**Unhealthy Brood Odor (UBeeO) Assay is strongly correlated with decreased virus, *Nosema* spp., and chalkbrood loads in honey bees (*Apis mellifera*).**

Samantha A. Alger, M. Sydney Miller, P. Alexander Burnham, Esmaeil Amiri, Corinne Jordan, Kaira Wagoner

*April 6, 2024*

**ABSTRACT**

Honey bee populations worldwide face numerous threats, including pests and pathogens. Selective breeding for traits associated with improved resistance to these stressors represents a promising approach for mitigating their impact on honey bee health. UBeeO, a novel hygiene eliciting selection tool, has previously been shown to identify colonies with reduced Varroa prevalence and loads in honey bees. Here, we show that UBeeO also identifies colonies with reduced RNA virus prevalence and loads, diminished chalkbrood severity, and lower *Nosema* spp*.* loads. Furthermore, the broad geographic range of our study sites underscores the robustness and applicability of UBeeO across varying environmental contexts. UBeeO-guided selection strategies have the potential to contribute significantly to honey bee breeding programs, fostering the development of resilient and pathogen-resistant bee populations.

**Statistical Analysis**

All data analyses were conducted in R (v4.2.3) (R Core Team 2019). All mixed models were built using the LME4 package (cite). Significance for all main effects and interaction terms was determined by conducting type II Wald chi-square tests using the "Anova" function in the CAR package (v3.1-0) (Fox and Weisberg 2019).

***Chalkbrood:*** To determine if the number of chalkbrood cell in a colony and chalkbrood spore type (white and black) were influenced by UBeeO score, we conducted a GLM with chalkbrood spore count as a function of chalkbrood type, UBeeO score, and their interaction term. As chalkbrood spore count was right-skewed integer count data, a Poisson distribution was used with a link “log” function.

***Nosema:*** In previous studies that examined the effect of UBeeO assay scores on Varroa mite loads, findings showed a high UBeeO response in colonies that had > 60% manipulated cells (cite). Here, we used this same metric to examine of there were differences in *Nosema* prevalence and load between high and low UBeeO testing colonies. For *Nosema* prevalence we examined a binary response variable (presence = 1, absence = 0) using a generalized linear mixed model (GLMM) with a binomial distribution and a link “logit” function. The binary UBeeO variable, month (June-September) and their interaction effect were used as predictors with apiary yard as a random control variable. The same model structure was used for the *Nosema* load data. Because *Nosema* load was skewed right, a Gamma distribution was used with a link “log” function. To identify significant UBeeO results over the four months of the study, we conducted pairwise comparisons on significant variables using the MULTCOMP package (cite).

In addition to testing UBeeO response as a binary variable, we examined *Nosema* loads response on a continuous UBeeO scale. For each of the four months of the study, we constructed a separate GLMM. All models followed the same structure with *Nosema* load modeled using a Gamma distribution with a link “log” function. UBeeO assay score (manipulated cells/total cells) was used as a predictor and apiary site as a random control variable. To test for an overall relationship between UBeeO score and *Nosema* load and to test for an effect of Month, we constructed a GLMM using the same response variable, distribution, and random effect as the models above. For this model, UBeeO assay score, Month, and their interaction term were used as predictors.

***RNA Viruses:*** Similar to the *Nosema* case study, we first examined Virus load and prevalence for six viruses (BQCV, DWV-A, DWV-B, IAPV, LSV, and SBV) by a binary UBeeO response variable with a 60% cell manipulation cutoff (cite). For both virus prevalence and load, separate models were created for each virus. For virus prevalence we tested a binary response variable (presence = 1, absence = 0) using a generalized linear model (GLM) with a binomial distribution and a link “logit” function. The binary UBeeO variable was the only predictor in each model. A similar model structure was used for virus load data (genome copies/ul) with the exception that a gamma distribution was used instead.

As in the *Nosema* case study, we also examined virus load on a continuous UBeeO scale. For each virus, we conducted a separate linear model. Virus load data were log10-transformed with continuous UBeeO score as the only predictor in each model. Significance for UBeeO score was assessed by conducting type II Wald chi-square tests using the "Anova" function in the CAR package (v3.1-0) (Fox and Weisberg 2019).

**RESULTS:**

**Chalkbrood**

When examining how the number and type of chalkbrood cells found in colonies changed in response to the UBeeO assay score of each colony we found that the number of spores detected was negatively correlated with increasing UBeeO score ( = 8.88, p = 0.003). Both black and white chalkbrood decreased with increasing UBeeO score, however, black spores decreased more steeply ( = 11.89, p < 0.001). Overall, White chalk cells were more common than black chalk spores ( = 342.47, p < 0.0001; Figure 1).

