**Title: Morphological adaptation of *Tephritis conura* in two stable host plants**

**Background**

This study investigates morphological differentiation among *Tephritis conura* (*T.conura*) between two different host plants. *T. conura* flies utilize and infest two different host plants. These two host plants are *Cirsium heterophyllum (C. heterophyllum) and Cirsium oleraceum (C. oleraceum*). Individuals of both host races were collected from eight different populations in northern Europe. Individuals from both sexes were obtained from each bud of both host plants. Host races from either individuals *C. heterophyllum* or *C. oleraceum or* *C.heterophyllum×C. oleraceum* hybrids were used to assess for their morphological differences. Morphological traits including body length, ovipositor length, wing length were investigated between two host plants. The ovipositor is a key trait and is the focus in this study.

**Methods**

For quantification of host-related morphological differences, 583 *T. conura* flies from both C.*heterophyllum* populations and *C.oleraceum* populations were studied. All subsequent statistical analysis were performed in the statistical software R. Three morphological traits (body length (BL), ovipositor length (OL), wing length) were quantified for 286 *T. conura* flies from *C.oleraceum and 297 T. conura* flies from C.*heterophyllum.* Two predictor variables (BL and wing length) were used in the analyses for OL of each host plants. In all model analysis, missing values (male) were deleted and disregard in the analysis. Multiple linear regression models were used to investigate the additive effects of two parameters (BL and wing length) against OL in each hostplant. Individual graphs were plotted using simple linear regression.

**Result**

Here is the combination of different models for *C.oleraceum* to investigate the morphological differences.

model\_1 = lm(OL ~ BL + Wing\_length, data = Oleraceum)

model\_2 = lm(OL ~ BL \* Wing\_length, data = Oleraceum)

model\_3 = lm(OL ~ BL, data = Oleraceum)

model\_4 = lm(OL ~ Wing\_length, data = Oleraceum)

Table 1. Model selection for variables OL, BL and wing length of *C.oleraceum*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | df | AIC | logLik | delta | w |
| model\_3 | 3 | -207.3541 | 106.6770 | 0.00 | 0.56 |
| model\_1 | 4 | -205.9213 | 106.9606 | 1.43 | 0.27 |
| model\_2 | 5 | -204.9134 | 107.4567 | 2.44 | 0.17 |
| model\_4 | 3 | -158.2546 | 82.1273 | 49.10 | 0.00 |

Model 1 was chosen among other models as it consisted of acceptable AIC and weight values (Table 1). Residuals from model 1 of *C.oleraceum* were normally distributed (supplementary Figure 1).

The linear regression analysis using the lm() function in R, lm(formula = OL ~ BL + Wing\_length, data = Oleraceum), to model the relationship between the response variable OL and predictor variables BL and Wing length using data from the Oleraceum dataset.

The estimated coefficients for the intercept is 0.548 and the predictor variables (BL (0.279) and Wing length(-0.030). Residual standard error estimates of the standard deviation of the residuals. This model indicates the average distance of data points from the fitted regression line is 0.1139. Multiple R-squared estimates the proportion of variance in the dependent variable (OL) explained by the independent variables (BL and Wing length). This model indicates that about 39.33% of the variance in OL is explained by BL and Wing length. This output suggests that while BL is strongly associated with OL, Wing length doesn't seem to have a significant relationship with OL in the context of this model.

Here is the combination of different models for *C.heterophyllum*to investigate the morphological differences.

model\_1 = lm(OL ~ BL + Wing\_length, data = Heterophyllum)

model\_2 = lm(OL ~ BL \* Wing\_length, data = Heterophyllum)

model\_3 = lm(OL ~ BL, data = Heterophyllum)

model\_4 = lm(OL ~ Wing\_length, data = Heterophyllum)

Table 2. Model selection for variables OL, BL and wing length of *C.oleraceum*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | df | AIC | logLik | delta | w |
| model\_1 | 4 | -182.7811 | 95.39053 | 0.00 | 0.45 |
| model\_2 | 5 | -182.5636 | 96.28180 | 0.22 | 0.40 |
| model\_3 | 3 | -180.2123 | 93.10613 | 2.57 | 0.12 |
| model\_4 | 3 | -176.9217 | 91.46084 | 5.86 | 0.02 |

Model 1 was chosen among other models as it consisted of acceptable AIC and weight values (Table 2). Residuals from model 1 of *C.heterophyllum* were normally distributed (supplementary Figure 2).

