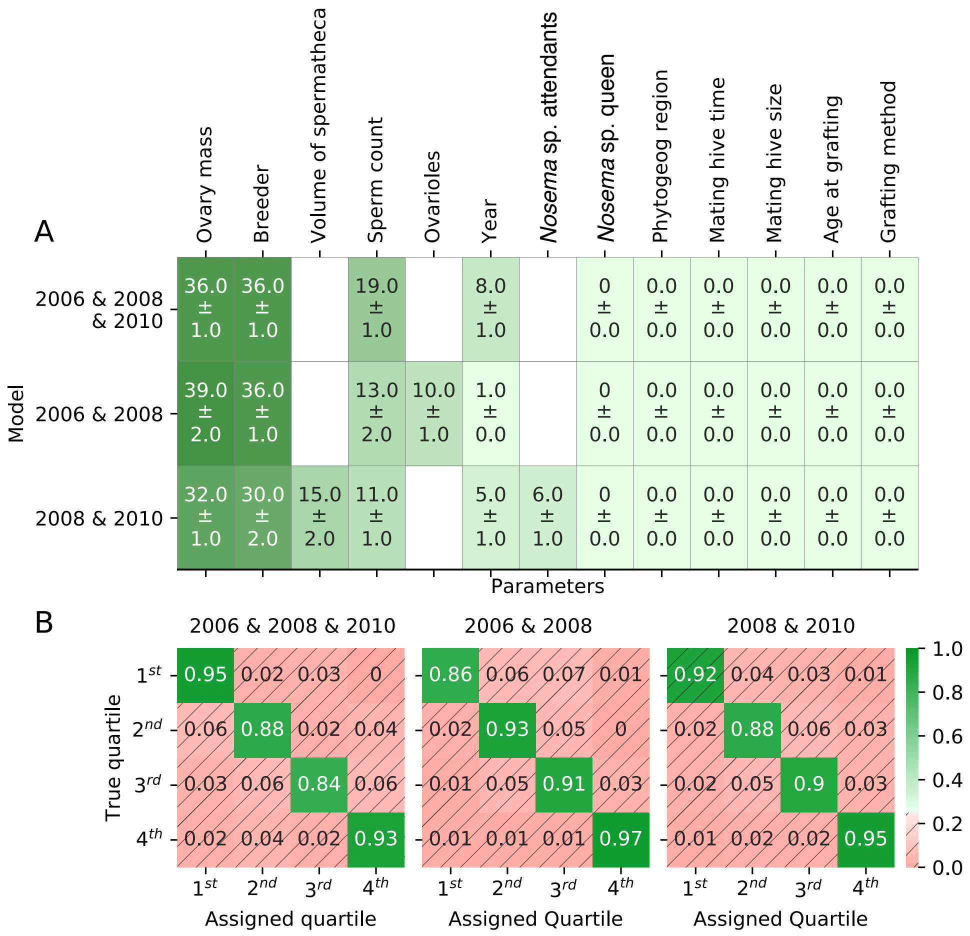

A
Model
Parameters
Ovary mass
Breeder
Volume of spermatheca
Sperm count
Ovarioles
Year
Nosema sp. attendants
Nosema sp. queen
Phytogeog region
Mating hive time
Mating hive size
Age at grafting
Grafting method
2006 & 2008 & 2010
36.0 ± 1.0
36.0 ± 1.0
19.0 ± 1.0
8.0 ± 1.0
0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
2006 & 2008
39.0 ± 2.0
36.0 ± 1.0
13.0 ± 2.0
10.0 ± 1.0
1.0 ± 0.0
0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
2008 & 2010
32.0 ± 1.0
30.0 ± 2.0
15.0 ± 2.0
11.0 ± 1.0
5.0 ± 1.0
6.0 ± 1.0
0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
0.0 ± 0.0
B
2006 & 2008 & 2010
2006 & 2008
2008 & 2010
True quartile
1st 0.95 0.02 0.03 0
2nd 0.06 0.88 0.02 0.04
3rd 0.03 0.06 0.84 0.06
4th 0.02 0.04 0.02 0.93
0.86 0.06 0.07 0.01
0.02 0.93 0.05 0
0.01 0.05 0.91 0.03
0.01 0.01 0.01 0.97