**Nosema**

We found that, high UBeeO colonies did not differ in overall *Nosema* prevalence from low UBeeO colonies (). However, we found an effect of month ( = 64.528, p < 0.0001) where *Nosema* prevalences where much higher in July (p = 0.003) and significantly lower in September (p < 0.0001; Figure 2a). *Nosema* loads were significantly lower in high UBeeO colonies ( = 10.861, p = 0.001) and a significant interaction effect indicates that loads decreased more quickly over time the high UBeeO group ( = 14.968, p = 0.002; Figure 2b). This, in part, seems to be driven by the significantly higher Nosema load measured for low UBeeO colonies in August (p < 0.0001).

We also tested *Nosema* load’s response to UBeeO on a continuous scale for each of the four months in the case study. Overall, controlling for month, we found that *Nosema* load decreases significantly with increased UBeeO assay score ( = 17.024, p < 0.0001). A significant UBeeO x Month interaction effect indicated that this relationship changed in slope for different months ( = 19.912, p < 0.001; Figure 3). In June, we found a negative trend ( = 3.277, p = 0.07) and a highly significant negative relationship in August ( = 15.274, p < 0.0001). No effect was detected for July or September ( < 1.3, p > 0.254).

**RNA Viruses**

We tested the prevalence and load of six RNA virus against high and low UBO colonies. We found a significantly lower prevalence in high UBeeO colonies for DWV-B ( = 7.651, p = 0.006) and LSV ( = 8.036, p = 0.005, Figure 4a). No significant difference between high and low UBeeO colonies was detected for BQCV, DWV-A, IAPV, or SBV ( < 1.857, p > 0.173, Figure 4a). For SBV, prevalence was 100% for UBeeO groups. For BQCV, prevalence was 100% for high UBeeO and 93.8% for low UBeeO indicating a high level of ubiquity for these two pathogens in our study apiaries. For load, we found significantly lower loads in high UBO colonies for three viruses: DWV-A ( = 4.517, p = 0.034), DWV-B ( = 10.392, p = 0.001), and IAPV ( = 9.621, p = 0.002; Figure 4b). We found evidence of a trend following a similar pattern for LSV ( = 3.333, p = 0.068) but no significant differences for BQCV or SBV ( < 1.577, p > 0.209).

In addition to the binary UBeeO variable (high and low colonies), we examined all size viruses by the continuous variable, UBeeO assay score. For four of the viruses, we found a significant negative relationship between virus load and UBeeO assay score: DWV-A (*F* = 7.944, p = 0.011), DWV-B (*F* = 22.14, p < 0.001), IAPV (*F* = 9.336, p = 0.007), and LSV (*F* = 18.797, p < 0.001; Figure 5). For the remaining two viruses (BQCV and SBV), we found no significant relationship (*F* < 0.373, p > 0.55).

**FIGURES:**

**Figure 1:** Prevalence and intensity of chalkbrood as a function of UBeeO score and spore type. Percent UBeeO score is shown on the x axis. Black chalkbrood (infective) spores are shown in black and white spores are shown in white. The black solid line and hatched line represent lines of best fit for black spores and white spores, respectively. The blue line represents the line of best fit for all spores, regardless of type.

**Figure 2:** *Nosema* prevalence and load by month for colonies that scored either high and low with the UBeeO assay. High UBeeO represents colonies scoring 60% or higher (blue) and low UBEEO represent colonies that scored less than 60% (orange). **(A)** *Nosema* prevalence (infected colonies/total colonies) by UBeeO score over time (months). Error bars represent a confidence interval generated by sampling from a binomial distribution parametrized by the number of infected (1) and uninfected (0) colonies for each UBeeO group at each month. **(B)** *Nosema* load (spores/bee) by UBeeO score over time (month). Error bars represent standard error.

**Figure 3:** *Nosema* load (spores/bee) by percent UBeeO score for June, July, August, and September. Green regression lines represent the line of best fit generated by a linear model.

**Figure 4:** Virus prevalence and load by virus and UBeeO status (low and high). High UBeeO is represented by blue and Low UBeeO by orange. Significance level for each virus grouping is shown above the bars. **(A)** Virus prevalence represents the number of infected samples over the total number of samples for each virus and UBeeO group multiplied by 100. Blue bars represent low UBeeO colonies, while orange bars represent high UBeeO colonies. Error bars represent a confidence interval generated by sampling from a binomial distribution parametrized by the number of infected (1) and uninfected (0) colonies for each virus and UBeeO group. **(B)** Virus load represents genome copies/bee and was log10 transformed. Horizontal black lines represent the median. Black dots represent outliers.

**Figure 5:** Virus load (genome copies/bee) by UBeeO score for 6 different viruses: BQCV, DWV-A, DWV-B, IAPV, LSV, and SBV. Green regression lines represent the line of best fit generated by a linear model.

**Figure 1**

****

**Figure 2**

****

**Figure 3**



**Figure 4**

****

**Figure 5**

