Effects of Body and Wing Size on Copulation Success in Damselflies

Introduction:

Damselflies have several morphological traits considered to be under sexual selection. To gain further inside into which specific traits and to what extent they contribute to an individual´s fitness I investigated the effect of three selected traits on copulation success: Total body length, forewing length, and hindwing length. The analysis comprises male and female damselfly individuals of two species: *Calopteryx splendens* and *Calopteryx virgo*. The data used in this study was collected in Sweden and represents a single population that was sampled each year in June and July from 2011-2015.

Research Question:

Do total body length, forewing length, and/or hindwing length affect copulation success in damselflies, and do the effects vary between species and/or sexes?

Analysis Methods:

I chose total body length (tbl), forewing length (fwl), and hindwing length (hwl) as variables to represent overall body size and wing size. All were measured in mm. Since body measurements are often correlated, and multicollinearity reduces the precision of estimated coefficients, I analysed each variable in a separate model to avoid misleading results.

I chose copulation status (cop) as the response variable to represent mating success and fitness of an individual. It was noted with either a 0, indicating the individual was never found in copulation with a partner, or a 1, indicating the individual was observed at least once in copulation. Given that the response variable is binary (0 or 1), I chose to fit a generalized linear model (GLM) with a logit link function to adjust for the binomial distribution (logistic regression).

I initially included lifespan as a covariate, to account for higher copulation success due to longer time spans for mating as well as increased observation time during the data collection. However, comparing the models with and without including lifespan as a covariate showed that there was no significant improvement of any of the models ($p > 0.05$). Considering that the models are nested, I did the comparisons using ANOVA (see Appendix 1). To reduce model complexity, I continued with the more parsimonious models, excluding lifespan as a covariate.

Since the provided dataset included both males and females from two species, I also tested whether the effects of tbl, fwl, and hwl on copulation success differed by sex and/or species by including these variables as interaction terms. The only significant effect was found for the variable 'species' in the hwl model (see Appendix 2). Hence, I kept species as an interaction term in the hwl model and continued with unchanged models for tbl and fwl.

For all variables the data was checked for NA values, none contained any.

The final binomial GLMs I selected for the data analysis are as follows:

cop ~ tbl

cop ~ fwl

cop ~ hwl * species

Results:

The final tbl-GLM indicates that total body length does not significantly influence copulation success in *C. splendens* and *C. virgo* (p > 0.05) (see Fig 1). The model predicts that, with 1 mm increase in tbl, the chances for copulation decrease by approximately 3.7%.

The final GLM for forewing length shows a significant positive effect on copulation success in *C. splendens* and *C. virgo* (p < 0.05) (see Fig 2). A 1 mm increase in fwl results in a 32.3% increase of the chances for copulation success.

The final GLM for hindwing length also shows a significant positive effect on copulation success (p < 0.05) with a species-specific variation (interaction p < 0.05) (see Fig 3). A 1mm increase in hwl results in *C. splendens* in a 27.24% and in *C. virgo* in a 41.23% increase of the chances for copulation success.

In summary, the models predict that fwl and hwl significantly influence copulation success, with hwl showing species-specific differences, while tbl has a negligible impact. However, these models have very low predictive and explanatory power, indicated by their low pseudo-R² values (see Table 2). None of the models explain more than 7% of the variance in copulation.

Table 1: Summary Statistics of the variables

Variable	Count	Mean	Min	Max
tbl		44.2	37.83	52.8
fwl		30.71	20.24	38.88
hwl		29.94	14	38.45
lifespan		2.022	0	41
cop = 0	2777			
cop = 1	841			
CSf	990			
CSm	2035			
CVf	244			
CVm	349			

Basic summary statistics of morphological traits and lifespan and counts for data groups. Total body length (tbl), forewing length (fwl), and hindwing length (hwl) are given in mm and rounded to 2 decimal places. Lifespan is given in days. Copulation status (cop) shows the number of individuals never observed in copulation (cop=0) and of individuals found at least once in copulation (cop=1). CSf, CSm, CVf, and CVm state the sample sizes by sex and species. CS refers to *Calopteryx splendens* and CV to *Calopteryx virgo*, f refers to females and m to males.