0.92 0.04 0.03 0.01
0.02 0.88 0.06 0.03
0.02 0.05 0.9 0.03
0.01 0.02 0.02 0.95
1st 2nd 3rd 4th
Assigned quartile
Assigned Quartile
Assigned Quartile
1.0
0.8
0.6
0.4
0.2
0.0

Figure 3

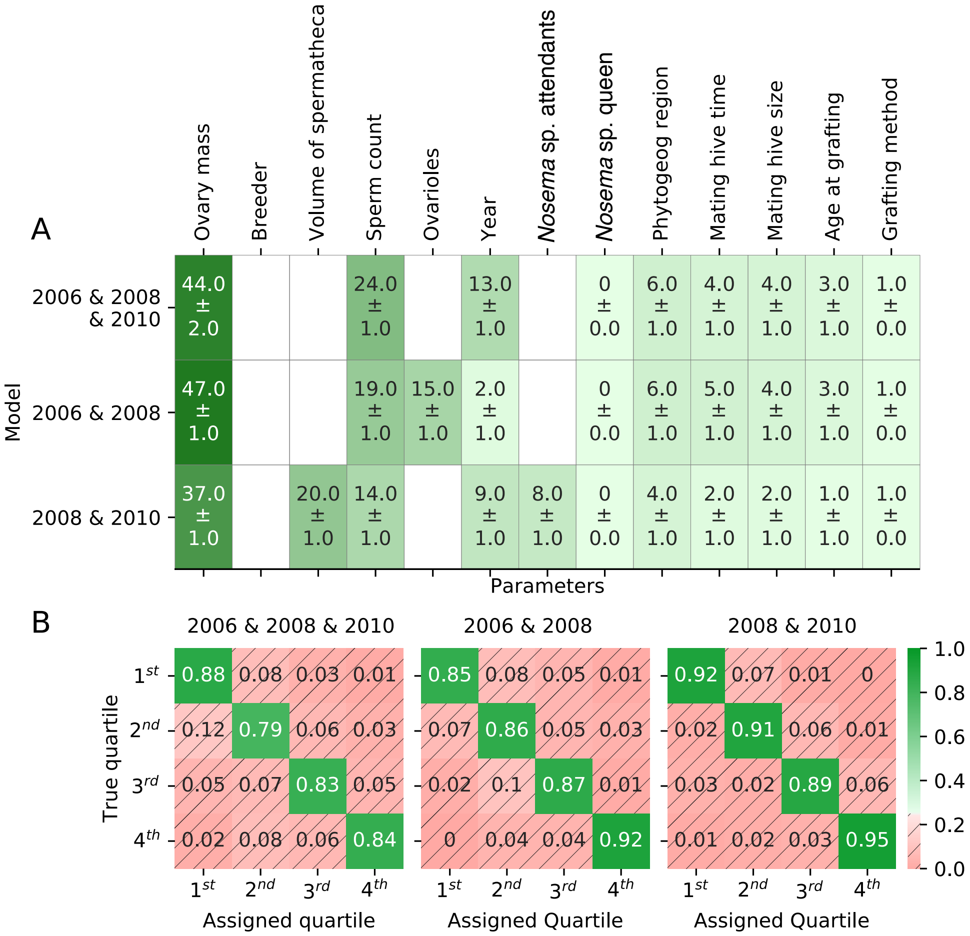
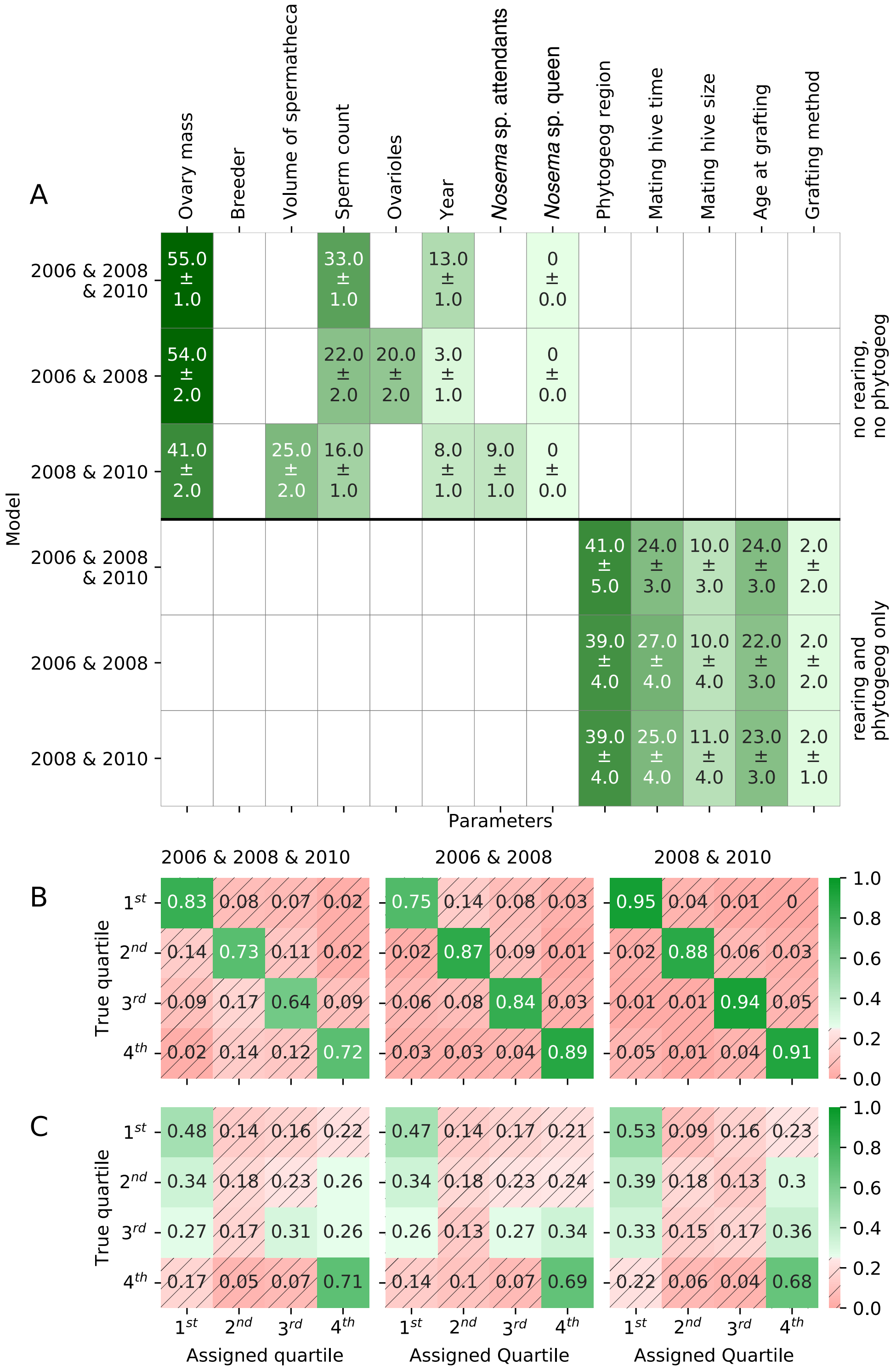

A
Ovary mass
Breeder
Volume of spermatheca
Sperm count
Ovarioles
Year
Nosema sp. attendants
Nosema sp. queen
Phytogeog region
Mating hive time
Mating hive size
Age at grafting
Grafting method
Model
2006 & 2008 & 2010
44.0 ± 2.0
24.0 ± 1.0
13.0 ± 1.0
0 ± 0.0
6.0 ± 1.0
4.0 ± 1.0
4.0 ± 1.0
3.0 ± 1.0
1.0 ± 0.0
2006 & 2008
47.0 ± 1.0
19.0 ± 1.0
15.0 ± 1.0
2.0 ± 1.0
0 ± 0.0
6.0 ± 1.0
5.0 ± 1.0
4.0 ± 1.0
3.0 ± 1.0
1.0 ± 0.0
2008 & 2010
37.0 ± 1.0
20.0 ± 1.0
14.0 ± 1.0
9.0 ± 1.0
8.0 ± 1.0
0 ± 0.0
4.0 ± 1.0
2.0 ± 1.0
2.0 ± 1.0
1.0 ± 1.0
1.0 ± 0.0
Parameters
B
2006 & 2008 & 2010
2006 & 2008
2008 & 2010
