Coming out from the shadows: facultative slave-making ants reveal their chemical identity during colony development

Behavioral Ecology and Sociobiology

Tomasz Włodarczyk* Thomas Schmitt†

Contents

*Faculty of Biology, University of Białystok, Poland, tomwlo@gmail.com, ORCID 0000-0003-1554-9699

†Department of Animal Ecology and Tropical Biology, University of Würzburg, Germany, thomas.schmitt@uni-wuerzburg.de, ORCID 0000-0002-6719-8635

1 Introduction

This document presents graphical results generated as output from analytic workflows written in *R* language and available in the public GitHub repository - https://github.com/TomVuod/sangchem/tree/supp.

2 Species discrimination and prediction

To classify the CHC samples collected from mixed colonies, we need a measure which positions a sample between the *F. sanguinea* and *F. fusca* CHC profiles. Therefore, using `mixOmics` R package [Le Cao et al., 2025], we performed a discriminant analysis training the model on the chemical data collected from the filed colonies in the earlier study [Włodarczyk and Szczepaniak, 2017]. The model was subsequently used to predict identity of new samples.

2.1 Species markers

To figure out the importance of each input variable in the prediction of the species identity made by discriminant model, we calculated the vector of weights using the following formula [Tenenhaus, 1998, Rohart et al. [2017]].

$$\mathbf{W}(\mathbf{P}^\top\mathbf{W})^{-1}\mathbf{c}$$

$$\mathbf{P} = \mathbf{X}\mathbf{V}$$

$$\mathbf{c} = \mathbf{V}^\top\mathbf{y},$$

where
W is the matrix of loading vectors with the number of rows corresponding to the number of input variables and the number of columns corresponding to the number of the latent components
V is the matrix of the coordinates of each observation on latent components
X is the input data matrix
y is the vector of zero-centered one-hot encoded class labels of the samples for a selected class (here, *F. sanguinea* sample)

To project the weights assigned to principal components onto the original variables one has to reverse data transformation. Technically, this is achieved by multiplying with the pseudoinverse of the rotation matrix produced by PCA. We use pseudoinverse because the rotation matrix has been reduced by removing principal components with low variance.

$$\mathbf{R}^+\mathbf{W}(\mathbf{P}^\top\mathbf{W})^{-1}\mathbf{c},$$

where $\mathbf{R}^+$ denotes the pseudoinverse of rotation matrix from PCA.

Table S1 : List of the peaks with the species identity score indicating their importance as a marker of *F. fusca* or *F. sanguinea* samples.

Peak ID	Compound	Marker score	fusca marker	sanguinea marker
15	4-MeC24	-0.0255654	TRUE	FALSE
60	13,17-diMeC29	-0.0254509	TRUE	FALSE
10	3-MeC23	-0.0199233	TRUE	FALSE
6	11-Me-C23 + 9-Me-C23	-0.0187906	TRUE	FALSE
32	14-; 10-MeC26 + 3,7,11-triMeC25	-0.0181863	TRUE	FALSE
12	3,13-; 3,11-; 3,9-; 3,7-diMeC23	-0.0176007	TRUE	FALSE
14	6-MeC24	-0.0174188	TRUE	FALSE
7	7-Me-C23	-0.0171787	TRUE	FALSE
13	12-; 11-; 10-; 9-; 8-MeC24	-0.0170539	TRUE	FALSE
69	C31	-0.0170109	TRUE	FALSE
44	7-MeC27	-0.0168715	TRUE	FALSE
37	7,x-; 5,x-; 10,14-diMeC26	-0.0168247	TRUE	FALSE
31	3,13-diMeC25	-0.0164856	TRUE	FALSE
19	4,12-diMeC24	-0.0160717	TRUE	FALSE
11	C24	-0.0151835	TRUE	FALSE
22	13-; 11-; 9-MeC25	-0.0150467	FALSE	FALSE
49	7,11-diMeC27	-0.0145835	FALSE	FALSE
5	C23	-0.0145465	FALSE	FALSE
33	6-MeC26	-0.0145167	FALSE	FALSE
27	7,11-diMeC25 + 3-MeC25	-0.0141605	FALSE	FALSE
55	C29-ene	-0.0120772	FALSE	FALSE

Table S1 : List of the peaks with the species identity score indicating their importance as a marker of *F. fusca* or *F. sanguinea* samples. *(continued)*

