

Disentangling the Effects of Sexual and Morphological Traits on Mating Success in Male Damselflies (*Calopteryx splendens*).

Student Chandrashekar CR

Course Processing and Analysis of Biological Data (BIOS14)

Introduction

Sexual selection and mating success in damselflies are fundamental to understanding their evolutionary biology (Serrano-Meneses et al., 2008)(Outomuro et al., 2016). While both morphological and behavioral traits influence reproductive success in *Calopteryx splendens*, the relative importance of different trait categories and their interactions remain poorly understood. This knowledge gap is particularly significant given the species' complex mating system, where multiple traits may simultaneously influence mate choice. Recent studies have begun to unravel these relationships, revealing unexpected patterns in trait-based sexual selection. For instance, Golab and Brodin (Golab & Brodin, 2024) demonstrated that while consistent courtship behavior enhances mating success, males with smaller wing patches—a key sexual trait—achieved greater reproductive success. Similarly, research on sympatric populations of *C. splendens* and *C. virgo* revealed complex selection pressures acting on both morphological and secondary sexual characteristics (Tynkkyne et al., 2010). Building on these findings, we investigate the relative contributions of sexual traits (forewing patch dimensions) and morphological characteristics to copulation success in male *C. splendens*. Our study draws from a comprehensive dataset collected over five summers (2011–2015) in Sweden, measuring multiple morphological variables: total body length, abdomen length, thorax dimensions (length and width), forewing length, hindwing length and sexual traits including forewing patch dimensions (length and width). By systematically analyzing trait relationships with mating success in male *C. splendens*, we aim to: (1) Evaluate how sexual traits, specifically wing patch dimensions, correlate with copulation success through detailed analysis of patch length and width variations between mating and non-mating males. (2) Develop and validate an integrated model that combines both sexual and morphological trait categories, providing biologically meaningful interpretations of how these traits collectively influence mating probability. This approach addresses a critical gap in our understanding of sexual selection in damselflies, potentially revealing new patterns in how multiple trait categories collectively influence reproductive success.

Methods

1. Sexual Traits and Copulation Status

The relationship between copulation status and wing patch dimensions was analyzed in male *C. splendens* to determine if differences in forewing patch dimensions exist between copulating and non-copulating individuals. Differences in forewing patch length (fpl) and width (fpw) between copulating and non-copulating males were assessed using a one-way analysis of variance (ANOVA).

2. Model Development and Selection

The following analysis employed a stepwise approach to evaluate factors influencing copulation success, progressing from simple to more complex models. Initial models focused on sexual traits, examining the predictive power of forewing patch dimensions through logistic regression. Four model configurations were tested: two single-predictor models examining forewing patch length and width separately, a combined model incorporating both dimensions, and an interaction model (see Appendix Table: 3 for model specifications). Model performance was assessed using pseudo R-squared values and Tjur's coefficient of discrimination (D) to evaluate their explanatory power.

A parallel analysis examined morphological traits' influence on copulation success. Prior to modeling, Variance Inflation Factor (VIF) analysis was conducted to address potential multicollinearity among morphological traits (see Appendix Table: 4 for model specifications). The final model incorporated morphological traits selected through bidirectional elimination based on AIC values.

The differential performance of these separate analyses informed the development of a comprehensive combined model. Many candidate models were constructed, systematically integrating sexual and morphological traits to explore their potential interactive effects on copulation success. Model selection was guided by AIC values, enabling identification of the most parsimonious model that captured both direct and interactive effects. To facilitate biological interpretation, predicted probabilities were calculated at mean predictor values and ± 1 standard deviation, while

threshold values were determined to identify critical points where changes in predictors might significantly influence mating behavior (see Appendix for calculation details).

Results

1. Sexual Traits and Copulation Status

Analysis of variance (ANOVA) was used to assess the relationship between copulation status and forewing patch dimensions (fpl and fpw). Copulation status was treated as a categorical responses (0 = not observed mating, 1 = observed mating), while fpl and fpw were treated as continuous variables. (See Appendix Figure: 3, 4)

Comparison of Forewing Patch Length and Width by Copulation Status

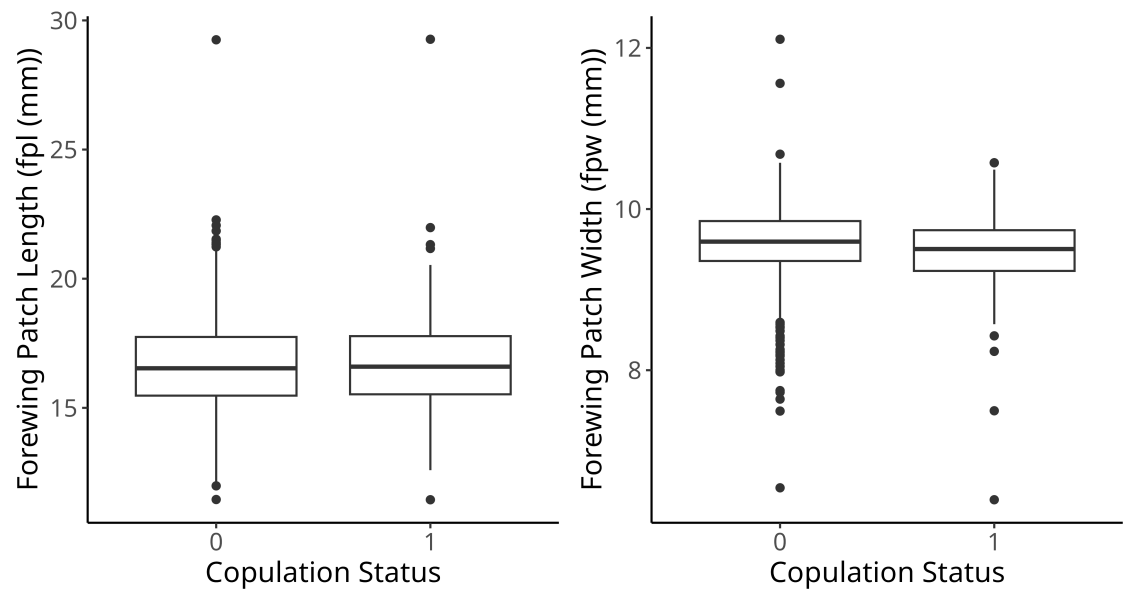


Figure 1: Analysis of sexual traits with respect to copulation status. **Copulation Status (cop)**: cop = 0 means that the individual was always found without a partner, 1 means that the individual was found while mating with a partner at least once.

Table 1: Comparison of Copulation Status with fpl and fpw

Analysis	Copulation Status vs. fpl	Copulation Status vs. fpw
Variance Explained (R^2)	0.00016	0.01104
Estimate for Copulation Status (1)	0.05961 ± 0.10599 mm	-0.12018 ± 0.025231 mm
Group Means	$mean_{cop=0} = 16.658 \pm 0.041$ mm $mean_{cop=1} = 16.718 \pm 0.041$ mm	$mean_{cop=0} = 9.585 \pm 0.010$ mm $mean_{cop=1} = 9.465 \pm 0.010$ mm
Percentage Difference	0.36% (cop=1 vs. cop=0)	-1.25% (cop=1 vs. cop=0)
p-value	0.5739 (Not significant)	2.04×10^{-6} (Significant)
Interpretation	No significant difference in fpl across copulation statuses.	Individuals found mating (cop=1) have slightly smaller fpw, with a 1.25% reduction compared to individuals without a partner (cop=0).

cop: Copulation status. **fpl**: Forewing patch length (mm). **fpw**: Forewing patch width (mm).

Forewing Patch Length (fpl): The ANOVA results revealed no statistically significant difference in fpl between copulating and non-copulating individuals ($p = 0.5739$). The model explained a negligible proportion of variance in fpl ($R^2 = 0.00016$), indicating that copulation status has virtually no influence on forewing patch length. The mean fpl for non-copulating individuals was 16.658 ± 0.041 mm, compared to 16.718 ± 0.041 mm for copulating individuals, a difference of only 0.36%. This small difference, coupled with the high p-value, suggests that forewing patch length

is unlikely to play a substantial role in determining mating success in this population. Refer to Table: 1 and Figure: 1)

Forewing Patch Width (fpw): In contrast, the ANOVA for fpw revealed a statistically significant difference between copulating and non-copulating individuals ($p = 2.04 \times 10^{-6}$). Although the model only explained 1.1% of the variance in fpw ($R^2 = 0.01104$), the strong statistical significance indicates a real, albeit small, effect. The mean fpw for non-copulating individuals was 9.585 ± 0.010 mm, while copulating individuals had a significantly smaller mean fpw of 9.465 ± 0.010 mm, representing a 1.25% reduction. This suggests that males with slightly narrower forewing patches are more likely to be observed mating (Golab & Brodin, 2024). This could potentially be due to female preference for narrower patches, or it could be related to other unmeasured traits that are correlated with fpw and influence mating success. While the effect size is small, the statistical significance suggests this difference is unlikely to be due to chance. (Refer to Table: 1 and Figure: 1)

2. Model Development and Selection

Initial analyses examined the individual and combined effects of forewing patch dimensions (sexual traits) on copulation success using logistic regression. A model incorporating both forewing patch length (fpl) and width (fpw) was selected ($cop \sim fpl + fpw$) based on AIC from a set of four candidate models (including single-predictor and interaction models; see Appendix Table 3). This model revealed contrasting effects: narrower patches (fpw) and greater length (fpl) were associated with increased copulation probability. However, this model explained minimal variance in copulation status (pseudo $R^2 = 0.029$, Tjur's $D = 0.014$), suggesting that sexual traits alone are insufficient to fully explain variation in mating success (see Appendix Table: 5 for model specifications and interpretation). A separate model examining morphological traits ($cop \sim (thorl \times thorw) + fwl + tbl$) demonstrated more complex relationships (see Appendix Table 4). Variance Inflation Factor (VIF) analysis revealed high collinearity between hindwing length and forewing length. Given the biological focus on forewing traits and the need to minimize redundancy, hindwing length was excluded to improve model interpretability. The final model incorporated thorax dimensions (width and length), forewing length, and total body length, selected through bidirectional elimination based on AIC values. Thorax dimensions (thorl and thorw) showed interactive effects suggesting optimal intermediate sizes, while forewing length (fwl) displayed positive associations with copulation success and total body length (tbl) showed negative associations (see Appendix Table: 6 for model specifications and interpretation). This morphological model explained substantially more variance (pseudo $R^2 = 0.175$, Tjur's $D = 0.119$) than the sexual traits model, highlighting the importance of considering other morphological traits.

A subsequent, more comprehensive logistic regression model was used to assess the combined influence of morphological and sexual traits on copulation success. This model ($cop \sim fpw + fpl + (thorl \times thorw) + fwl + tbl$) was selected based on having the lowest AIC score (1519.9) compared to the other models tested (Table 2) (See Appendix Figure: 5), indicating a better fit to the data. The model equation is as follows:

$$\begin{aligned} \text{logit}(P) = \ln \left(\frac{P}{1 - P} \right) = \text{log-odds} = & \beta_0 + \beta_1 \cdot fpw + \beta_2 \cdot fpl + \beta_3 \cdot thorl \\ & + \beta_4 \cdot thorw + \beta_5 \cdot fwl + \beta_6 \cdot tbl + \beta_7 \cdot (thorl \cdot thorw), \end{aligned} \tag{1}$$

where P is the probability of the outcome of interest (copulation success), $1 - P$ is the probability of failure (no copulation), and $\frac{P}{1 - P}$ is the odds of success.

Table 2: All Logistic Regression Models for Copulation Status with AIC Scores.