Table 2: Logistic Regression Models Outcomes

Predictor	Odd Ratio	p-value	Pseudo R^2
tbl	-0.04	0.074	0.0008
fwl	0.32	2.50e-53	0.0668
hwl	CS:0.27 CV:0.41	2.29e-28	0.0560
sp	-0.97	0.038	
hwl : sp	0.11	0.049	

Results of the logistic regression models predicting copulation success by total body length (tbl), forewing length (fwl), hindwing length (hwl), species (sp), and the interaction between hwl and sp (hwl:sp). The odd ratios show the change for a 1-unit increase in a predictor; the values are rounded to 2 decimal places. The p-values state the significance level, p-values <0.05 being considered statistically significant, and are rounded to 3 decimal places. The pseudo R^2 values indicate the fit of each model, i.e., the amount of variance in the data explained by the model.
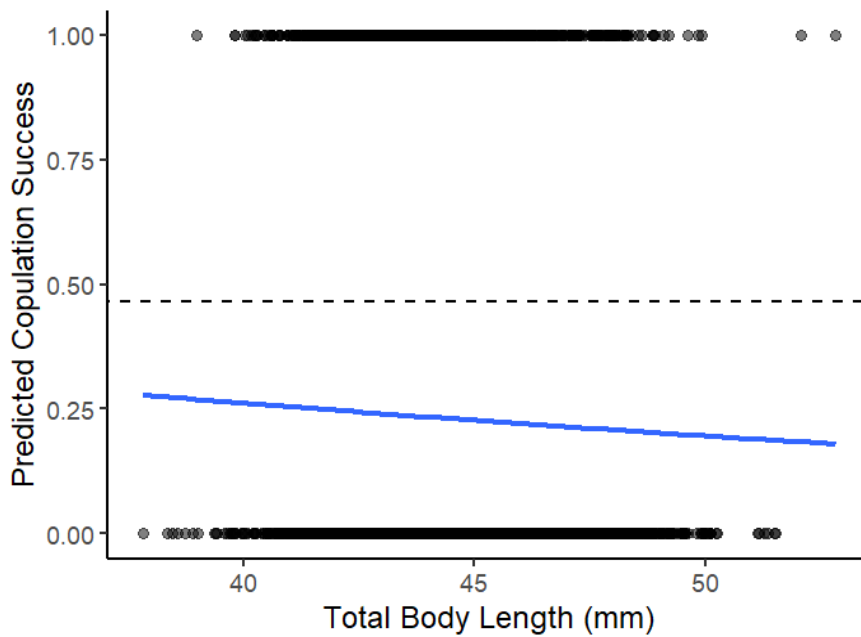
Figure 1: Predicted copulation success based on total body length
The blue line represents the fitted values of the logistic regression model and shows the predicted probability of copulation success depending on the total body length. Its flat slope indicates little effect of the predictor. The black, dashed line represents the intercept and shows the copulation success if no variables were considered.
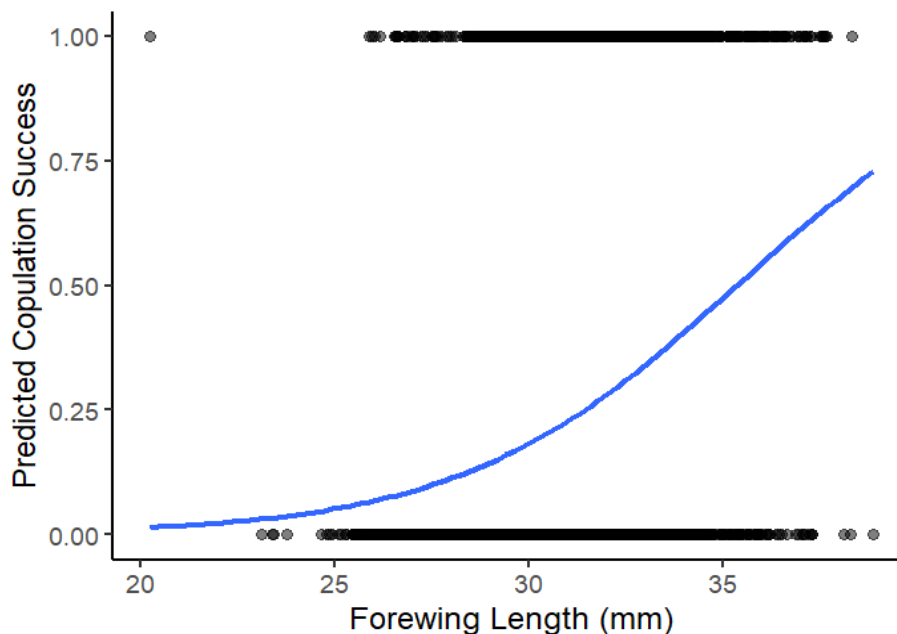