Peak ID	Compound	Marker score	fusca marker	sanguinea marker
40	4,12-diMeC26	-0.0117546	FALSE	FALSE
21	4,12,16-triMeC24	-0.0114977	FALSE	FALSE
65	3,7-; 3,9-; 3,11-diMeC29	-0.0111030	FALSE	FALSE
3	C23-ene	-0.0105679	FALSE	FALSE
8	5-Me C23	-0.0099760	FALSE	FALSE
57	4,8,12-triMeC28	-0.0097011	FALSE	FALSE
42	4,8,14-MeC26	-0.0092556	FALSE	FALSE
26	11,15-; 9,13-diMeC25	-0.0074470	FALSE	FALSE
63	C30-ene	-0.0073432	FALSE	FALSE
34	5-MeC26	-0.0072183	FALSE	FALSE
53	14-; 12-; 10-MeC28	-0.0059951	FALSE	FALSE
16	10,14-diMeC24	-0.0050215	FALSE	FALSE
56	C29	-0.0026027	FALSE	FALSE
24	7-MeC25	-0.0024487	FALSE	FALSE
58	15-; 13-; 11-; 9-; 7-MeC29	-0.0017338	FALSE	FALSE
23	9-MeC25	-0.0006181	FALSE	FALSE
20	C25	-0.0003093	FALSE	FALSE
43	13-MeC27	0.0000836	FALSE	FALSE
77	C33-ene	0.0000884	FALSE	FALSE
51	C28 + 3,15-diMeC27	0.0002487	FALSE	FALSE
47	9,17-; 9,15-diMeC27	0.0027655	FALSE	FALSE
59	11,17-diMeC29	0.0032331	FALSE	FALSE
66	14-; 13-; 12-; 11-MeC30	0.0036963	FALSE	FALSE
48	9,13-diMeC27	0.0051897	FALSE	FALSE
28	5,13-diMeC25	0.0054064	FALSE	FALSE
9	9,13-diMeC23	0.0059222	FALSE	FALSE
4	C23-ene	0.0061225	FALSE	FALSE
80	11,21-; 11,19-diMeC33	0.0091702	FALSE	FALSE
36	10,16-; 10,14-diMeC26	0.0094485	FALSE	FALSE
52	3,9-; 3,7-diMeC27	0.0095821	FALSE	FALSE
54	8,12,16-triMeC28	0.0097574	FALSE	FALSE
72	13,19-;13,17-diMeC31	0.0098296	FALSE	FALSE
18	C25-ene	0.0098422	FALSE	FALSE
30	C26	0.0102945	FALSE	FALSE
62	5,9-; 5,11-; 5,13-diMeC29	0.0103835	FALSE	FALSE
75	x-MeC32	0.0107013	FALSE	TRUE
79	x-MeC32	0.0132174	FALSE	TRUE
46	11,15-diMeC27	0.0143520	FALSE	TRUE
74	(di)MeC31, mix	0.0146347	FALSE	TRUE
70	15-; 13-; 11-; 9-MeC31	0.0150594	FALSE	TRUE
45	5-MeC27	0.0154209	FALSE	TRUE
35	4-MeC26	0.0156890	FALSE	TRUE
76	13,17-diMeC32 + x,y-diMeC32	0.0157811	FALSE	TRUE
17	C25-ene	0.0177864	FALSE	TRUE
71	11,19-; 11,17-; 11,15-diMeC31	0.0181046	FALSE	TRUE
41	C27	0.0196727	FALSE	TRUE

Table S1 : List of the peaks with the species identity score indicating their importance as a marker of *F. fusca* or *F. sanguinea* samples. *(continued)*

Peak ID	Compound	Marker score	fusca marker	sanguinea marker
38	C27-ene + 6,12-diMeC26	0.0222393	FALSE	TRUE
67	(di)MeC30 (mix)	0.0296645	FALSE	TRUE
25	5-MeC25	0.0309652	FALSE	TRUE
50	C28-ene	0.0333505	FALSE	TRUE

2.2 Species Identity Score

Ants in the mixed colonies exchanged their CHC to produce a hybrid chemical identity. We developed a method to assign each sample a value measuring similarity to one or the other species. It is based on a discriminant model that predicts the class of new samples using `predict` method from `mixOmics` R package. The input data are processed in the same way as the training set, i.e. central-log transformation is followed by projection into principal component space. In the latter step, the same rotation matrix and scaling vector is applied as those used for the training set are applied. This is achieved by using `predict` method on the object of `prcomp` class.

2.2.1 Predicted species of different categories of ants as a function of the proportion of *F. sanguinea* ants in a colony

The Species Identity Score was computed for *F. sanguinea* and *F. fusca* samples from mixed colonies and used as a response variable in linear mixed models. As a fixed effect we used the proportion of *F. sanguinea* workers among all workers in a colony. Colony ID and sampling occasion were incorporated as random effects.

