Environmental Drivers of *Eulaema nigrita* Abundance in the Atlantic Forest

Nadine Pigida

November 25, 2025

1 Introduction

Orchid bees (Euglossini) are vital pollinators in Neotropical ecosystems. Understanding the factors that drive their abundance is critical for conservation efforts, particularly in fragmented landscapes like the Atlantic Forest. *Eulaema nigrita* is a common species in this region, and we will use abundance data of this species to conduct our analysis.

In this report, we ask: How do climate, landscape composition, and sampling methodology influence the abundance of *E. nigrita*? Based on the species' known characteristics, we predict that, as a generalist species often associated with open areas, *E. nigrita* abundance may negatively correlate with dense forest cover. Furthermore, we expect that sampling effort will positively predict count data, and that abiotic factors, such as temperature and precipitation, will significantly structure the distribution of this species.

2 Analysis Methods

To assess the drivers of bee abundance, we analyzed abundance data of *E. nigrita* from 178 sampling sites. The response variable was the abundance count of *E. nigrita*. Predictors included sampling effort (log hours), altitude (m), mean annual temperature (MAT), mean annual precipitation (MAP), temperature seasonality, precipitation seasonality, the proportion of forest cover, and land use heterogeneity defined as the Shannon diversity of local land-use classes ($-\sum_i p_i \ln(p_i)$). MAT was recorded as degrees Celsius multiplied by 10, and MAP in millimeters. Temperature Seasonality was measured as 100 times the standard deviation of the monthly temperature, and precipitation seasonality was given as the coefficient of variation (CV) of monthly precipitation as a percentage. MAT and temperature seasonality were divided by 10 and 100, respectively, for better interpretation. We also included a factor for the sampling methodology, which consisted of three types: hand nets, traps, and both nets and traps.

Sampling effort varied substantially among sites (min = 1.609 log hours $\approx$ 5 hours, max = 10.020 log hours $\approx$ 22,000 hours). Raw bee counts vary among sites (Figure 1, left panel). In a model without accounting for effort, sampling time showed the strongest positive effect on observed bee counts (Parameter Estimate = 0.61, $p < 0.001$), indicating that a one standard deviation increase in sampling time corresponds to an 84% increase in expected counts. This percentage was calculated by back-transforming the parameter estimate from the log scale, using $(e^{\beta}-1)\times 100$. To make comparisons among sites meaningful, we therefore included sampling effort as an offset in the negative binomial model, to model bees per hour instead of raw counts (bees/hour; Figure 1, right panel).