Figure 2: Predicted copulation success based on forewing length
The blue line represents the fitted values of the logistic regression model and shows the predicted probability of copulation success depending on the forewing length. Its slope suggests that the predictor has a significant effect. The intercept is not visible in this plot as it located at ~ -10, contrasting the difference the predictor makes to the baseline of the model.
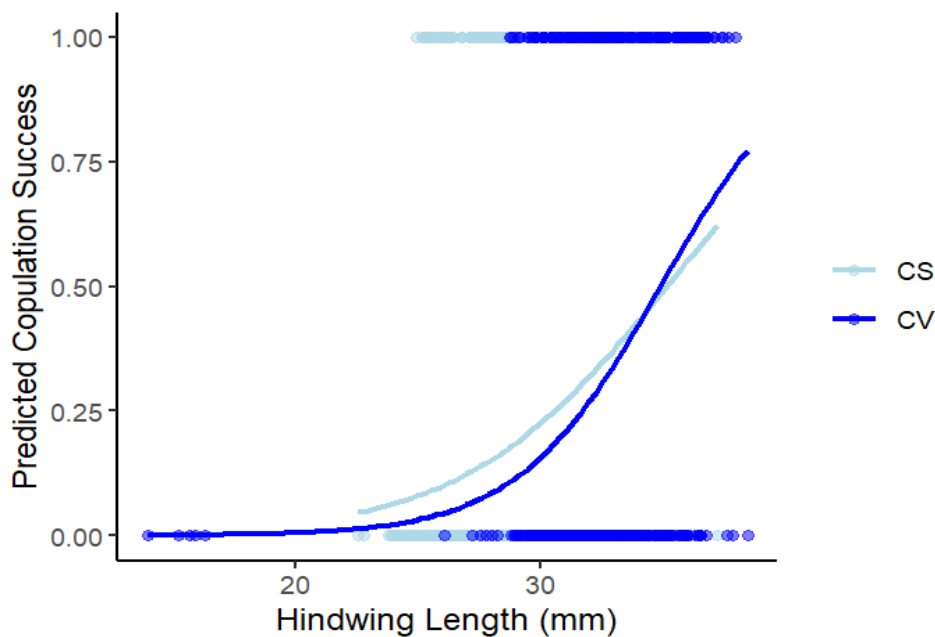
Figure 3: Predicted copulation success based on hindwing length for *C. splendens* and *C. virgo*
The blue lines represent the fitted values of the logistic regression model and show the predicted probability of copulation success depending on the hindwing length. The light blue line represents *Calopteryx splendens*, the dark blue line *Calopteryx virgo*. Their slopes suggest that the predictor has a significant effect in both species. The intercept is not visible in this plot as it located at ~ -8.5, contrasting the difference the predictor makes to the baseline of the model.