Model Equation	Traits	AIC Score
$cop \sim fpw + fpl$	Sexual	1700.8
$cop \sim (thorl \times thorw) + fwl + tbl$	Morphological	1531.2
$cop \sim fpw + fpl(thorl \times thorw) + fwl + tbl$	Sexual and Morphological	1519.9

cop: Copulation status. **tbl:** Total body length (mm). **thorl:** Thorax length (mm). **thorw:** Thorax width (mm). **fwl:** Forewing length (mm). **fpl:** Forewing patch length (mm). **fpw:** Forewing patch width (mm). **AIC:** Akaike Information Criterion, measures the relative fit of the statistical models.

At mean predictor values, the baseline probability of copulation success was 9.4%. Varying each predictor by one standard deviation (SD) while holding others constant revealed distinct effects (Figure 2). Forewing patch width (fpw) exhibited a negative relationship with copulation success; increasing fpw by one SD decreased the probability to 8.2%, while decreasing fpw by one SD increased it to 10.8%, suggesting that males with narrower patches are more successful

in mating. This aligns with previous research suggesting a potential female preference for smaller wing patches (Golab & Brodin, 2024). Forewing patch length (fpl) showed a positive association; a one SD increase raised the probability to 12.2%, consistent with findings that larger forewing patches enhance visual signaling during courtship (Allison, 2016). Forewing length (fwl) also positively influenced copulation success, with a one SD increase raising the probability to 13.5%, likely reflecting improved flight performance advantageous for mate pursuit (Outomuro et al., 2021). Thorax length (thorl) and thorax width (thorw) demonstrated complex effects due to their significant interaction. Increasing thorl or thorw by one SD reduced copulation probabilities to 8.2% and 7.6%, respectively, whereas decreasing them increased probabilities to 10.8% and 11.7%, suggesting an optimal range of thorax dimensions for mating success, potentially related to flight mechanics or other performance traits (Golab & Brodin, 2024) (See Appendix Figure: 7, 8). Finally, total body length (tbl) was negatively associated with copulation success; increasing tbl by one SD reduced the probability to 7.6%, while decreasing it increased the probability to 11.6%. This finding is consistent with the small male advantage hypothesis observed in some non-territorial damselfly species, where smaller males benefit from increased agility and reduced visibility to predators (de Almeida et al., 2022). (see Appendix Table: 7 for model specifications and interpretation)

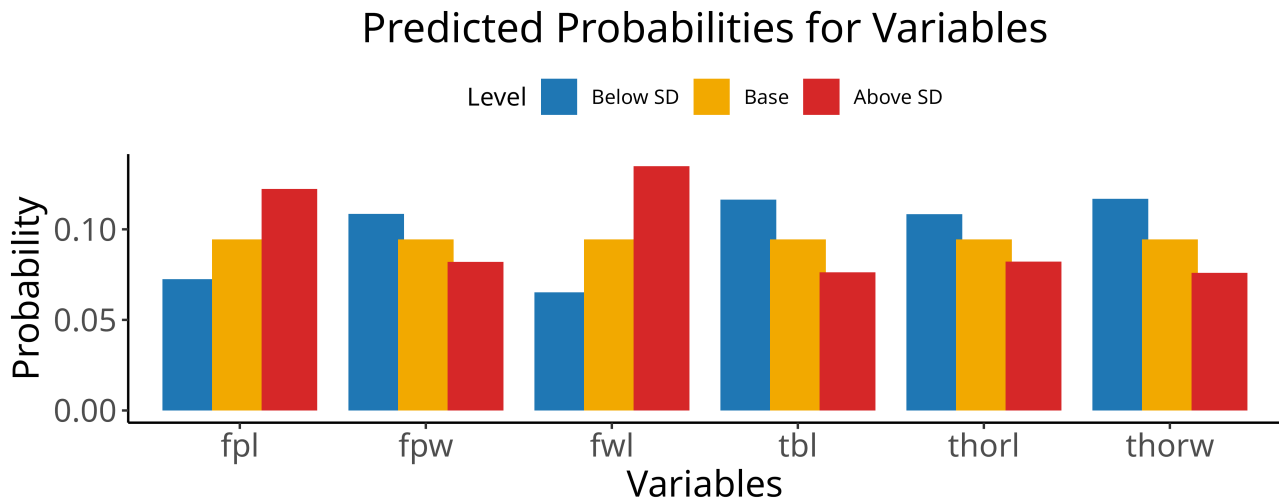


Figure 2: Predicted probabilities for different levels of predictor variables (Below SD, Base, and Above SD (SD: Standard Deviation)). Predictor variables: **tbl** (Total body length, mm), **thorl** (Thorax length, mm), **thorw** (Thorax width, mm), **fwl** (Forewing length, mm), **fpl** (Forewing patch length, mm; males only), and **fpw** (Forewing patch width, mm; males only). "Below SD," "Base," and "Above SD" denote one SD below the mean, the mean, and one SD above the mean, respectively.

This combined model explained 18.5% of the variance in copulation status (pseudo $R^2 = 0.185$), a considerable improvement compared to the models considering sexual or morphological traits alone. Tjur's D of 0.128 indicates that the average predicted probability of copulation for individuals observed mating was 12.8% higher than for those not observed mating, demonstrating moderate discriminatory power. Threshold values, representing the predictor values at which the predicted probability of copulation is 50%, were calculated for each predictor while holding all others at their mean values (See Appendix Figure: 6). Only the thresholds for fwl (38.15 mm) and fpl (30.03 mm) fell within or at the edge of their respective observed ranges. This suggests that variation in these traits, particularly in combination with other traits, could plausibly influence copulation probability within the observed data. (See Appendix Figure: 6 8)

Conclusion

This study provides evidence that mating success in male *C. splendens* is determined by a complex interplay of both sexual and morphological traits, rather than either category alone. Our analyses revealed that while sexual traits (forewing patch dimensions) influence copulation success, their impact is relatively modest compared to morphological characteristics. The superior performance of the combined model suggests that sexual selection in *C. splendens* operates on multiple trait categories simultaneously, supporting earlier findings about the complexity of mate selection in this species (Golab & Brodin, 2024). Particularly notable was the discovery that intermediate thorax dimensions and smaller total body size confer mating advantages, while longer forewings enhance copulation success. These patterns suggest trade-offs between flight performance and body size (Tynkkynen et al., 2010), while the contrasting effects of forewing patch dimensions indicate that patch shape, rather than just size, influences mate selection. Our

results extend beyond earlier studies by demonstrating quantitatively that the most comprehensive understanding of mating success emerges from considering both morphological and sexual traits together. This integrated approach reveals patterns that were not apparent when examining trait categories in isolation, highlighting the need to consider multiple aspects of male phenotype in sexual selection studies. Future research could benefit from investigating how these trait relationships vary across different populations and environmental conditions, potentially revealing the ecological contexts that shape the relative importance of different trait categories in sexual selection.

References

- Allison, S. E. (2016). Variation in female mating behavior and success in the damselfly, *calopteryx maculata*. *Masters Theses, 2010-2019*, 471. <https://commons.lib.jmu.edu/master201019/471>
- de Almeida, T. R., Salomoni, S., Vilela, D. S., & Guillermo-Ferreira, R. (2022). Male agility in relation to mating success in two non-territorial damselflies. *Austral Ecology*, 47(4), 1569–1577.
- Golab, M. J., & Brodin, T. (2024). Looks or personality: What drives damselfly male mating success in the wild? *The European Zoological Journal*, 91(1), 81–93. <https://doi.org/10.1080/24750263.2023.2294090>
- Outomuro, D., Golab, M. J., Johansson, F., & Sniegula, S. (2021). Body and wing size, but not wing shape, vary along a large-scale latitudinal gradient in a damselfly. *Scientific Reports*, 11(1), 18642. <https://doi.org/10.1038/s41598-021-97829-9>
- Outomuro, D., Söderquist, L., Nilsson-Örtman, V., Cortázar-Chinarro, M., Lundgren, C., & Johansson, F. (2016). Antagonistic natural and sexual selection on wing shape in a scrambling damselfly. *Evolution*, 70(7), 1582–1595. <https://doi.org/10.1111/evo.12951>
- Serrano-Meneses, M. A., Córdoba-Aguilar, A., Azpilicueta-Amorín, M., González-Soriano, E., & Székely, T. (2008). Sexual selection, sexual size dimorphism and rensch's rule in odonata. *Journal of Evolutionary Biology*, 21(5), 1259–1273. <https://doi.org/10.1111/j.1420-9101.2008.01567.x>
- Tynkkynen, K., Kotiaho, J., Luojumki, M., & Suhonen, J. (2010). Selection on size and secondary sexual characters of the damselfly *calopteryxsplendens* when sympatric with the congener *calopteryxvirgo*. *Canadian Journal of Zoology*, 89, 1–9. <https://doi.org/10.1139/Z10-090>

Appendix

Explanation of predicted probability of copulation

The predicted probability of copulation represents the likelihood, as estimated by the model, that an individual has a copulation status of 1 (i.e., was observed mating at least once). While copulation status is binary (0 or 1), logistic regression does not predict these values directly. Instead, it estimates the probability (P) that the outcome is 1. Logistic regression models the log-odds (logit) of the outcome as a linear combination of predictors:

$$\text{logit}(P) = \ln \left(\frac{P}{1-P} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

where β_0 is the intercept, β_1, β_2, \dots are the coefficients, and x_1, x_2, \dots are the predictor values. The log-odds are transformed into probabilities using:

$$P = \frac{1}{1 + e^{-\text{logit}}}$$

ensuring P always lies between 0 and 1. A threshold probability of $P = 0.5$ serves as the decision boundary, where $P > 0.5$ indicates a predicted outcome of 1 (copulation) and $P < 0.5$ indicates 0 (no copulation). Threshold values for predictors correspond to the points where $P = 0.5$, assuming other predictors are held at their mean values. These thresholds reveal critical values at which changes in a predictor significantly affect mating likelihood. For instance, if the threshold for thorax length (thorl) is 32 mm, individuals with a thorax length of 32 mm have a 50% chance of mating. This analysis highlights how specific traits influence copulation probability, providing insights into biological or behavioral mechanisms underlying mating success.

Additional Tables and Figures

Table 3: Logistic Regression Models (Sexual Traits) for Copulation Status with AIC Scores.

Model Equation	AIC Score
$cop \sim fpl$	1726.7
$cop \sim fpw$	1705.6
$cop \sim fpw + fpl$	1700.8
$cop \sim fpw + fpl + (fpw \times fpl)$	1702.8

cop: Copulation status. **fpl:** Forewing patch length (mm). **fpw:** Forewing patch width (mm). **AIC:** Akaike Information Criterion, measures the relative fit of the statistical models.

Table 4: Logistic Regression Models (Morphological Traits) for Copulation Status with AIC Scores.

Model Equation	AIC Score
$cop \sim tbl + abl + thorl + thorw + fwl$	1556.6
$cop \sim tbl + thorl + thorw + fwl$	1554.6
$cop \sim (thorl \times thorw) + fwl + tbl$	1531.2

cop: Copulation status. **tbl:** Total body length (mm). **abl:** Abdomen length (mm). **thorl:** Thorax length (mm). **thorw:** Thorax width (mm). **fwl:** Forewing length (mm). **AIC:** Akaike Information Criterion, measures the relative fit of the statistical models.

Table 5: Standard Deviation Effects and Threshold Values for Copulation Probability (Sexual Traits Only: fpw and fpl)

Predictor	Effect of SD Change on Copulation Probability	Threshold for 50% Probability of Copulation
Forewing Patch Width (fpw)	Baseline: 14.6% (at mean fpw and fpl). One SD below mean: 19.4%. One SD above mean: 10.8%. Negative relationship: Smaller fpw is associated with higher copulation probability.	5.4167 mm (Biologically meaningful, well below the observed mean)
Forewing Patch Length (fpl)	Baseline: 14.6% (at mean fpw and fpl). One SD below mean: 12.6%. One SD above mean: 17.0%. Positive relationship: Larger fpl is associated with higher copulation probability.	-44.1492 mm (Biologically meaningless, outside realistic range)

fpw: Forewing patch width (mm). **fpl:** Forewing patch length (mm).