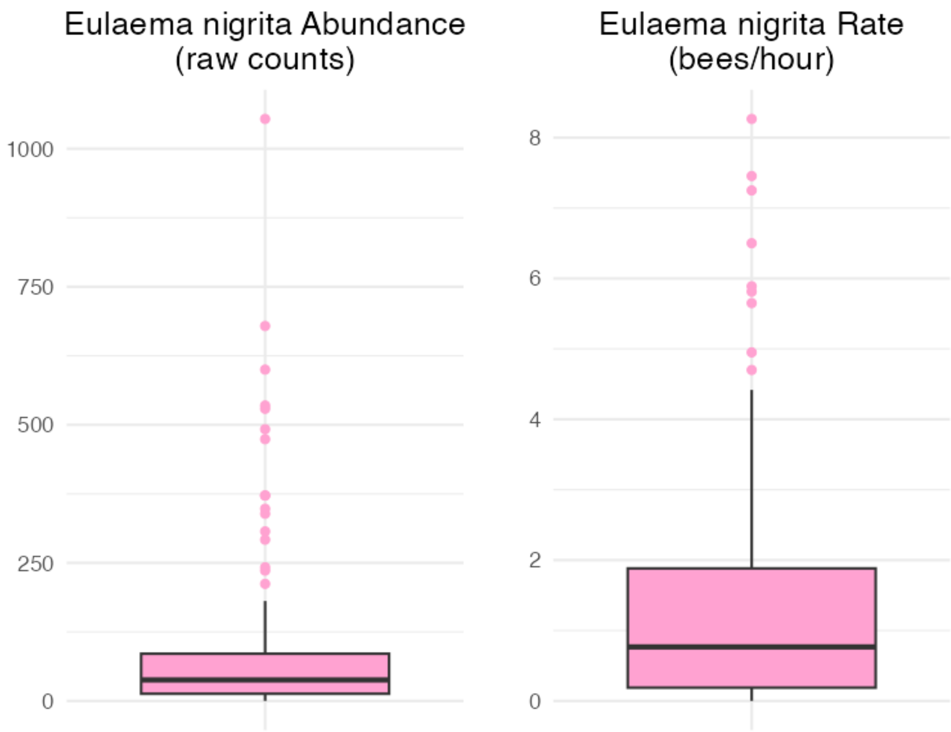

Figure 1: Data distribution for Eulaema nigrita abundance and outlier identification. Left panel: raw counts of bees per site. Right panel: predicted abundance standardized by sampling effort (bees/hour). The central line in each box represents the median (50th percentile), and the box shows the interquartile range (Q1–Q3, middle 50% of the data). Whiskers extend to 1.5× the interquartile range (IQR). Points beyond the whiskers indicate relatively high or low values compared to most sites. Standardizing by effort accounts for differences in sampling time, allowing fair comparison of bee abundance among sites.

The remaining predictors were scaled (standardized) to have a mean of zero and a standard deviation of one to allow direct comparison of effect sizes (Table 1). Netting was used as the reference category for the methodological factor. Then, we fitted a Generalized Linear Model (GLM). Initial Poisson modeling indicated significant overdispersion (variance > mean); consequently, we fitted a Negative Binomial model to account for the aggregated variance structure typical of count data. Then, we performed backward selection to derive the most parsimonious model. The Land Use Heterogeneity predictor was excluded during this selection process as it was found to be non-significant (Parameter Estimate $= -0.073$, $p = 0.3805$).

Thus, our final model was fitted to the data, and in R syntax took the form:

```
Eulaema_nigrita ~ altitude + MAT_C + MAP + Tseason_C + Pseason +
forest. + method + offset(effort)
```

3 Results

The Negative Binomial model successfully corrected for overdispersion (residual deviance / degrees of freedom ≈ 1.207). The final model demonstrated a strong fit, explaining 39.1% of the null deviance (Pseudo-$R^2 = 0.391$). Since the predictors were standardized, the estimates (β) in Table 1 represent the change in the log mean capture rate for a one standard deviation increase in the respective predictor. The percentage changes reported in the text were calculated from the standardized coefficients using the formula $(e^{\beta} - 1) \times 100$, which gives the proportional change in expected capture rate for a one standard deviation change in the predictor.

Climatic, Geographic, and Landscape Drivers

After accounting for detection method, the abundance of *Eulaema nigrita* was most strongly influenced by climatic factors. Temperature Seasonality (Tseason) had the largest negative effect: a one standard deviation increase in temperature variability was associated with a 54.8% decrease in expected capture rate ($\beta = -0.79477$, $p < 0.001$, Table 1). Similarly, higher Mean Annual Temperature ($\beta = -0.69997$, $p = 0.0123$ Table 1) and greater Mean Annual Precipitation (MAP; $\beta = -0.64296$, $p < 0.001$, Table 1) predicted 50.3% and 47.4% reductions in abundance, respectively.

In contrast, Precipitation Seasonality (Pseason; $\beta = 0.38154$, $p < 0.001$, Table 1) had a positive effect, with a one standard deviation increase in rainfall variability corresponding to a 46.5% increase in expected capture rate. This suggests that *E. nigrita* favors climates with distinct wet and dry periods. Geographic location also mattered: higher altitude was linked to lower abundance, with a one standard deviation increase in altitude predicting a 43.9% decrease in expected capture rate ($\beta = -0.57790$, $p = 0.0193$, Table 1).