Discussion:

The results indicate that forewing length and hindwing length, but not total body length, significantly affect copulation success in *Calopteryx splendens* and *Calopteryx virgo*. This supports the hypothesis that wing morphology influences an individual´s fitness and is subject to sexual selection. The effect of hindwing length varied between species which could be due to different evolutionary pressures and might have contributed to the speciation process, shaping the morphological traits of the two species in different ways. The results do not support the hypothesis that body size influences copulation success. A more detailed analysis of the different main body segments could reveal if the single parts influence copulation success. Against the initial expectations, lifespan did not significantly influence copulation success. This, however, could be due to the collected data. Lifespan depended on sighting an individual again and did not necessarily reflect the actual 'lifespan'. Additionally, the data distribution and resolution could have caused problems as well as assigning a 0-lifespan to all individuals not sighted more than once.
The sex of an individual has also shown no significant effect or interaction in any of the final models, suggesting that the observed morphological traits influence female and male copulation success similarly. As all models show a weak fit, the exact magnitude of the variable effects should be considered with caution. For further research and analysis more variables and factors should be monitored and considered for improvement of the model fits. Nevertheless, the conclusions on which of the three investigated variables are significant and relevant remain valid.

In conclusion, forewing length and hindwing length appear to be important predictors of copulation success in *Calopteryx splendens* and *Calopteryx virgo*, with no sex-specific differences and with species-specific differences observed only for hindwing length. Overall, these results add to our understanding of the evolution of damselfly species and highlight the complexity of their morphological traits under sexual selection.

1. ANOVA outputs

```
Model 1: cop ~ tbl
Model 2: cop ~ tbl + lifespan
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      3616      3920.3
2      3615      3920.2  1 0.086301   0.7689


Model 1: cop ~ fwl
Model 2: cop ~ fwl + lifespan
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      3616      3661.4
2      3615      3660.9  1  0.52122   0.4703


Model 1: cop ~ hwl
Model 2: cop ~ hwl + lifespan
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      3616      3710.8
2      3615      3710.3  1  0.42741   0.5133
```

2. Coefficient outputs for models with interaction terms

Species:

```
glm(formula = cop ~ tbl * sp, family = binomial, data = all_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.64736    1.31682   4.289 1.80e-05 ***
tbl         -0.15902    0.03023  -5.259 1.45e-07 ***
spCV        -0.59993    2.63110  -0.228    0.820
tbl:spCV     0.03414    0.05754   0.593    0.553
---
```

```
glm(formula = cop ~ fwl * sp, family = binomial, data = all_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.13949    0.68752 -14.748   <2e-16 ***
fwl           0.28892    0.02215  13.042   <2e-16 ***
spCV         -1.90977    1.75533  -1.088    0.277
fwl:spCV      0.05081    0.05315   0.956    0.339
---
```

```
glm(formula = cop ~ hwl * sp, family = binomial, data = all_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.46594    0.65785 -12.869   <2e-16 ***
hwl          0.24088    0.02181  11.046   <2e-16 ***
spCV        -3.57905    1.72144  -2.079   0.0376 *
hwl:spCV     0.10435    0.05301   1.968   0.0490 *
---
```

Sex:

```
glm(formula = cop ~ tbl * sex, family = binomial, data = all_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.592891   1.297196  -1.228    0.219
tbl          0.024803   0.029672   0.836    0.403
sexMale     -1.556849   1.898524  -0.820    0.412
tbl:sexMale  0.008453   0.042971   0.197    0.844


glm(formula = cop ~ fwl * sex, family = binomial, data = all_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.71504    1.06947   -6.279 3.41e-10 ***
fwl          0.18890    0.03242    5.827 5.66e-09 ***
sexMale     -1.50342    1.44972   -1.037     0.30
fwl:sexMale  0.03024    0.04586    0.659     0.51
---

glm(formula = cop ~ hwl * sex, family = binomial, data = all_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.98142    0.94756   -6.312 2.75e-10 ***
hwl          0.17122    0.02951    5.803 6.52e-09 ***
sexMale      0.17322    1.27612    0.136    0.892
hwl:sexMale -0.02921    0.04142   -0.705    0.481
---
```