Table 6: Standard Deviation Effects and Threshold Values for Copulation Probability (Morphological Traits Only: thorl, thorw, fwl, and tbl)

Predictor	Effect of SD Change on Copulation Probability	Threshold for 50% Probability of Copulation
Thorax Length (thorl)	Baseline: 9.6% (at mean thorl, thorw, fwl, and tbl). One SD below mean: 11.7%. One SD above mean: 7.9%. Complex relationship: Interaction with thorw likely influences copulation probability.	−10.46 mm (Biologically meaningless, negative value)
Thorax Width (thorw)	Baseline: 9.6% (at mean thorl, thorw, fwl, and tbl). One SD below mean: 12.0%. One SD above mean: 7.7%. Complex relationship: Interaction with thorl likely influences copulation probability.	−3.62 mm (Biologically meaningless, negative value)
Forewing Length (fwl)	Baseline: 9.6% (at mean thorl, thorw, fwl, and tbl). One SD below mean: 6.9%. One SD above mean: 13.3%. Positive relationship: Larger fwl is associated with a higher probability of copulation.	−44.05 mm (Biologically meaningless, negative value)
Total Body Length (tbl)	Baseline: 9.6% (at mean thorl, thorw, fwl, and tbl). One SD below mean: 11.8%. One SD above mean: 7.8%. Negative relationship: Larger tbl is associated with a lower probability of copulation.	147.67 mm (Likely biologically unrealistic, far larger than observed range)

thorl: Thorax length (mm). **thorw:** Thorax width (mm). **fwl:** Forewing length (mm). **tbl:** Total body length (mm).

Table 7: Standard Deviation Effects and Threshold Values for Copulation Probability (Combined Model Sexual and Morphological Traits)

Predictor	Effect of SD Change on Copulation Probability	Threshold for 50% Probability of Copulation
Forewing Patch Width (fpw)	One SD below mean: 10.8%, One SD above mean: 8.2%, Negative relationship: Smaller fpw is associated with higher copulation probability.	3.56 mm (Biologically meaningless, below observed range)
Forewing Patch Length (fpl)	One SD below mean: 7.2%, One SD above mean: 12.2%, Positive relationship: Larger fpl is associated with higher copulation probability.	30.03 mm (At the high end of the observed range)
Thorax Length (thorl)	One SD below mean: 10.8%, One SD above mean: 8.2%, Complex relationship due to interaction with thorax width (thorw).	-18.14 mm (Biologically meaningless, outside observed range)
Thorax Width (thorw)	One SD below mean: 11.7%, One SD above mean: 7.6%, Complex relationship due to interaction with thorax length (thorl).	-4.01 mm (Biologically meaningless, negative value)
Forewing Length (fwl)	One SD below mean: 6.5%, One SD above mean: 13.5%, Positive relationship: Larger fwl is associated with higher copulation probability.	38.15 mm (At the upper limit of the observed range)
Total Body Length (tbl)	One SD below mean: 11.6%, One SD above mean: 7.6%, Negative relationship: Larger tbl is associated with lower copulation probability.	30.33 mm (Below the observed range, biologically meaningless)

fpw: Forewing patch width (mm). **fpl:** Forewing patch length (mm). **thorl:** Thorax length (mm). **thorw:** Thorax width (mm). **fwl:** Forewing length (mm). **tbl:** Total body length (mm).

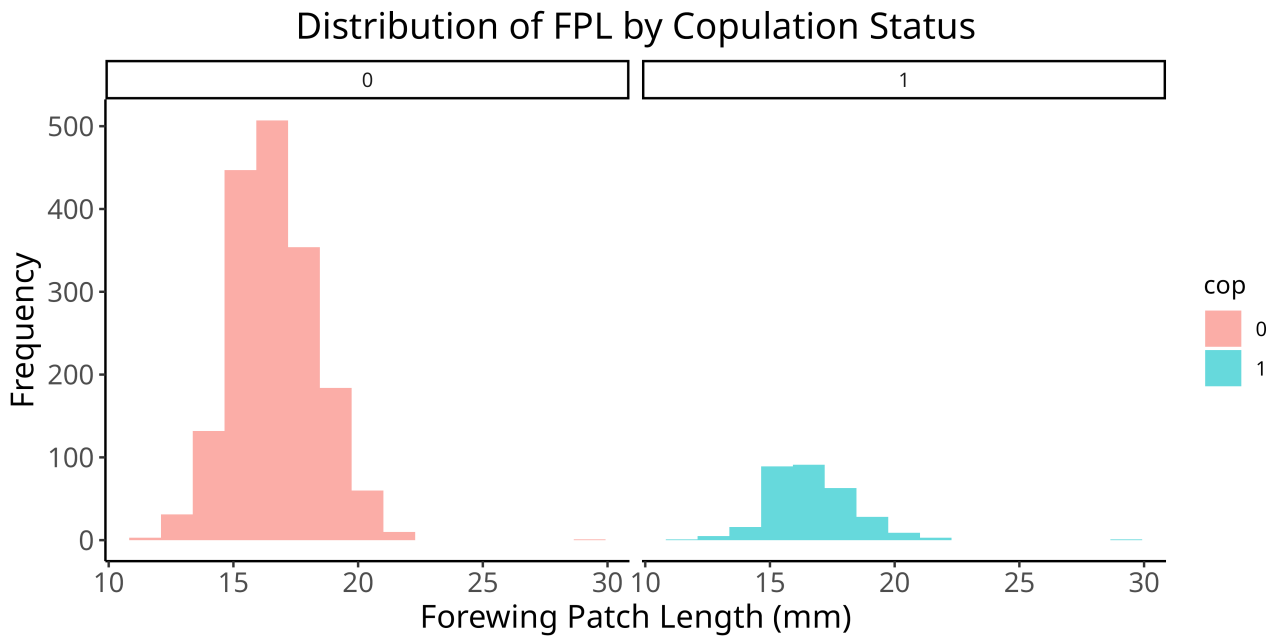


Figure 3: Distribution of Forewing Patch Length (fpl) by Copulation Status. This figure displays histograms of forewing patch length (fpl, in mm) for male individuals, separated by copulation status (cop): 0 (never observed mating) and 1 (observed mating at least once). The distributions of fpl for both copulation statuses appear approximately normally distributed, suggesting the applicability of a linear model, such as ANOVA, for statistical analysis. This analysis aims to investigate whether there is a significant difference in mean fpl between males with different copulation statuses.

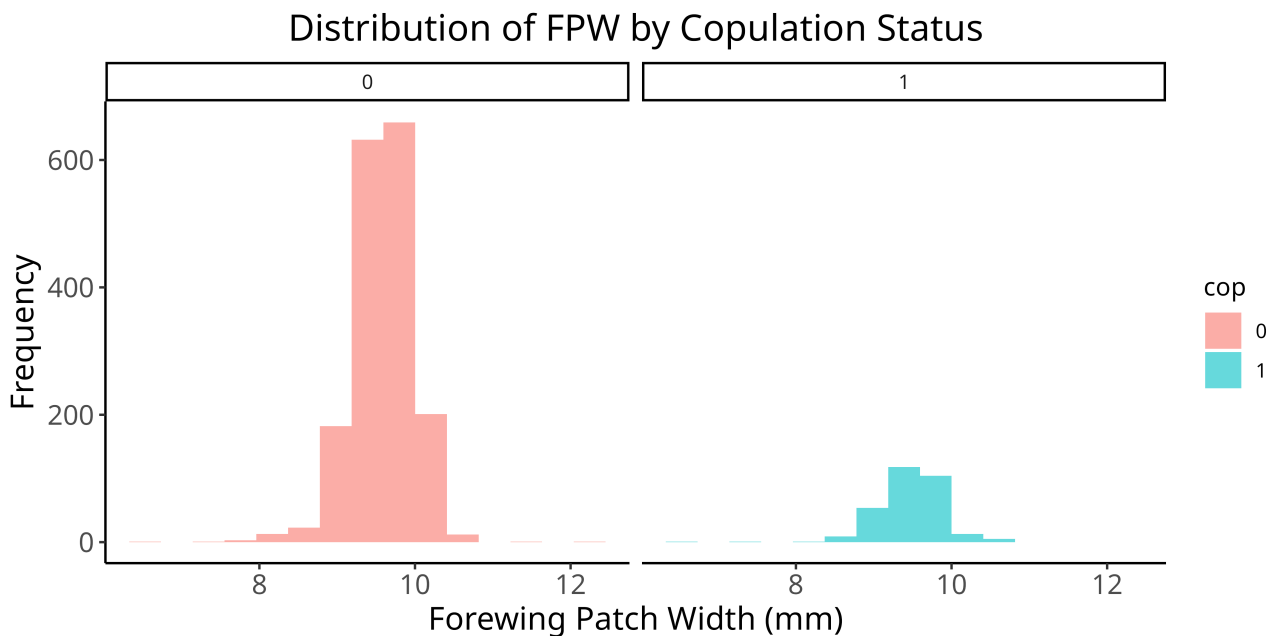


Figure 4: Distribution of Forewing Patch Width (fpw) by Copulation Status. This figure displays histograms of forewing patch width (fpw, in mm) for male individuals, separated by copulation status (cop): 0 (never observed mating) and 1 (observed mating at least once). The distributions of fpw for both copulation statuses appear approximately normally distributed, suggesting the applicability of a linear model, such as ANOVA, for statistical analysis. This analysis aims to investigate whether there is a significant difference in mean fpw between males with different copulation statuses.