In the following model specifications 1|random_factor denotes random intercept (separate for each level of the random factor) and 1|random_factor_1:random_factor_2 denotes random intercepts generated from the combinations of two factors. β_0 and ϵ denote intercept and error term, respectively. Included are summary reports, results of the Shapiro-Wilk test of nornalizty of conditional residuals, as well as diagnostic plots and tests generated with the use of `DHARMa` R package[Hartig, 2024]. The response variable often needed to be transformed to meet model assumptions. Random terms with no variance have been dropped.

2.2.1.1 Mature *F. sanguinea*

$$\text{Species\_Identity\_Score} = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon,$$

where Species_Identity_Score denotes the Species Identity Score of the mature *F. sanguinea* workers.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: predicted_species ~ sang_prop + (1 | colony) + (1 | colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: -26
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9734 -0.4164  0.1299  0.4207  1.2731
##
## Random effects:
```

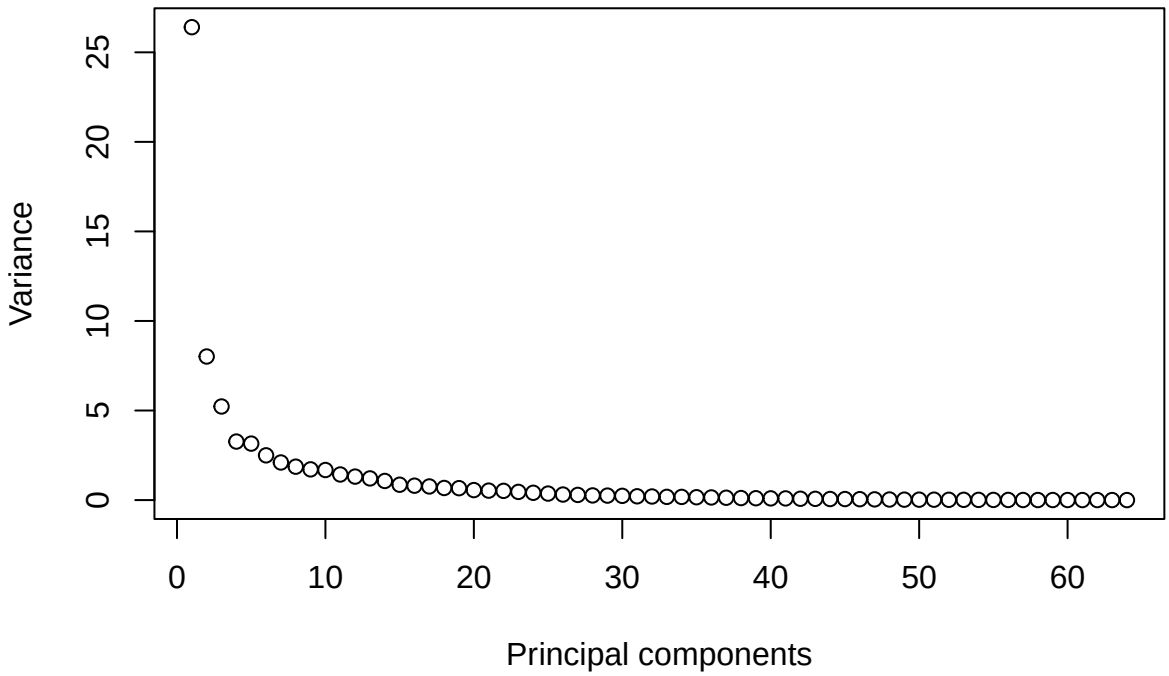