Code:

```
# Total body length: tbl
# Fore wing length: fwl
# Hind wing length: hwl
# Copulation success: cop
# Species: sp

# Imports ####

library(ggplot2)
library(dplyr)
library(MASS)


# Data handling and checks: ####

# Load data frames
CSf =
read.csv("C:/Users/lea/OneDrive/Dokumente/UniLund/BIOS14/ExamBIOS14/fema
le_CS.csv")
CSm =
read.csv("C:/Users/lea/OneDrive/Dokumente/UniLund/BIOS14/ExamBIOS14/male
_CS.csv")
CVf =
read.csv("C:/Users/lea/OneDrive/Dokumente/UniLund/BIOS14/ExamBIOS14/fema
le_CV.csv")
CVm =
read.csv("C:/Users/lea/OneDrive/Dokumente/UniLund/BIOS14/ExamBIOS14/male
_CV.csv")

# Combine into 1 data frame
all_data <- bind_rows(CSf, CSm, CVf, CVm)

# Remove irrelevant columns
all_data <- all_data[, !(colnames(all_data) %in% c("year", "id",
"marked", "abl", "thorl", "thorw", "fpl", "fpw"))]

# Check for missing values
any(is.na(all_data))
# --> no NA values


# Summary statistics: ####

summary(all_data)
all_data %>%
  group_by(sp, sex) %>%
  summarise(count = n())
table(all_data$cop)


# Data exploration: ####

# Basic GLMs for tbl, fwl, and hwl
glm_tbl <- glm(cop ~ tbl, family = binomial, data = all_data)
glm_fwl <- glm(cop ~ fwl, family = binomial, data = all_data)
glm_hwl <- glm(cop ~ hwl, family = binomial, data = all_data)
summary(glm_tbl) # --> no significant effect of tbl on cop (p>0.05)
summary(glm_fwl) # --> significant effect of fwl on cop (p<0.05)
summary(glm_hwl) # --> significant effect of hwl on cop (p<0.05)
```

```
# GLMs with lifespan as a covariate
glm_tbl_lifespan <- glm(cop ~ tbl + lifespan, family = binomial, data =
all_data)
glm_fwl_lifespan <- glm(cop ~ fwl + lifespan, family = binomial, data =
all_data)
glm_hwl_lifespan <- glm(cop ~ hwl + lifespan, family = binomial, data =
all_data)
summary(glm_tbl_lifespan) # --> no significant effect of tbl on cop
(p>0.05)
summary(glm_fwl_lifespan) # --> significant effect of fwl on cop
(p<0.05)
summary(glm_hwl_lifespan) # --> significant effect of hwl on cop
(p<0.05)
# --> lifespan has no significant effect in any of the models (p>0.05)

# Model comparison with ANOVA
anova(glm_tbl, glm_tbl_lifespan, test = "Chisq")
anova(glm_fwl, glm_fwl_lifespan, test = "Chisq")
anova(glm_hwl, glm_hwl_lifespan, test = "Chisq")
# --> no significant difference between models in any case (p>0.05)

# GLMs with species as a factor
glm_tbl_sp <- glm(cop ~ tbl * sp, family = binomial, data = all_data)
glm_fwl_sp <- glm(cop ~ fwl * sp, family = binomial, data = all_data)
glm_hwl_sp <- glm(cop ~ hwl * sp, family = binomial, data = all_data)
summary(glm_tbl_sp) # --> significant effect of tbl on cop (p<0.05)
summary(glm_fwl_sp) # --> significant effect of fwl on cop (p<0.05)
summary(glm_hwl_sp) # --> significant effect of hwl on cop (p<0.05)
# sp has no significant effect in the tbl and fwl models (p>0.05)
# but has a significant effect and interaction in the hwl model (p<0.05)

# GLMs with sex as a factor
glm_tbl_sex <- glm(cop ~ tbl * sex, family = binomial, data = all_data)