Among the landscape predictors, Forest Cover had the smallest effect, yet still significant: sites with more forested habitat showed a 29.1% reduction in expected capture rate ($\beta = -0.34322$, $p < 0.001$, Table 1). Together, these results indicate that *E. nigrita* thrives in open or disturbed habitats and in climates with relatively stable temperatures but pronounced seasonal rainfall.

Sampling Methodology Effects

Detection method was included as a covariate in the model to account for differences in capture efficiency among sites. Using active Netting as the baseline category, passive Traps were significantly less effective (Estimate = -0.85075, $p < 0.001$, Table 1), reducing the expected capture rate by approximately 57.3% compared to the Netting baseline. This highlights the need to account for detection bias in biodiversity surveys. In contrast, combining both methods in the NetTraps approach did not significantly differ from the Netting baseline (Estimate = -0.39355, $p = 0.1140$, Table 1), suggesting that passive traps added little statistical value to the active search effort.

Table 1: **Standardized Parameter Estimates from the Final Negative Binomial GLM with Effort Offset.** Estimates (β) represent the expected change in the log mean capture rate (bees/hour) for a one standard deviation increase in the predictor. Std. Error is the standard error of the estimate. The z value is the Z-statistic (Estimate/SE). Pr($> |z|$) is the P-value associated with the Z-statistic. **Note:** The standardized effect of sampling effort is fixed at 1 and is excluded from the table as it is included as an offset in the model.

Parameter	**Estimate (β)**	**Std. Error**	**z value**	**Pr($> \|z\|$)**
(Intercept)	4.22135	0.10570	39.936	< 0.001
Altitude	-0.57790	0.24707	-2.339	0.0193
MAT	-0.69997	0.27945	-2.505	0.0123
MAP	-0.64296	0.09508	-6.763	< 0.001
Temp Seasonality	-0.79477	0.12957	-6.134	< 0.001
Precip Seasonality	0.38154	0.09243	4.128	< 0.001
Forest Cover	-0.34322	0.08616	-3.984	< 0.001
Method [NetTraps]	-0.39355	0.24902	-1.580	0.1140
Method [Traps]	-0.85075	0.19366	-4.393	< 0.001