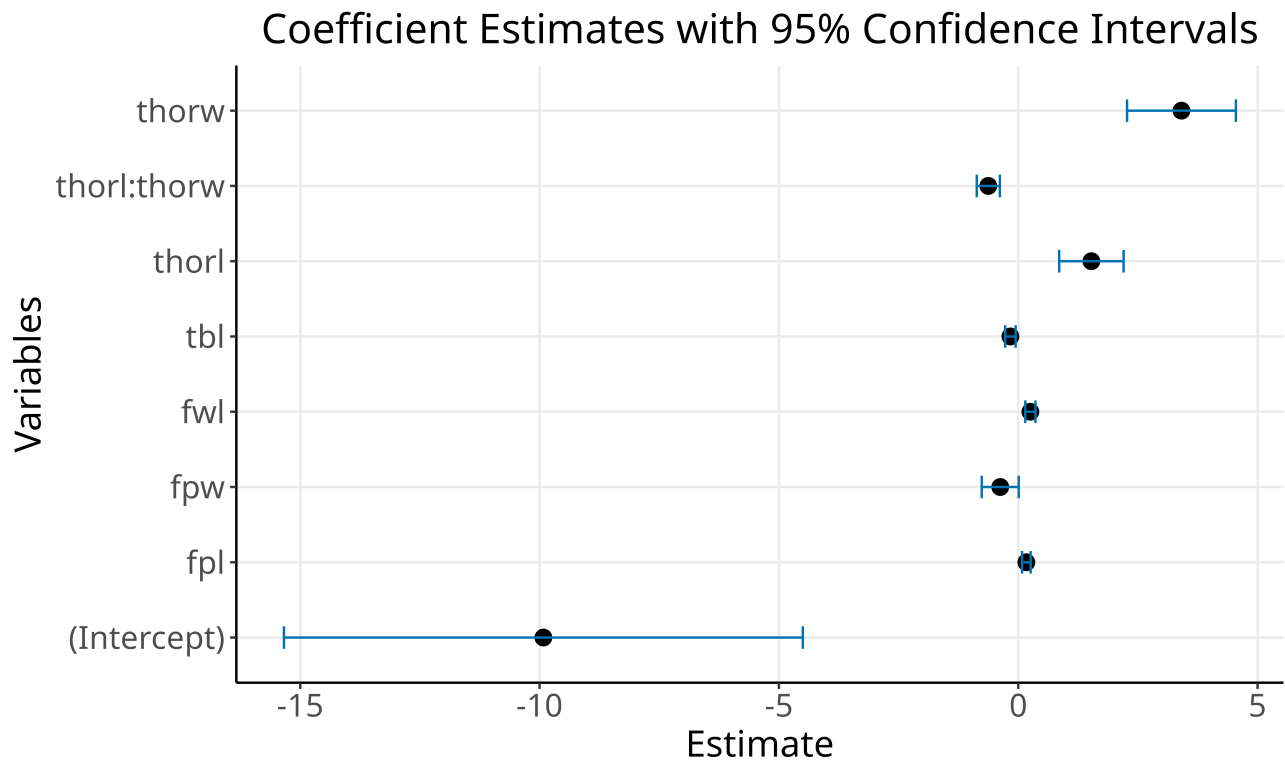


Figure 5: Coefficient Estimates with 95% Confidence Intervals from a Generalized Linear Model. This figure displays the coefficient estimates and their associated 95% confidence intervals from a binomial generalized linear model (GLM) predicting copulation status (*cop*). The model was fitted using the formula: $cop \sim fpw + fpl + (thorl * thorw) + fwl + tbl$. Here, **cop** represents copulation status (0 = never observed mating, 1 = observed mating at least once); **fpw** represents forewing patch width (mm); **fpl** represents forewing patch length (mm); **thorl** represents thorax length (mm); **thorw** represents thorax width (mm); **fwl** represents forewing length (mm); and **tbl** represents total body length (mm). The interaction term ($thorl * thorw$) includes both main effects of **thorl** and **thorw** as well as their interaction. The plot shows the estimated effect of each predictor on the log-odds of copulation. Confidence intervals that do not cross zero indicate statistically significant effects at the 5% level. The model was fit to male *C. splendens* data.

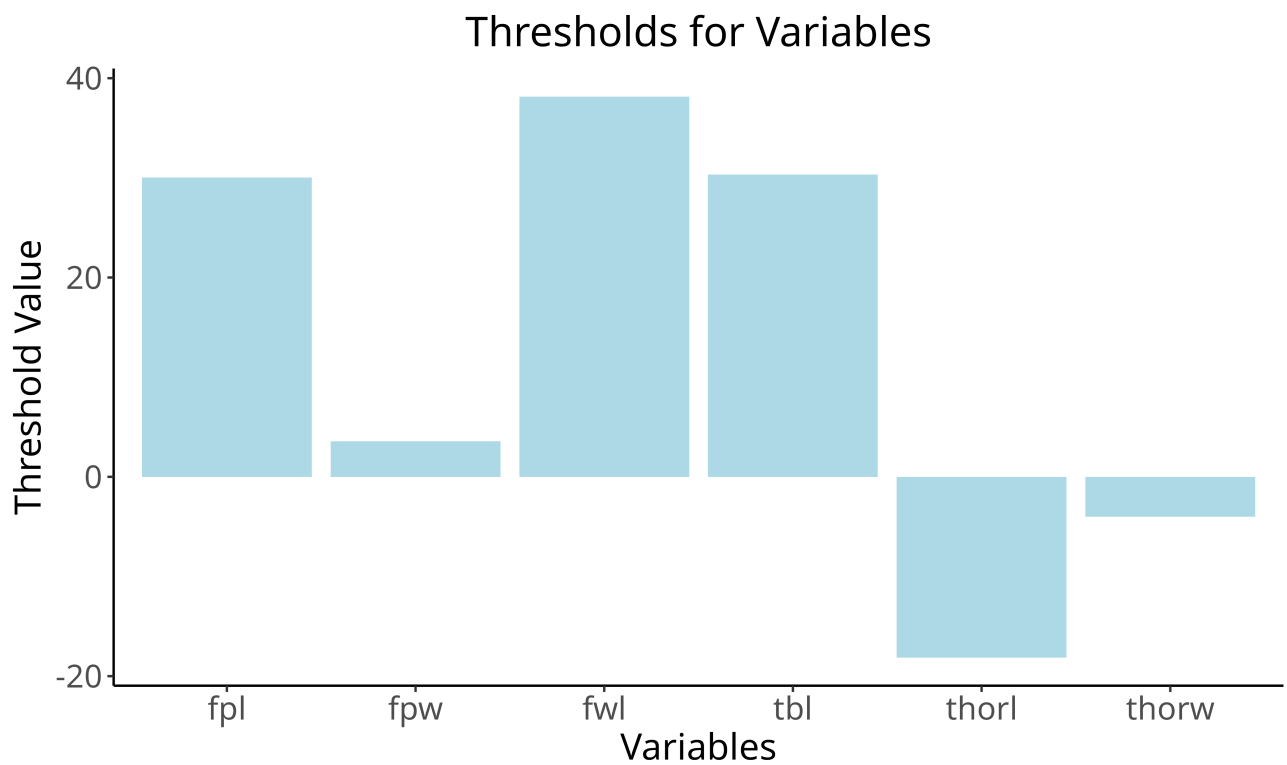


Figure 6: Thresholds for Predictor Variables. This figure displays the calculated threshold values for various predictor variables: tbl (Total body length, mm); thori (Thorax length, mm); thorw (Thorax width, mm); fwl (Forewing length, mm); fpl (Forewing patch length, mm; males only); and fpw (Forewing patch width, mm; males only).

Effect of thori and thorw Interaction on Predicted Probabilities

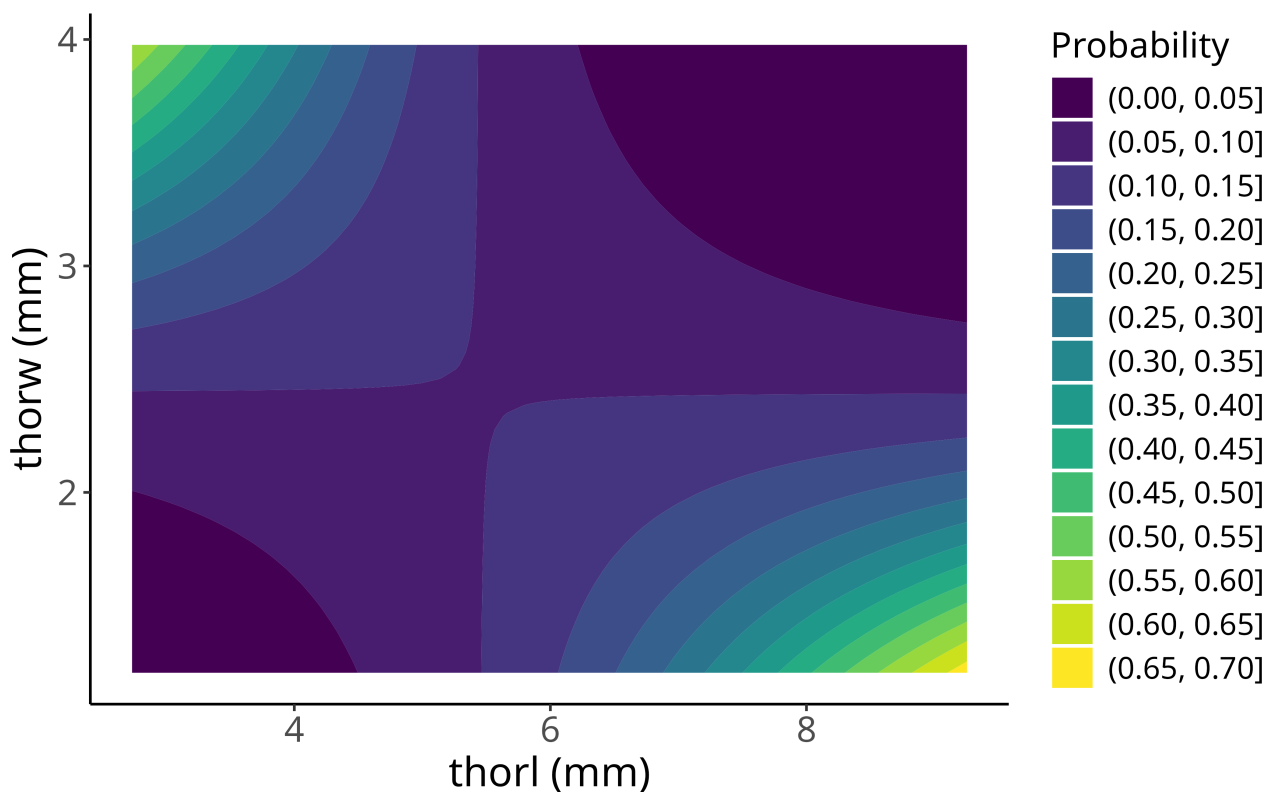


Figure 7: Interaction between thorax width (thorw) and thorax length (thori) on predicted probabilities. This contour plot illustrates the combined effect of thorax width (thorw, in mm) and thorax length (thori, in mm) on the predicted probability. The color gradient represents different probability ranges, as indicated in the legend. While high copulation probabilities are predicted at extreme combinations of thorw and thori, these combinations are biologically implausible (as shown in Figure 8).