glm_fwl_sex <- glm(cop ~ fwl * sex, family = binomial, data = all_data)
glm_hwl_sex <- glm(cop ~ hwl * sex, family = binomial, data = all_data)
summary(glm_tbl_sex) # --> no significant effect of tbl on cop (p>0.05)
summary(glm_fwl_sex) # --> significant effect of fwl on cop (p<0.05)
summary(glm_hwl_sex) # --> significant effect of hwl on cop (p<0.05)
# sex has no significant effect in any of the models (p>0.05)


# Extract values for selected models: ####

# Coefficients & p-values
coef_tbl <- summary(glm_tbl)$coef
exp_coef_tbl <- exp(coef_tbl["tbl", "Estimate"]) - 1
p_value_tbl <- coef_tbl["tbl", "Pr(>|z|)"]
# --> -0.03686729 & 0.074
coef_fwl <- summary(glm_fwl)$coef
exp_coef_fwl <- exp(coef_fwl["fwl", "Estimate"]) - 1
p_value_fwl <- coef_fwl["fwl", "Pr(>|z|)"]
# --> 0.3233487 & 2.5e-53
coef_hwl_sp <- summary(glm_hwl_sp)$coef
exp_coef_hwl_spCS <- exp(coef_hwl_sp["hwl", "Estimate"]) - 1
p_value_hwl_spCS <- coef_hwl_sp["hwl", "Pr(>|z|)"]
# --> 0.2723689 & 2.29e-28
exp_coef_hwl_spCV <- exp(coef_hwl_sp["hwl", "Estimate"] +
coef_hwl_sp["hwl:spCV", "Estimate"]) - 1
# --> 0.4123115
exp_coef_hwl_sp_sp <- exp(coef_hwl_sp["spCV", "Estimate"]) - 1
p_value_hwl_sp_sp <- coef_hwl_sp["spCV", "Pr(>|z|)"]
```

```
# --> -0.9720979 & 0.038
exp_coef_hwl_sp_int <- exp(coef_hwl_sp["hwl:spCV", "Estimate"]) - 1
p_value_hwl_sp_int <- coef_hwl_sp["hwl:spCV", "Pr(>|z|)"]
# --> 0.1099858 & 0.049

# Pseudo-R²
pseudo_r2_tbl <- 1 - (glm_tbl$deviance / glm_tbl$null.deviance)
# --> 0.0008184671
pseudo_r2_fwl <- 1 - (glm_fwl$deviance / glm_fwl$null.deviance)
# --> 0.06680286
pseudo_r2_hwl_sp <- 1 - (glm_hwl_sp$deviance / glm_hwl_sp$null.deviance)
# --> 0.05604784


# Plots: ####

# tbl
ggplot(all_data, aes(x = tbl, y = cop)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "glm", method.args = list(family = "binomial"),
se = FALSE) +
  geom_abline(intercept = coef_tbl["(Intercept)", "Estimate"], slope =
0, linetype = "dashed", color = "black") +
  labs(x = "Total Body Length (mm)", y = "Predicted Copulation Success")
+
  theme_classic()

# fwl
ggplot(all_data, aes(x = fwl, y = cop)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "glm", method.args = list(family = "binomial"),
se = FALSE) +
  geom_abline(intercept = coef_fwl["(Intercept)", "Estimate"], slope =
0, linetype = "dashed", color = "black") +
  labs(x = "Forewing Length (mm)", y = "Predicted Copulation Success") +
  theme_classic()

# hwl
ggplot(all_data, aes(x = hwl, y = cop, color = sp)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "glm", method.args = list(family = "binomial"),
se = FALSE, aes(color = sp)) +
  scale_color_manual(values = c("CS" = "lightblue", "CV" = "blue")) +
  geom_abline(intercept = coef_hwl_sp["(Intercept)", "Estimate"], slope
= 0, linetype = "dashed", color = "black") +
  labs(x = "Hindwing Length (mm)", y = "Predicted Copulation Success") +
  guides(color = guide_legend(title = NULL)) +
  theme_classic()
```