The linear regression analysis using the lm() function in R, lm(formula = OL ~ BL + Wing\_length, data = Heterophyllum), to model the relationship between the response variable OL and predictor variables BL and Wing length using data from the Heterophyllum dataset. The estimated coefficients for the intercept is 0.681 and the predictor variables (BL (0.131) and Wing length (0.098). Residual standard error estimates of the standard deviation of the residuals. This model indicates the average distance of data points from the fitted regression line is 0.1266. This output suggests that BL and Wing length have a statistically significant association with OL, with the model explaining about 26.01% of the variance in OL.

The total variance (0.1139, Oleraceum; 0.1266, Heterophyllum) were explained by the model. The variance explained by each of the predictors BL and wing length was computed using the predicted values associated only with BL, while wing length was kept constant at its mean, and vice versa for the variance associated with BL.

The linear regression analysis using the z-transformation (standardize predictor variables is scaled to zero mean and unit variance), lm(formula = Oleraceum$OL ~ Oleraceum$BL\_z + Oleraceum$Wing\_length\_z) was used.

The model fit (the r2) has not changed, but the parameter estimates have. The intercept (1.64466) can now be interpreted as the mean of OL, because it represents the value of OL when both predictors have a value of 0 (i.e. their mean after the z-transform). This effect can be obtained also by mean-centering the variables without scaling them to a standard deviation of 1. The slopes now have units of standard deviations, this shows directly that the predictor BL (slope = 0.09301) explains more variance in OL than does wing length ( -0.00957).

The same linear regression analysis using the z-transformation (standardize predictor variables is scaled to zero mean and unit variance), lm(formula = Heterophyllum$OL ~ Heterophyllum$BL\_z + Heterophyllum$Wing\_length\_z) was used. The intercept for Heterophyllum is 1.74748. The slope for predictor BL (0.04141), and wing length (0.03316) explains the variance in OL.

**A graph of blue and black dots

Description automatically generated**

Figure 1. Dependent OL (mm) is plotted against BL (mm) and fitted linear regression line (solid: *C. oleraceum* (light blue circle); dotted: *C.heterophyllum* (blue circle).

A linear regression model is fitted using the lm() function in R to model the relationship between the outcome variable OL and the predictor variable BL using data from the *C. oleraceum* dataset or the *C. heterophyllum.*  The fitted line is plotted over the scatterplot. The intercept and slope for *C. oleraceum* (0.48086, 0.26208±0.02785 mm/mm), while the intercept and slope for *C. heterophyllum* (0.82363, 0.20672 ± 0.03107 mm/mm). The small standard error (relative to the slope estimate) directly indicates the strong statistical support in both hostplants (Figure 1). The ratio of average OL size over the average BL size is bigger (0.3920825) in *C. heterophyllum* as compared to *C. oleraceum* (0.3730107).

A graph of different sizes of plants

Description automatically generated with medium confidence

Figure 2. Dependent OL (mm) is plotted against wing length (mm) and fitted linear regression line (solid: *C. oleraceum* (orchid1 circle); dotted: *C.heterophyllum* (slateblue circle).

A linear regression model is fitted using the lm() function in R to model the relationship between the outcome variable OL and the predictor variable wing length using data from the *C. oleraceum* dataset or the *C. heterophyllum.*  The fitted line is plotted over the scatterplot. The intercept and slope for *C. oleraceum* (0.82580, 0.17132 ± 0.03692 mm/mm), while the intercept and slope for *C. heterophyllum* (0.79835, 0.19693 ± 0.03111 mm/mm). The small standard error (relative to the slope estimate) directly indicates the strong statistical support in both hostplants.

This output suggests that while wing length is significantly associated with OL, it explains a relatively small proportion of the variance in OL, as indicated by the modest Multiple R-squared value of 0.135 (*C. oleraceum*) , 0.2189 (*C. heterophyllum*)(Figure 2). The ratio of average OL size over the average wing length is larger (0.3489645) in *C. heterophyllum* as compared to *C. oleraceum* (0.3730107).

**Conclusion**

The ovipositor length differences among *T. conura* host races are consistent in BL and wing length. Ovipositor is longer in host races in *C. heterophyllum*. The additive effects BL and wing length are correlate better in *C. heterophyllum.* Host plant in given published articles indicate host flower-head size in *C. heterophyllum* is larger than *C. oleraceum.* This suggests that host-related differences in ovipositor length may reflect adaptations to the respective host-plant species.

Appendix:

**A graph of a number of individuals

Description automatically generated**

Supplementary figure 1. Histogram of residuals for *C.oleraceum* hostplant

A graph of a number of individuals

Description automatically generated

Supplementary figure 2. Histogram of residuals for *C*.*heterophyllum* hostplant