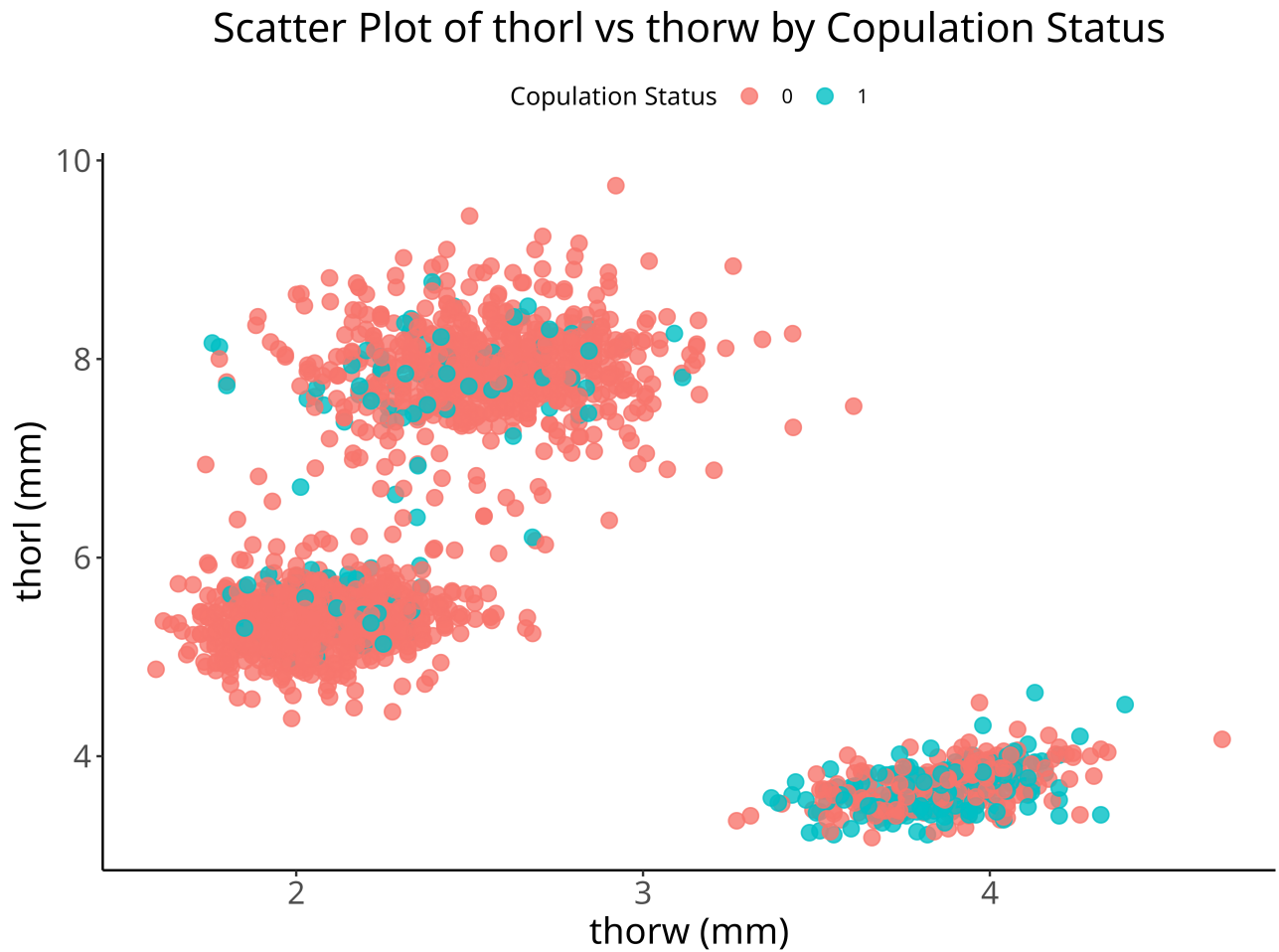


Figure 8: Relationship between thorax length (thorl) and thorax width (thorw) by copulation status. This scatter plot displays the distribution of individuals based on their thorax length (thorl, in mm) and thorax width (thorw, in mm), differentiated by copulation status (0: never observed mating, 1: observed mating at least once). The plot demonstrates the observed combinations of thorl and thorw, highlighting that certain extreme combinations (e.g., high thorw with low thorl, or vice versa) are not found in the dataset, indicating biological constraints. This figure provides empirical support for the biologically plausible region of the interaction plot shown in Figure 7.

Scripts

Damselflies BIOS14 GitHub Repository - Click this to go to the Github Repository.

```
1 # Clear any pre-stored variables in the environment
2 rm(list = ls())
3
4 # Import Libraries
5 library(MASS)
6 library(tidyverse)
7 library(ggplot2)
8 library(patchwork)
9 library(dplyr)
10 library(MuMIn)
11
12 # Load the data
13 male_CS = read.csv("../data/male_CS.csv")
14 male_CV = read.csv("../data/male_CV.csv")
15 female_CS = read.csv("../data/female_CS.csv")
16 female_CV = read.csv("../data/female_CV.csv")
17
18 data = bind_rows(male_CS, male_CV, female_CS, female_CV)
19
20 head(data)
```

Listing 1: R code to load and combine datasets of male and female *Calopteryx splendens*

```
1
2 # Clear the environment
3 rm(list = ls())
4
5
6 # Read and prepare the data
7 male_CS = read.csv("../data/male_CS.csv")
8 male_CS = male_CS %>%
9   select(fpl, fpw, cop) %>%
10   mutate(cop = as.factor(cop))
11
12 # Verify normality within copulation status groups
13
14 # 1. ----- Forewing Patch Width (fpw) -----
15
16 # Histogram
17 fpw_histogram = male_CS %>%
18   ggplot(aes(x = fpw, fill = cop)) +
19   geom_histogram(alpha = 0.6, bins = 15, position = "identity") +
20   facet_wrap(~cop) +
21   theme_classic() +
22   labs(
23     title = "Distribution of fpw by Copulation Status",
24     x = "Forewing Patch Width (mm)",
25     y = "Frequency"
26   ) +
27   theme(
28     axis.text = element_text(size = 12),
29     axis.title = element_text(size = 14),
30     plot.title = element_text(size = 16, hjust = 0.5)
31   )
32
33 # Save histogram
34 ggsave(
35   "../results/fpw_histogram.png",
36   fpw_histogram,
```

```

37   width = 200,
38   height = 100,
39   units = "mm",
40   dpi = 600
41 )
42
43 # Q-Q Plot
44 fpw_qq = male_CS %>%
45   ggplot(aes(sample = fpw, color = cop)) +
46     stat_qq() +
47     stat_qq_line() +
48     facet_wrap(~cop) +
49     theme_classic() +
50     labs(
51       title = "Q-Q Plot for fpw by Copulation Status",
52       x = "Theoretical Quantiles",
53       y = "Sample Quantiles"
54     ) +
55     theme(
56       axis.text = element_text(size = 12),
57       axis.title = element_text(size = 14),
58       plot.title = element_text(size = 16, hjust = 0.5),
59       legend.position = "none"
60     )
61
62 # Save Q-Q plot
63 ggsave(
64   "../results/fpw_qq.png",
65   fpw_qq,
66   width = 200,
67   height = 100,
68   units = "mm",
69   dpi = 600
70 )
71
72 # 2. ----- Forewing Patch Length (fpl) -----
73
74 # Histogram
75 fpl_histogram = male_CS %>%
76   ggplot(aes(x = fpl, fill = cop)) +
77   geom_histogram(alpha = 0.6, bins = 15, position = "identity") +
78   facet_wrap(~cop) +
79   theme_classic() +
80   labs(
81     title = "Distribution of fpl by Copulation Status",
82     x = "Forewing Patch Length (mm)",
83     y = "Frequency"
84   ) +
85   theme(
86     axis.text = element_text(size = 12),
87     axis.title = element_text(size = 14),
88     plot.title = element_text(size = 16, hjust = 0.5)
89   )
90
91 # Save histogram
92 ggsave(
93   "../results/fpl_histogram.png",
94   fpl_histogram,
95   width = 200,
96   height = 100,
97   units = "mm",
98   dpi = 600
99 )

```

```

100
101 # Q-Q Plot
102 fpl_qq = male_CS %>%
103   ggplot(aes(sample = fpl, color = cop)) +
104     stat_qq() +
105     stat_qq_line() +
106     facet_wrap(~cop) +
107     theme_classic() +
108     labs(
109       title = "Q-Q Plot for fpl by Copulation Status",
110       x = "Theoretical Quantiles",
111       y = "Sample Quantiles"
112     ) +
113     theme(
114       axis.text = element_text(size = 12),
115       axis.title = element_text(size = 14),
116       plot.title = element_text(size = 16, hjust = 0.5),
117       legend.position = "none"
118     )
119
120 # Save Q-Q plot
121 ggsave(
122   "../results/fpl_qq.png",
123   fpl_qq,
124   width = 200,
125   height = 100,
126   units = "mm",
127   dpi = 600
128 )
129
130 # Perform ANOVA for fpl
131 anova_fpl = lm(fpl ~ as.factor(cop), data = male_CS)
132 anova_result_fpl = anova(anova_fpl)
133 summary_fpl = summary(anova_fpl)
134
135 # Perform ANOVA for fpw
136 anova_fpw = lm(fpw ~ as.factor(cop), data = male_CS)
137 anova_result_fpw = anova(anova_fpw)
138 summary_fpw = summary(anova_fpw)
139
140 # Plot the results
141 fpw_histogram
142 fpw_qq
143 fpl_histogram
144 fpl_qq
145
146 # Print ANOVA results
147 print(anova_result_fpl)
148 print(summary_fpl)
149 print(anova_result_fpw)
150 print(summary_fpw)

```

Listing 2: R code for evaluating the normality and group differences in forewing patch width (fpw) and forewing patch length (fpl) among male *Calopteryx splendens* based on copulation status. The script includes data preparation, visualization (histograms and Q-Q plots), and statistical analysis using ANOVA to test for significant differences in fpw and fpl between copulation status groups. The plots are saved as high-resolution PNG files for further interpretation.

```

1 # Clear environment
2 rm(lsist = ls())
3
4 # Read and prepare data
5 male_CS = read.csv("../data/male_CS.csv")

```

```

6 male_CS = male_CS %>%
7   select(fpl, fpw, cop) %>%
8   mutate(cop = as.factor(cop))
9
10 # Boxplot for fpl
11 box_fpl = ggplot(male_CS, aes(x = as.factor(cop), y = fpl)) +
12   geom_boxplot() +
13   labs(x = "Copulation Status", y = "Forewing Patch Length (fpl (mm))") +
14   theme_classic() +
15   theme(
16     axis.text.x = element_text(size = 12), # Adjusted size of x-axis tick labels
17     axis.text.y = element_text(size = 12), # Adjusted size of y-axis tick labels
18     axis.title = element_text(size = 14), # Adjusted axis titles
19     plot.title = element_text(size = 16, hjust = 0.5) # Adjusted plot title
20   )
21
22 # Boxplot for fpw
23 box_fpw = ggplot(male_CS, aes(x = as.factor(cop), y = fpw)) +
24   geom_boxplot() +
25   labs(x = "Copulation Status", y = "Forewing Patch Width (fpw (mm))") +
26   theme_classic() +
27   theme(
28     axis.text.x = element_text(size = 12), # Adjusted size of x-axis tick labels
29     axis.text.y = element_text(size = 12), # Adjusted size of y-axis tick labels
30     axis.title = element_text(size = 14), # Adjusted axis titles
31     plot.title = element_text(size = 16, hjust = 0.5) # Adjusted plot title
32   )
33
34 # Combine plots side by side using patchwork
35 combined_plots = box_fpl + box_fpw + plot_layout(ncol = 2) +
36   plot_annotation(
37     title = "Comparison of Forewing Patch Length and Width by Copulation Status",
38     theme = theme(plot.title = element_text(size = 17, hjust = 0.5)) # Adjusted annotation title
39   )
40
41 # Display combined plots
42 combined_plots
43
44 # Save the plot as a high-resolution PNG for LaTeX document
45 ggsave(
46   "../results/anova_combined_plot.png",
47   combined_plots,
48   width = 200, # Width in mm
49   #height = 80, # Adjusted height to maintain aspect ratio
50   units = "mm",
51   dpi = 600
52 )

```

Listing 3: Boxplots comparing forewing patch length (fpl) and forewing patch width (fpw) across copulation status groups in male *Calopteryx splendens*. The combined plots illustrate the distribution of fpl (left) and fpw (right) for individuals observed mating at least once versus those never observed mating. Each boxplot highlights the median, interquartile range, and variability within groups, providing insights into potential morphological differences associated with mating success.

```

1
2 # Clear the environment
3 rm(list = ls())
4
5 # Read the data and filter accordingly
6 male_CS = read.csv("../data/male_CS.csv")
7
8 # Selects only the columns fpl, fpw, and cop from the dataset.