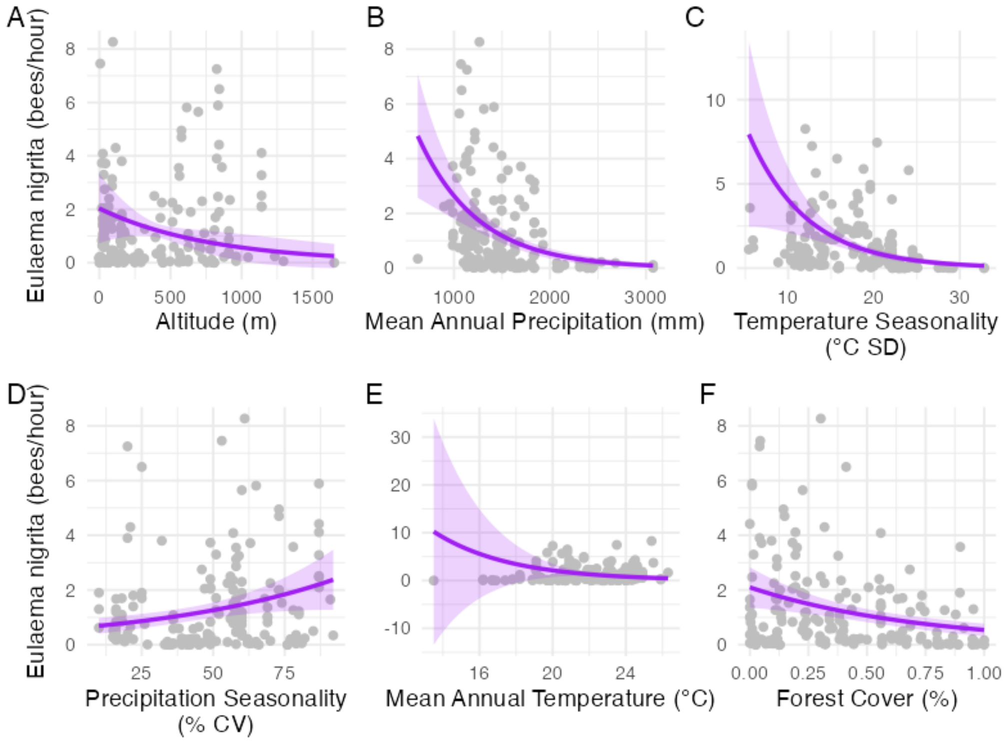

Figure 2: Predicted partial effects of significant continuous predictors on the mean abundance of *Eulaema nigrita*. Panels show Altitude (A), Mean Annual Precipitation (B), Temperature Seasonality (C), Precipitation Seasonality (D), Mean Annual Temperature (E), and Forest Cover (F). Solid purple lines indicate model-predicted mean bees/hour, and shaded areas represent 95% confidence intervals. Grey points show the observed bees/hour for each site.

4 Conclusion

Our study demonstrates that the abundance of the generalist orchid bee *Eulaema nigrita* is strongly shaped by climatic conditions, particularly temperature stability and rainfall seasonality, with geographic and local land-use factors playing smaller but meaningful roles. The species prefers stable temperatures, lower mean annual temperature and precipitation, and areas with distinct wet and dry periods, while showing a preference for reduced forest cover. Detection methodology also had a substantial impact on observed abundance, highlighting the importance of accounting for sampling method and survey effort when comparing bee populations across sites. Standardizing counts by effort (bees/hour) allowed for meaningful comparisons. Overall, *E. nigrita* is resilient to habitat disturbance but sensitive to climatic conditions, highlighting the need to account for both climate and methodology in monitoring and conservation.