True quartile
1st
2nd
3rd
4th
0.88 0.08 0.03 0.01
0.12 0.79 0.06 0.03
0.05 0.07 0.83 0.05
0.02 0.08 0.06 0.84
0.85 0.08 0.05 0.01
0.07 0.86 0.05 0.03
0.02 0.1 0.87 0.01
0 0.04 0.04 0.92
0.92 0.07 0.01 0
0.02 0.91 0.06 0.01
0.03 0.02 0.89 0.06
0.01 0.02 0.03 0.95
1st 2nd 3rd 4th
Assigned quartile
Assigned Quartile
Assigned Quartile
1.0
0.8
0.6
0.4
0.2
0.0

Figure 4


A
Ovary mass
Breeder
Volume of spermatheca
Sperm count
Ovarioles
Year
Nosema sp. attendants
Nosema sp. queen
Phytogeog region
Mating hive time
Mating hive size
Age at grafting
Grafting method
Model
2006 & 2008 & 2010
2006 & 2008
2008 & 2010
2006 & 2008 & 2010
2006 & 2008
2008 & 2010
no rearing, no phytogeog
rearing and phytogeog only
55.0 ± 1.0
33.0 ± 1.0
13.0 ± 1.0
0 ± 0.0
54.0 ± 2.0
22.0 ± 2.0
20.0 ± 2.0
3.0 ± 1.0
0 ± 0.0
41.0 ± 2.0
25.0 ± 2.0
16.0 ± 1.0
8.0 ± 1.0
9.0 ± 1.0
0 ± 0.0
41.0 ± 5.0