```



```

9 # Converts the 'cop' variable (copulation status) into a categorical variable (factor).
10 male_CS = male_CS %>%
11   select(fpl, fpw, cop) %>%
12   mutate(cop = as.factor(cop))
13
14
15 # Logistic regression model: Copulation status as a function of forewing patch length (fpl)
16 logit_model_fpl = glm(cop ~ fpl, data = male_CS, family = binomial(link = "logit"))
17 #summary(logit_model_fpl)
18
19
20 # Logistic regression model: Copulation status as a function of forewing patch width (fpw)
21 logit_model_fpw = glm(cop ~ fpw, data = male_CS, family = binomial(link = "logit"))
22 #summary(logit_model_fpw)
23
24 # Logistic regression model: Copulation status as a function of both fpl and fpw
25 logit_model_fpl_fpw = glm(cop ~ fpw + fpl, data = male_CS, family = binomial(link = "logit"))
26 summary(logit_model_fpl_fpw)
27
28 # Logistic regression model: Copulation status with interaction between fpw and fpl
29 logit_model_interaction = glm(cop ~ fpw + fpl + (fpw * fpl), data = male_CS, family = binomial(link =
  "logit"))
30 #summary(logit_model_interaction) # Only the interaction term (fpw*fpl) also gave the AIC score as
  this model.
31
32
33 # Pseudo R-squared (using MuMIn)
34 r.squaredGLMM(logit_model_fpl_fpw)
35
36 # Tjur's D
37 y_hat = predict(logit_model_fpl_fpw, type = "response") # Get predicted probabilities
38
39 tjur_d = mean(y_hat[male_CS$cop == 1]) - mean(y_hat[male_CS$cop == 0])
40 tjur_d
41
42 # Calculate mean and standard deviation of predictors
43 fpw_mean = mean(male_CS$fpw, na.rm = TRUE)
44 fpw_sd = sd(male_CS$fpw, na.rm = TRUE)
45 fpl_mean = mean(male_CS$fpl, na.rm = TRUE)
46 fpl_sd = sd(male_CS$fpl, na.rm = TRUE)
47
48 # Predicted probabilities for fpw and fpl (mean wrt SD)
49 fpw_low = fpw_mean - fpw_sd
50 fpw_high = fpw_mean + fpw_sd
51 fpl_low = fpl_mean - fpl_sd
52 fpl_high = fpl_mean + fpl_sd
53
54 # Logistic function for probabilities
55 logit_to_prob = function(log_odds) {
56   return(1 / (1 + exp(-log_odds)))
57 }
58
59 # Base line probabilities for fpw and fpl
60 prob_fpw_fpl_base = logit_to_prob(4.54030 + (-0.83821 * fpw_mean) + (0.10284 * fpl_mean))
61
62 # Predicted probabilities for fpw changes
63 prob_fpw_low = logit_to_prob(4.54030 + (-0.83821 * fpw_low) + (0.10284 * fpl_mean))
64 prob_fpw_high = logit_to_prob(4.54030 + (-0.83821 * fpw_high) + (0.10284 * fpl_mean))
65
66 # Predicted probabilities for fpl changes
67 prob_fpl_low = logit_to_prob(4.54030 + (-0.83821 * fpw_mean) + (0.10284 * fpl_low))
68 prob_fpl_high = logit_to_prob(4.54030 + (-0.83821 * fpw_mean) + (0.10284 * fpl_high))
69

```

```

70 # Threshold for P(cop=1)=0.5 for fpw and fpl
71 fpw_threshold = -(4.54030 + (0.10284*fpl_mean))/-0.83821
72 fpl_threshold = (-4.54030 + (-0.83821*fpw_mean))/ 0.1028
73
74 # Results
75 list(
76   fpw_fpl_base = c(prob_fp_w_fpl_base),
77   fpw_prob_change = c(low = prob_fp_w_low, high = prob_fp_w_high),
78   fpl_prob_change = c(low = prob_fpl_low, high = prob_fpl_high),
79   thresholds = c(fpw = fpw_threshold, fpl = fpl_threshold)
80 )

```

Listing 4: This analysis examines the relationship between copulation status and two morphological traits, forewing patch length (fpl) and forewing patch width (fpw), in male *Calopteryx splendens*. Logistic regression models were fitted to predict copulation status as a function of fpl, fpw, their combination, and their interaction. Key results include: Pseudo R-squared and Tjur's D, Predicted Probabilities and Thresholds for Copulation. This approach quantifies the independent and combined effects of fpl and fpw, offering insights into their roles in sexual selection.

```

1
2 # Clear the environment
3 rm(list = ls())
4
5 # Import libraries
6 library(car)
7
8 # Read the data and filter accordingly
9 male_CS = read.csv("../data/male_CS.csv")
10 male_CS = male_CS %>%
11   select(tbl, abl, thorl,thorw,fwl,hwl,cop) %>%
12   mutate(cop = as.factor(cop))
13
14 # Calculate VIF: VIF > 5 or 10: Indicates multicollinearity; consider removing the variable.
15 # Fit an initial logistic regression model
16 initial_model = glm(cop ~ tbl + abl + thorl + thorw + fwl, data = male_CS, family = binomial) #
17   Removed hwl (hind wing length) as that had the largest value.
18
19 # Calculate VIF
20 vif(initial_model)
21
22 # Stepwise selection
23 final_model = stepAIC(cop ~ tbl + abl + thorl + thorw + fwl,
24   data = male_CS, family = binomial),
25   direction = "both")
26
27 #summary(final_model) The interaction model has a better AIC score. So we will analyze only that.
28
29 # Step 3: Add interaction terms
30 interaction_model = glm(cop ~ (thorl * thorw) + fwl + tbl,
31   data = male_CS, family = binomial)
32 summary(interaction_model)
33
34 # Confusion Matrix
35 male_CS = male_CS %>%
36   mutate(predicted_prob = predict(interaction_model, type = "response"),
37     predicted_class = ifelse(predicted_prob > 0.5, 1, 0))
38
39 # Pseudo R-squared (using MuMIn)
40 r.squaredGLMM(interaction_model)
41
42 # Tjur's D
43 y_hat = predict(interaction_model, type = "response") # Get predicted probabilities

```

```

44
45 tjur_d = mean(y_hat[male_CS$cop == 1]) - mean(y_hat[male_CS$cop == 0])
46 tjur_d

```

Listing 5: Logistic regression analysis of male *Calopteryx splendens* copulation status, incorporating key morphological traits and interaction terms. Multicollinearity was addressed, and model fit was evaluated using stepwise selection, Pseudo R-squared, and Tjur's D.

```

1
2 # Clear the environment
3 rm(list = ls())
4
5 # Read the data and filter
6 male_CS = read.csv("../data/male_CS.csv") %>%
7   select(tbl, abl, thorl, thorw, fwl, hwl, cop) %>%
8   mutate(cop = as.factor(cop))
9
10 # Interaction model
11 interaction_model = glm(cop ~ (thorl * thorw) + fwl + tbl,
12   data = male_CS, family = binomial)
13
14 # Mean and standard deviation of predictors
15 thorl_mean = mean(male_CS$thorl, na.rm = TRUE)
16 thorl_sd = sd(male_CS$thorl, na.rm = TRUE)
17 thorw_mean = mean(male_CS$thorw, na.rm = TRUE)
18 thorw_sd = sd(male_CS$thorw, na.rm = TRUE)
19 fwl_mean = mean(male_CS$fwl, na.rm = TRUE)
20 fwl_sd = sd(male_CS$fwl, na.rm = TRUE)
21 tbl_mean = mean(male_CS$tbl, na.rm = TRUE)
22 tbl_sd = sd(male_CS$tbl, na.rm = TRUE)
23
24 # Define logistic function
25 logit_to_prob = function(log_odds) {
26   return(1 / (1 + exp(-log_odds)))
27 }
28
29 # Base probability
30 prob_base = logit_to_prob(-9.51364 +
31   (1.44451 * thorl_mean) +
32   (3.29692 * thorw_mean) +
33   (0.22905 * fwl_mean) +
34   (-0.16208 * tbl_mean) +
35   (-0.61031 * thorl_mean * thorw_mean))
36
37 # Predicted probabilities for thorl wrt SD
38 prob_thorl_low = logit_to_prob(-9.51364 +
39   (1.44451 * (thorl_mean - thorl_sd)) +
40   (3.29692 * thorw_mean) +
41   (0.22905 * fwl_mean) +
42   (-0.16208 * tbl_mean) +
43   (-0.61031 * (thorl_mean - thorl_sd) * thorw_mean))
44
45 prob_thorl_high = logit_to_prob(-9.51364 +
46   (1.44451 * (thorl_mean + thorl_sd)) +
47   (3.29692 * thorw_mean) +
48   (0.22905 * fwl_mean) +
49   (-0.16208 * tbl_mean) +
50   (-0.61031 * (thorl_mean + thorl_sd) * thorw_mean))
51
52 # Predicted probabilities for thorw wrt SD
53 prob_thorw_low = logit_to_prob(-9.51364 +
54   (1.44451 * thorl_mean) +

```

```

55         (3.29692 * (thorw_mean - thorw_sd)) +
56         (0.22905 * fwl_mean) +
57         (-0.16208 * tbl_mean) +
58         (-0.61031 * thorl_mean * (thorw_mean - thorw_sd)))
59
60 prob_thorw_high = logit_to_prob(-9.51364 +
61         (1.44451 * thorl_mean) +
62         (3.29692 * (thorw_mean + thorw_sd)) +
63         (0.22905 * fwl_mean) +
64         (-0.16208 * tbl_mean) +
65         (-0.61031 * thorl_mean * (thorw_mean + thorw_sd)))
66
67 # Predicted probabilities for fwl wrt SD
68 prob_fwl_low = logit_to_prob(-9.51364 +
69         (1.44451 * thorl_mean) +
70         (3.29692 * thorw_mean) +
71         (0.22905 * (fwl_mean - fwl_sd)) +
72         (-0.16208 * tbl_mean) +
73         (-0.61031 * thorl_mean * thorw_mean))
74
75 prob_fwl_high = logit_to_prob(-9.51364 +
76         (1.44451 * thorl_mean) +
77         (3.29692 * thorw_mean) +
78         (0.22905 * (fwl_mean + fwl_sd)) +
79         (-0.16208 * tbl_mean) +
80         (-0.61031 * thorl_mean * thorw_mean))
81
82 # Predicted probabilities for tbl wrt SD
83 prob_tbl_low = logit_to_prob(-9.51364 +
84         (1.44451 * thorl_mean) +
85         (3.29692 * thorw_mean) +
86         (0.22905 * fwl_mean) +
87         (-0.16208 * (tbl_mean - tbl_sd)) +
88         (-0.61031 * thorl_mean * thorw_mean))
89
90 prob_tbl_high = logit_to_prob(-9.51364 +
91         (1.44451 * thorl_mean) +
92         (3.29692 * thorw_mean) +
93         (0.22905 * fwl_mean) +
94         (-0.16208 * (tbl_mean + tbl_sd)) +
95         (-0.61031 * thorl_mean * thorw_mean))
96
97
98 # Thresholds for P(cop=1)=0.5
99 thorl_threshold = (-(-9.51364 + (3.29692 * thorw_mean) +
100         (0.22905 * fwl_mean) +
101         (-0.16208 * tbl_mean)) /
102         (1.44451 + (-0.61031 * thorw_mean)))
103
104 thorw_threshold = (-(-9.51364 + (1.44451 * thorl_mean) +
105         (0.22905 * fwl_mean) +
106         (-0.16208 * tbl_mean)) /
107         (3.29692 + (-0.61031 * thorl_mean)))
108
109 # Threshold for forewing length (fwl)
110 fwl_threshold = (-(-(-9.51364) +
111         (1.44451 * thorl_mean) +
112         (3.29692 * thorw_mean) +
113         (-0.16208 * tbl_mean) +
114         (-0.61031 * thorl_mean * thorw_mean)) /
115         0.22905)
116
117 # Threshold for tibia length (tbl)

```

```

118 tbl_threshold = (-(-(-9.51364) +
119     (1.44451 * thorl_mean) +
120     (3.29692 * thorw_mean) +
121     (0.22905 * fwl_mean) +
122     (-0.61031 * thorl_mean * thorw_mean)) /
123     -0.16208)
124
125 # Results
126 list(
127     base_probability = prob_base,
128     thorl_prob_change = c(low = prob_thorl_low, high = prob_thorl_high),
129     thorw_prob_change = c(low = prob_thorw_low, high = prob_thorw_high),
130     fwl_prob_change = c(low = prob_fwl_low, high = prob_fwl_high),
131     tbl_prob_change = c(low = prob_tbl_low, high = prob_tbl_high),
132     thresholds = c(thorl = thorl_threshold, thorw = thorw_threshold, fwl = fwl_threshold, tbl =
        tbl_threshold)
133 )