Appendix: R Code

```r
# Load data and required libraries
dat = read.csv("../datasets/Eulaema.csv")
library(dplyr)
library(MASS)
library(broom)
library(ggplot2)
library(patchwork)

# Data Prep
# MAT and Tseason unit adjustment
dat$MAT_C <- dat$MAT / 10
dat$Tseason_C <- dat$Tseason / 100

# Create 'rate' column for visualization
dat$rate <- dat$Eulaema_nigrita / exp(dat$effort)

# Factorize the categorical variable
dat$method <- factor(dat$method)

# Standardization
num_vars <- c("effort", "altitude", "MAT_C", "MAP", "Tseason_C", "Pseason", "forest.", "lu_het")
dat_scaled <- dat
dat_scaled[num_vars] <- scale(dat[num_vars])

# ---------------------------------------------------------
# 1. MODEL FITTING (on standardized data)
m_nb_1_scaled_effort = glm.nb(Eulaema_nigrita ~ altitude + MAT_C + MAP + Tseason_C + Pseason + forest. + method + offset(effort),
                  data = dat_scaled)
summary(m_nb_1_scaled_effort)

# dispersion and pseudo R2 calculation
dispersion <- m_nb_1_scaled_effort$deviance / m_nb_1_scaled_effort$df.residual
# dispersion is approx. 1.207
# Pseudo R2 is approx. 0.391
1-(m_nb_1_scaled_effort$deviance/m_nb_1_scaled_effort$null.deviance)

# Calculation of percent change in expected rate (for Results section)
tidy_nb <- broom::tidy(m_nb_1_scaled_effort, conf.int = TRUE, exponentiate = TRUE) %>%
  dplyr::mutate(
    percent_change = (estimate - 1) * 100
  ) %>%
  dplyr::select(term, estimate, percent_change)
tidy_nb

# ---------------------------------------------------------
# 2. FIGURE 1 CODE: Abundance Boxplots
# Raw counts plot
p1 <- ggplot(dat, aes(x = "", y = Eulaema_nigrita)) +
  geom_boxplot(fill = "#ffa2d1ff", outlier.color = "#ffa2d1ff") +
  labs(title = "Eulaema nigrita Abundance\n(raw counts)", y = "") +
  theme_minimal(base_size = 16) +
  theme(axis.title.x = element_blank(), axis.text.x = element_blank(), axis.ticks.x = element_blank(), plot.title = element_text(hjust = 0.5))

# Bees/hour plot
p2 <- ggplot(dat, aes(x = "", y = rate)) +
  geom_boxplot(fill = "#ffa2d1ff", outlier.color = "#ffa2d1ff") +
  labs(title = "Eulaema nigrita Rate\n(bees/hour)", y = "") +
```

```r
  theme_minimal(base_size = 16) +
  theme(axis.title.x = element_blank(), axis.text.x = element_blank(), axis.ticks.x =
     element_blank(), plot.title = element_text(hjust = 0.5))

panel_boxplots <- p1 | p2
panel_boxplots

# ------------------------------------------------------------
# 3. FIGURE 2 CODE: Partial Effects Plots

# Refit model on unscaled data for plotting
m_nb_1 = glm.nb(Eulaema_nigrita ~ altitude + MAT_C + MAP + Tseason_C + Pseason +
    forest. + method + offset(effort),
               data = dat)
summary(m_nb_1)

# Variables and labels for plots:
key_vars <- c("altitude", "MAP", "Tseason_C", "Pseason", "MAT_C", "forest.")
nice_labels <- c("Altitude (m)",
                 "Mean Annual Precipitation (mm)",
                 "Temperature Seasonality\n(C SD)",
                 "Precipitation Seasonality\n(% CV)",
                 "Mean Annual Temperature (C)",
                 "Forest Cover (%)")

plots <- list()
mean_effort <- mean(dat$effort, na.rm = TRUE)
mean_effort_hours <- exp(mean_effort)

for(i in seq_along(key_vars)) {
  var <- key_vars[i]

  # Create new data frame for mean values for prediction
  new_data <- data.frame(
    effort = mean_effort,
    altitude = mean(dat$altitude, na.rm=TRUE),
    MAT_C = mean(dat$MAT_C, na.rm=TRUE),
    MAP = mean(dat$MAP, na.rm=TRUE),
    Tseason_C = mean(dat$Tseason_C, na.rm=TRUE),
    Pseason = mean(dat$Pseason, na.rm=TRUE),
    forest. = mean(dat$forest., na.rm=TRUE),
    method = factor("Net", levels=levels(dat$method))
  )

  new_data <- new_data[rep(1, 100), ]
  new_data[[var]] <- seq(min(dat[[var]], na.rm=TRUE),
                         max(dat[[var]], na.rm=TRUE), length.out=100)

  # Predict mean counts and CIs from unscaled model (m_nb_1)
  pred <- predict(m_nb_1, newdata = new_data, type="response", se.fit = TRUE)

  # Convert to rate (bees/hour)
  new_data$fit_rate <- pred$fit / mean_effort_hours
  new_data$upper_rate <- (pred$fit + 1.96 * pred$se.fit) / mean_effort_hours
  new_data$lower_rate <- (pred$fit - 1.96 * pred$se.fit) / mean_effort_hours

  # Plot
  p <- ggplot(new_data, aes(x = .data[[var]], y = fit_rate)) +
    geom_point(data = dat, aes(x = .data[[var]], y = rate), color = "grey") +
    geom_line(color = "purple", linewidth = 1.2) +
    geom_ribbon(aes(ymin = lower_rate, ymax = upper_rate), alpha = 0.2, fill = "
   purple") +
    labs(x = nice_labels[i], y = "Eulaema nigrita (bees/hour)") +
```

```
    theme_minimal(base_size = 14) +
    theme(plot.margin = unit(c(0.2,0.2,0.2,0.2), "cm"),
          axis.title.y = if(i == 1 || i == 4) element_text(margin=margin(r=10)) else
    element_blank())

  plots[[i]] <- p
}

panel_effects <- (plots[[1]] | plots[[2]] | plots[[3]]) /
                 (plots[[4]] | plots[[5]] | plots[[6]]) &
                 plot_annotation(tag_levels = 'A')
panel_effects
```