24.0 ± 3.0
10.0 ± 3.0
24.0 ± 3.0
2.0 ± 2.0
39.0 ± 4.0
27.0 ± 4.0
10.0 ± 4.0
22.0 ± 3.0
2.0 ± 2.0
39.0 ± 4.0
25.0 ± 4.0
11.0 ± 4.0
23.0 ± 3.0
2.0 ± 1.0
Parameters
B
2006 & 2008 & 2010
2006 & 2008
2008 & 2010
True quartile
1st 0.83 0.08 0.07 0.02 0.75 0.14 0.08 0.03 0.95 0.04 0.01 0
2nd 0.14 0.73 0.11 0.02 0.02 0.87 0.09 0.01 0.02 0.88 0.06 0.03
3rd 0.09 0.17 0.64 0.09 0.06 0.08 0.84 0.03 0.01 0.01 0.94 0.05
4th 0.02 0.14 0.12 0.72 0.03 0.03 0.04 0.89 0.05 0.01 0.04 0.91
1.0
0.8
0.6
0.4
0.2
0.0
C
True quartile
1st 0.48 0.14 0.16 0.22 0.47 0.14 0.17 0.21 0.53 0.09 0.16 0.23
2nd 0.34 0.18 0.23 0.26 0.34 0.18 0.23 0.24 0.39 0.18 0.13 0.3
3rd 0.27 0.17 0.31 0.26 0.26 0.13 0.27 0.34 0.33 0.15 0.17 0.36
4th 0.17 0.05 0.07 0.71 0.14 0.1 0.07 0.69 0.22 0.06 0.04 0.68
1.0
0.8
0.6
0.4
0.2
0.0
1st 2nd 3rd 4th
Assigned quartile
Assigned Quartile
Assigned Quartile