```

Listing 6: Predicted probabilities and thresholds for copulation status based on variations in thorax length, thorax width, forewing length, and tibia length in *Calopteryx splendens*. Results highlight changes relative to ± 1 SD and thresholds for $P(\text{cop}=1)=0.5$.

```

1
2 # Clear environment
3 rm(list = ls())
4
5 # Read the data and filter accordingly
6 male_CS = read.csv("../data/male_CS.csv")
7 male_CS = male_CS %>%
8     mutate(cop = as.factor(cop))
9
10 # Logistic regression model
11 # Note to self - I have tried all these model to get an AIC score below 1519.9 and it is just not
    possible -:
12 # # Model 1: Simple Main Effects Model (Body Size and Wing Length)
13 #model1 = glm(cop ~ tbl + thorl + thorw + fwl + hwl, data = male_CS, family = binomial)
14 # Model 2: Adding Interaction Terms Between Size and Shape
15 #model2 = glm(cop ~ tbl * thorl + thorl * thorw + fwl * hwl + tbl * fwl, data = male_CS, family =
    binomial)
16 # Model 3: Including Lifespan
17 #model3 = glm(cop ~ tbl + thorl + thorw + fwl + lifespan, data = male_CS, family = binomial)
18 # Model 4: Combining Forewing Patch Length and Width
19 #model4 = glm(cop ~ tbl + fpl * fpw + thorl + thorw + fwl, data = male_CS, family = binomial)
20 # Model 5: Full Model with All Interactions
21 #model5 = glm(cop ~ tbl * abl * thorl * thorw * fwl * hwl * fpl * fpw * lifespan, data = male_CS,
    family = binomial)
22 # Model 6: Size and Shape Focused Model
23 #model6 = glm(cop ~ thorl * thorw + fpl * fpw + tbl + lifespan, data = male_CS, family = binomial)
24 # Model 7: Testing the Effects of Individual Wing Dimensions
25 #model7 = glm(cop ~ fwl * fpl + fwl * fpw + hwl + tbl + thorl, data = male_CS, family = binomial)
26 # Model 8: Alternative Model with Only Morphological Variables
27 #model8 = glm(cop ~ tbl + thorl * thorw + fwl * hwl + fpl + fpw, data = male_CS, family = binomial)
28
29 comprehensive_model = glm(cop ~ fpw + fpl + (thorl * thorw) + fwl + tbl,
30     data = male_CS, family = binomial)
31
32 # Summary of the model
33 summary(comprehensive_model)
34
35
36 # Pseudo R-squared (using MuMIn)
37 r.squaredGLMM(comprehensive_model)

```

```

38
39 # Tjur's D
40 y_hat = predict(comprehensive_model, type = "response") # Get predicted probabilities
41
42 tjur_d = mean(y_hat[male_CS$cop == 1]) - mean(y_hat[male_CS$cop == 0])
43 tjur_d
44
45 # Note to self:
46 # I tried scaling the predictor variables and applying standard scaler, and nothing change AIC was
    the same and the thresholds also were the same but were scaled differently. So I don't think I
    need to perform any scaling for this.

```

Listing 7: Analysis of copulation success in *Calopteryx splendens* using a comprehensive logistic regression model. Key predictors include forewing patch dimensions, thorax length and width interactions, forewing length, and tibia length. Model performance metrics: pseudo R^2 (variance explained) and Tjur's D (separation of predicted probabilities between copulation outcomes). Scaling predictors did not impact AIC or thresholds, affirming robustness of raw data.

```

1
2 # Clear environment
3 rm(list = ls())
4
5 # Read the data and filter
6 male_CS = read.csv("../data/male_CS.csv") %>%
7   mutate(cop = as.factor(cop))
8
9 # Comprehensive Interaction Model
10 comprehensive_model = glm(cop ~ fpw + fpl + (thorl * thorw) + fwl + tbl,
11                           data = male_CS, family = binomial)
12
13 # Extract coefficients
14 intercept = -9.9211330
15 coef_fpw = -0.3765803
16 coef_fpl = 0.1691122
17 coef_thorl = 1.5277422
18 coef_thorw = 3.4099390
19 coef_fwl = 0.2536333
20 coef_tbl = -0.1642272
21 coef_thorl_thorw = -0.6260531
22
23 # Mean and standard deviation of predictors
24 fpw_mean = mean(male_CS$fpw, na.rm = TRUE)
25 fpl_mean = mean(male_CS$fpl, na.rm = TRUE)
26 thorl_mean = mean(male_CS$thorl, na.rm = TRUE)
27 thorw_mean = mean(male_CS$thorw, na.rm = TRUE)
28 fwl_mean = mean(male_CS$fwl, na.rm = TRUE)
29 tbl_mean = mean(male_CS$tbl, na.rm = TRUE)
30
31 fpw_sd = sd(male_CS$fpw, na.rm = TRUE)
32 fpl_sd = sd(male_CS$fpl, na.rm = TRUE)
33 thorl_sd = sd(male_CS$thorl, na.rm = TRUE)
34 thorw_sd = sd(male_CS$thorw, na.rm = TRUE)
35 fwl_sd = sd(male_CS$fwl, na.rm = TRUE)
36 tbl_sd = sd(male_CS$tbl, na.rm = TRUE)
37
38 # Define logistic function
39 logit_to_prob = function(log_odds) {
40   return(1 / (1 + exp(-log_odds)))
41 }
42
43 # Base probability
44 prob_base = logit_to_prob(intercept +
45                           (coef_fpw * fpw_mean) +

```

```

46         (coef_fpl * fpl_mean) +
47         (coef_thorl * thorl_mean) +
48         (coef_thorw * thorw_mean) +
49         (coef_fwl * fwl_mean) +
50         (coef_tbl * tbl_mean) +
51         (coef_thorl_thorw * thorl_mean * thorw_mean))
52
53 # Predicted probabilities for fpw wrt SD
54 prob_fpw_low = logit_to_prob(intercept +
55                             (coef_fpw * (fpw_mean - fpw_sd)) +
56                             (coef_fpl * fpl_mean) +
57                             (coef_thorl * thorl_mean) +
58                             (coef_thorw * thorw_mean) +
59                             (coef_fwl * fwl_mean) +
60                             (coef_tbl * tbl_mean) +
61                             (coef_thorl_thorw * thorl_mean * thorw_mean))
62
63 prob_fpw_high = logit_to_prob(intercept +
64                              (coef_fpw * (fpw_mean + fpw_sd)) +
65                              (coef_fpl * fpl_mean) +
66                              (coef_thorl * thorl_mean) +
67                              (coef_thorw * thorw_mean) +
68                              (coef_fwl * fwl_mean) +
69                              (coef_tbl * tbl_mean) +
70                              (coef_thorl_thorw * thorl_mean * thorw_mean))
71
72 # Predicted probabilities for fpl wrt SD
73 prob_fpl_low = logit_to_prob(intercept +
74                             (coef_fpw * fpw_mean) +
75                             (coef_fpl * (fpl_mean - fpl_sd)) +
76                             (coef_thorl * thorl_mean) +
77                             (coef_thorw * thorw_mean) +
78                             (coef_fwl * fwl_mean) +
79                             (coef_tbl * tbl_mean) +
80                             (coef_thorl_thorw * thorl_mean * thorw_mean))
81
82 prob_fpl_high = logit_to_prob(intercept +
83                              (coef_fpw * fpw_mean) +
84                              (coef_fpl * (fpl_mean + fpl_sd)) +
85                              (coef_thorl * thorl_mean) +
86                              (coef_thorw * thorw_mean) +
87                              (coef_fwl * fwl_mean) +
88                              (coef_tbl * tbl_mean) +
89                              (coef_thorl_thorw * thorl_mean * thorw_mean))
90
91 # Predicted probabilities for thorl wrt SD
92 prob_thorl_low = logit_to_prob(intercept +
93                               (coef_fpw * fpw_mean) +
94                               (coef_fpl * fpl_mean) +
95                               (coef_thorl * (thorl_mean - thorl_sd)) +
96                               (coef_thorw * thorw_mean) +
97                               (coef_fwl * fwl_mean) +
98                               (coef_tbl * tbl_mean) +
99                               (coef_thorl_thorw * (thorl_mean - thorl_sd) * thorw_mean))
100
101 prob_thorl_high = logit_to_prob(intercept +
102                                (coef_fpw * fpw_mean) +
103                                (coef_fpl * fpl_mean) +
104                                (coef_thorl * (thorl_mean + thorl_sd)) +
105                                (coef_thorw * thorw_mean) +
106                                (coef_fwl * fwl_mean) +
107                                (coef_tbl * tbl_mean) +
108                                (coef_thorl_thorw * (thorl_mean + thorl_sd) * thorw_mean))

```

```

109
110 # Predicted probabilities for thorw wrt SD
111 prob_thorw_low = logit_to_prob(intercept +
112                                (coef_fpw * fpw_mean) +
113                                (coef_fpl * fpl_mean) +
114                                (coef_thorl * thorl_mean) +
115                                (coef_thorw * (thorw_mean - thorw_sd)) +
116                                (coef_fwl * fwl_mean) +
117                                (coef_tbl * tbl_mean) +
118                                (coef_thorl_thorw * thorl_mean * (thorw_mean - thorw_sd)))
119
120 prob_thorw_high = logit_to_prob(intercept +
121                                 (coef_fpw * fpw_mean) +
122                                 (coef_fpl * fpl_mean) +
123                                 (coef_thorl * thorl_mean) +
124                                 (coef_thorw * (thorw_mean + thorw_sd)) +
125                                 (coef_fwl * fwl_mean) +
126                                 (coef_tbl * tbl_mean) +
127                                 (coef_thorl_thorw * thorl_mean * (thorw_mean + thorw_sd)))
128
129 # Predicted probabilities for fwl wrt SD
130 prob_fwl_low = logit_to_prob(intercept +
131                               (coef_fpw * fpw_mean) +
132                               (coef_fpl * fpl_mean) +
133                               (coef_thorl * thorl_mean) +
134                               (coef_thorw * thorw_mean) +
135                               (coef_fwl * (fwl_mean - fwl_sd)) +
136                               (coef_tbl * tbl_mean) +
137                               (coef_thorl_thorw * thorl_mean * thorw_mean))
138
139 prob_fwl_high = logit_to_prob(intercept +
140                               (coef_fpw * fpw_mean) +
141                               (coef_fpl * fpl_mean) +
142                               (coef_thorl * thorl_mean) +
143                               (coef_thorw * thorw_mean) +
144                               (coef_fwl * (fwl_mean + fwl_sd)) +
145                               (coef_tbl * tbl_mean) +
146                               (coef_thorl_thorw * thorl_mean * thorw_mean))
147
148 # Predicted probabilities for tbl wrt SD
149 prob_tbl_low = logit_to_prob(intercept +
150                               (coef_fpw * fpw_mean) +
151                               (coef_fpl * fpl_mean) +
152                               (coef_thorl * thorl_mean) +
153                               (coef_thorw * thorw_mean) +
154                               (coef_fwl * fwl_mean) +
155                               (coef_tbl * (tbl_mean - tbl_sd)) +
156                               (coef_thorl_thorw * thorl_mean * thorw_mean))
157
158 prob_tbl_high = logit_to_prob(intercept +
159                               (coef_fpw * fpw_mean) +
160                               (coef_fpl * fpl_mean) +
161                               (coef_thorl * thorl_mean) +
162                               (coef_thorw * thorw_mean) +
163                               (coef_fwl * fwl_mean) +
164                               (coef_tbl * (tbl_mean + tbl_sd)) +
165                               (coef_thorl_thorw * thorl_mean * thorw_mean))
166
167
168 # Threshold for fpw
169 fpw_threshold = -(
170     intercept +
171     (coef_fpl * fpl_mean) +

```



```

172 (coef_thorl * thorl_mean) +
173 (coef_thorw * thorw_mean) +
174 (coef_fwl * fwl_mean) +
175 (coef_tbl * tbl_mean) +
176 (coef_thorl_thorw * thorl_mean * thorw_mean)
177 ) / coef_fpw)
178
179 # Threshold for fpl
180 fpl_threshold = -(
181 intercept +
182 (coef_fpw * fpw_mean) +
183 (coef_thorl * thorl_mean) +
184 (coef_thorw * thorw_mean) +
185 (coef_fwl * fwl_mean) +
186 (coef_tbl * tbl_mean) +
187 (coef_thorl_thorw * thorl_mean * thorw_mean)
188 ) / coef_fpl)
189
190 # Threshold for thorl
191 thorl_threshold = -(
192 intercept +
193 (coef_fpw * fpw_mean) +
194 (coef_fpl * fpl_mean) +
195 (coef_thorw * thorw_mean) +
196 (coef_fwl * fwl_mean) +
197 (coef_tbl * tbl_mean)
198 ) / (coef_thorl + (coef_thorl_thorw * thorw_mean)))
199
200 # Threshold for thorw
201 thorw_threshold = -(
202 intercept +
203 (coef_fpw * fpw_mean) +
204 (coef_fpl * fpl_mean) +
205 (coef_thorl * thorl_mean) +
206 (coef_fwl * fwl_mean) +
207 (coef_tbl * tbl_mean)
208 ) / (coef_thorw + (coef_thorl_thorw * thorl_mean)))
209
210 # Threshold for fwl
211 fwl_threshold = -(
212 intercept +
213 (coef_fpw * fpw_mean) +
214 (coef_fpl * fpl_mean) +
215 (coef_thorl * thorl_mean) +
216 (coef_thorw * thorw_mean) +
217 (coef_tbl * tbl_mean) +
218 (coef_thorl_thorw * thorl_mean * thorw_mean)
219 ) / coef_fwl)
220
221 # Threshold for tbl
222 tbl_threshold = -(
223 intercept +
224 (coef_fpw * fpw_mean) +
225 (coef_fpl * fpl_mean) +
226 (coef_thorl * thorl_mean) +
227 (coef_thorw * thorw_mean) +
228 (coef_fwl * fwl_mean) +
229 (coef_thorl_thorw * thorl_mean * thorw_mean)
230 ) / coef_tbl)
231
232
233 # Results
234 list(

```

```

235 base_probability = prob_base,
236 fpw_prob_change = c(low = prob_fpw_low, high = prob_fpw_high),
237 fpl_prob_change = c(low = prob_fpl_low, high = prob_fpl_high),
238 thorl_prob_change = c(low = prob_thorl_low, high = prob_thorl_high),
239 thorw_prob_change = c(low = prob_thorw_low, high = prob_thorw_high),
240 fwl_prob_change = c(low = prob_fwl_low, high = prob_fwl_high),
241 tbl_prob_change = c(low = prob_tbl_low, high = prob_tbl_high),
242 thresholds = c(thorl = thorl_threshold, thorw = thorw_threshold, fwl = fwl_threshold, tbl =
          tbl_threshold, fpw = fpw_threshold, fpl = fpl_threshold)
243 )

```

Listing 8: R code for biological analysis for the combined model.

```

1
2 # Data for plotting predicted probabilities with SD and Base
3 plot_data = data.frame(
4   Variable = rep(c("fpw", "fpl", "thorl", "thorw", "fwl", "tbl"), each = 3),
5   Level = rep(c("Below SD", "Base", "Above SD"), times = 6),
6   Probability = c(
7     prob_fpw_low, prob_base, prob_fpw_high,
8     prob_fpl_low, prob_base, prob_fpl_high,
9     prob_thorl_low, prob_base, prob_thorl_high,
10    prob_thorw_low, prob_base, prob_thorw_high,
11    prob_fwl_low, prob_base, prob_fwl_high,
12    prob_tbl_low, prob_base, prob_tbl_high
13  )
14 )
15
16 # Adjust the order of factor levels to make "Below SD" to the left, and "Above SD" to the right of
   "Base"
17 plot_data$Level = factor(plot_data$Level, levels = c("Below SD", "Base", "Above SD"))
18
19 # Create the plot
20 plot = ggplot(plot_data, aes(x = Variable, y = Probability, fill = Level)) +
21   geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
22   labs(
23     title = "Predicted Probabilities for Predictor Variables",
24     y = "Probability",
25     x = "Predictor Variable"
26   ) +
27   scale_fill_manual(
28     values = c("Below SD" = "#1f77b4", "Base" = "#f2a900", "Above SD" = "#d62728"),
29     breaks = c("Below SD", "Base", "Above SD"),
30     labels = c("Below SD", "Base", "Above SD")
31   ) +
32   theme_classic() +
33   theme(
34     legend.position = "top",
35     plot.title = element_text(hjust = 0.5, size = 18),
36     axis.text = element_text(size = 14),
37     axis.title = element_text(size = 16)
38   )
39
40 # Display the plot
41 print(plot)
42
43 # Save the plot as a high-resolution PNG
44 ggsave(
45   filename = "../results/predicted_probabilities_plot.png",
46   plot = plot,
47   width = 200, # Width in mm
48   height = 80, # Adjusted height for document proportions
49   units = "mm",

```

```

50   dpi = 600
51 )
52
53 # Generate grid of values for thorl and thorw
54 thorl_seq = seq(thorl_mean - 2 * thorl_sd, thorl_mean + 2 * thorl_sd, length.out = 50)
55 thorw_seq = seq(thorw_mean - 2 * thorw_sd, thorw_mean + 2 * thorw_sd, length.out = 50)
56
57 interaction_data = expand.grid(thorl = thorl_seq, thorw = thorw_seq)
58
59 # Predict probabilities
60 interaction_data$Probability = logit_to_prob(
61   intercept +
62     coef_thorl * interaction_data$thorl +
63     coef_thorw * interaction_data$thorw +
64     coef_thorl_thorw * interaction_data$thorl * interaction_data$thorw +
65     coef_fpw * fpw_mean +
66     coef_fpl * fpl_mean +
67     coef_fwl * fwl_mean +
68     coef_tbl * tbl_mean
69 )
70
71 # Create contour plot
72 contour_plot = ggplot(interaction_data, aes(x = thorl, y = thorw, z = Probability)) +
73   geom_contour_filled() +
74   labs(
75     title = "Effect of thorl and thorw Interaction on Predicted Probabilities",
76     x = "thorl (mm)",
77     y = "thorw (mm)",
78     fill = "Probability"
79   ) +
80   theme_classic() +
81   theme(
82     plot.title = element_text(hjust = 0.3, size = 16), # Centered and adjusted title size
83     axis.title = element_text(size = 16),             # Axis title size
84     axis.text = element_text(size = 14),              # Axis text size
85     legend.title = element_text(size = 14),          # Legend title size
86     legend.text = element_text(size = 12)            # Legend text size
87   )
88
89 # Display the plot
90 print(contour_plot)
91
92 # Save the plot as a high-resolution PNG
93 ggsave(
94   filename = "../results/interaction_contour_plot.png",
95   plot = contour_plot,
96   width = 180, # Width in mm
97   height = 120, # Adjusted height for document proportions
98   units = "mm",
99   dpi = 600
100 )
101
102 # Extract coefficients and confidence intervals
103 coef_summary = summary(comprehensive_model)$coefficients
104 coef_data = data.frame(
105   Term = rownames(coef_summary),
106   Estimate = coef_summary[, "Estimate"],
107   StdError = coef_summary[, "Std. Error"],
108   LowerCI = coef_summary[, "Estimate"] - 1.96 * coef_summary[, "Std. Error"],
109   UpperCI = coef_summary[, "Estimate"] + 1.96 * coef_summary[, "Std. Error"]
110 )
111
112 # Plot coefficients

```

```

113 coef_plot = ggplot(coef_data, aes(x = Term, y = Estimate)) +
114   geom_point(size = 3) +
115   geom_errorbar(aes(ymin = LowerCI, ymax = UpperCI), width = 0.3, color = "#0072B2") +
116   labs(
117     title = "Coefficient Estimates with 95% Confidence Intervals",
118     y = "Estimate",
119     x = "Predictor"
120   ) +
121   theme_classic() +
122   theme(
123     plot.title = element_text(hjust = 0.5, size = 18), # Centered and scaled title
124     axis.title = element_text(size = 16),             # Axis title size
125     axis.text = element_text(size = 14),              # Axis text size
126     panel.grid.major = element_line(linewidth = 0.5), # Adjust grid line thickness
127     panel.grid.minor = element_blank()                # Hide minor grid lines
128   ) +
129   coord_flip() # Flip coordinates for better readability
130
131 # Display the plot
132 print(coef_plot)
133
134 # Save the plot as a high-resolution PNG
135 ggsave(
136   filename = "../results/coefficients_plot.png",
137   plot = coef_plot,
138   width = 200, # Width in mm for LaTeX integration
139   height = 120, # Adjusted height for readability
140   units = "mm",
141   dpi = 600
142 )
143
144 # Create a dataframe for thresholds
145 threshold_data = data.frame(
146   Variable = c("fpw", "fpl", "thorl", "thorw", "fwl", "tbl"),
147   Threshold = c(fpw_threshold, fpl_threshold, thorl_threshold, thorw_threshold, fwl_threshold,
148                 tbl_threshold)
149 )
150
151 # Plot thresholds
152 threshold_plot = ggplot(threshold_data, aes(x = Variable, y = Threshold)) +
153   geom_bar(stat = "identity", fill = "lightblue") +
154   labs(
155     title = "Thresholds for Predictor Variables",
156     y = "Threshold Value",
157     x = "Predictor Variable"
158   ) +
159   theme_classic() +
160   theme(
161     plot.title = element_text(hjust = 0.5, size = 18), # Centered and scaled title
162     axis.title = element_text(size = 16),             # Axis title size
163     axis.text = element_text(size = 14),              # Axis text size
164     panel.grid.major = element_blank(),               # Adjust grid line thickness
165     panel.grid.minor = element_blank()                # Hide minor grid lines
166   )
167
168 # Display the plot
169 print(threshold_plot)
170
171 # Save the plot as a high-resolution PNG
172 ggsave(
173   filename = "../results/threshold_plot.png",
174   plot = threshold_plot,
175   width = 200, # Width in mm for LaTeX integration

```

```

175     height = 120, # Adjusted height for readability
176     units = "mm",
177     dpi = 600
178 )
179
180 # Read the data and filter accordingly
181 male_CS = read.csv("../data/male_CS.csv")
182 scatter_plot = male_CS %>%
183   mutate(cop = as.factor(cop)) %>%
184   ggplot(aes(x = thorw, y = thorl, colour = cop)) +
185   geom_point(size = 3, alpha = 0.8) + # Adjusted point size and transparency
186   theme_classic() +
187   labs(
188     title = "Scatter Plot of thorl vs thorw by Copulation Status",
189     x = "thorw (mm)", # Added unit for clarity
190     y = "thorl (mm)", # Added unit for clarity
191     colour = "Copulation Status"
192   ) +
193   theme(
194     legend.position = "top", # Position legend at the top
195     plot.title = element_text(hjust = 0.5, size = 18), # Centered and scaled title
196     axis.title = element_text(size = 16), # Adjusted axis title size
197     axis.text = element_text(size = 14) # Adjusted axis text size
198   )
199
200 # Display the plot
201 print(scatter_plot)
202
203
204 # Save the plot as a high-resolution PNG
205 ggsave(
206   filename = "../results/thorl_vs_thorw_scatter.png",
207   plot = scatter_plot,
208   width = 200, # Width in mm for LaTeX integration
209   height = 150, # Adjusted height for readability
210   units = "mm",
211   dpi = 600
212 )

```

Listing 9: Plots for the combined model.