Figure S1 : Variance of each principal compnent. Data was fed into the discriminat model to find features separating both species.

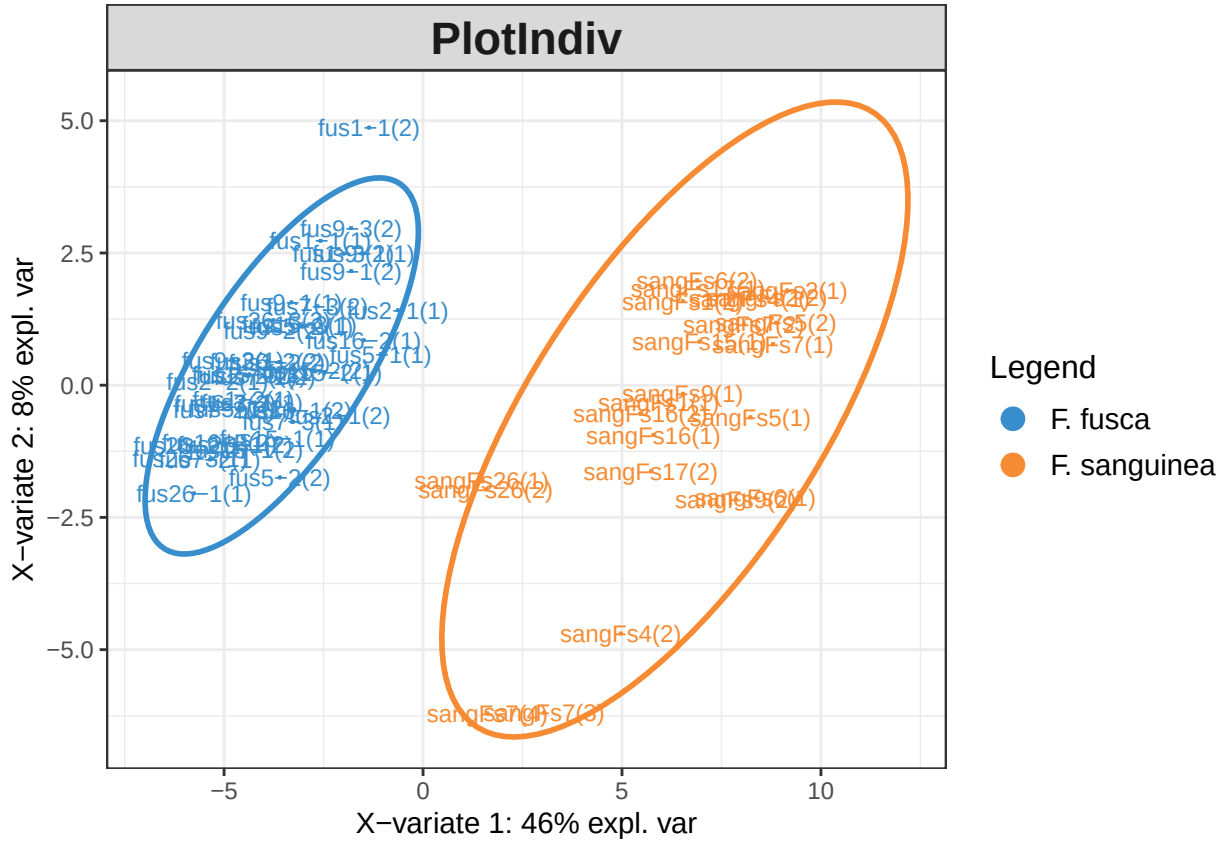


Figure S2 : Distribution of the samples projected onto two first latent components before optimizing the number of model paramaters.

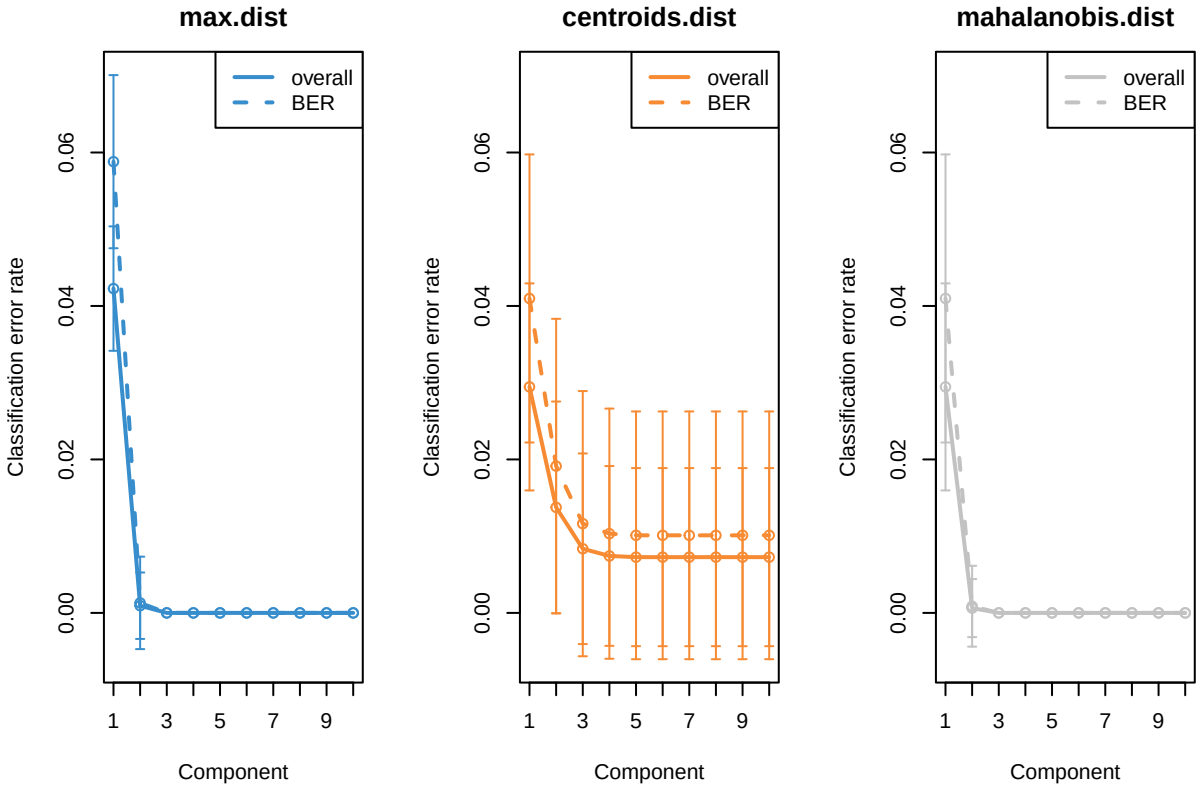


Figure S3 : Results of the performance test of the discriminant model with different number of the latent components. For details see the documunetation of R mixOmics package.

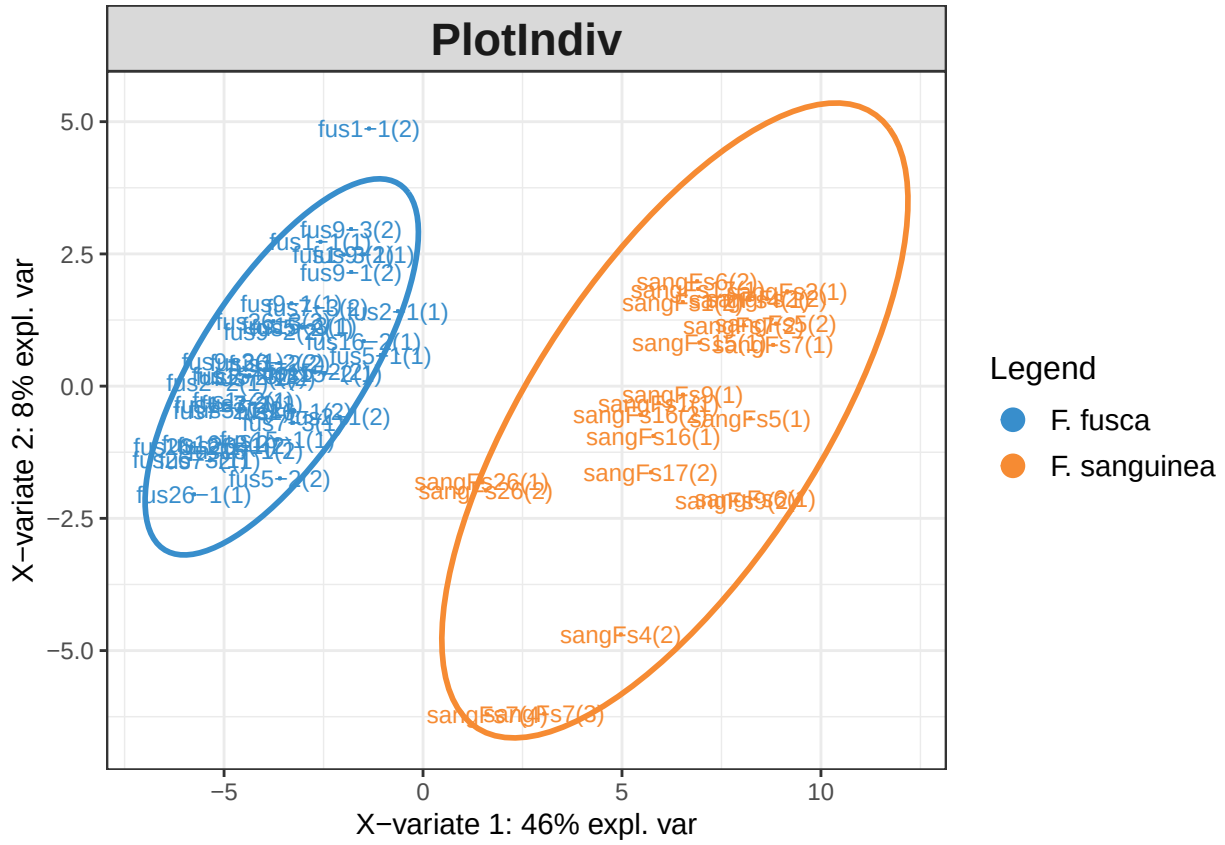
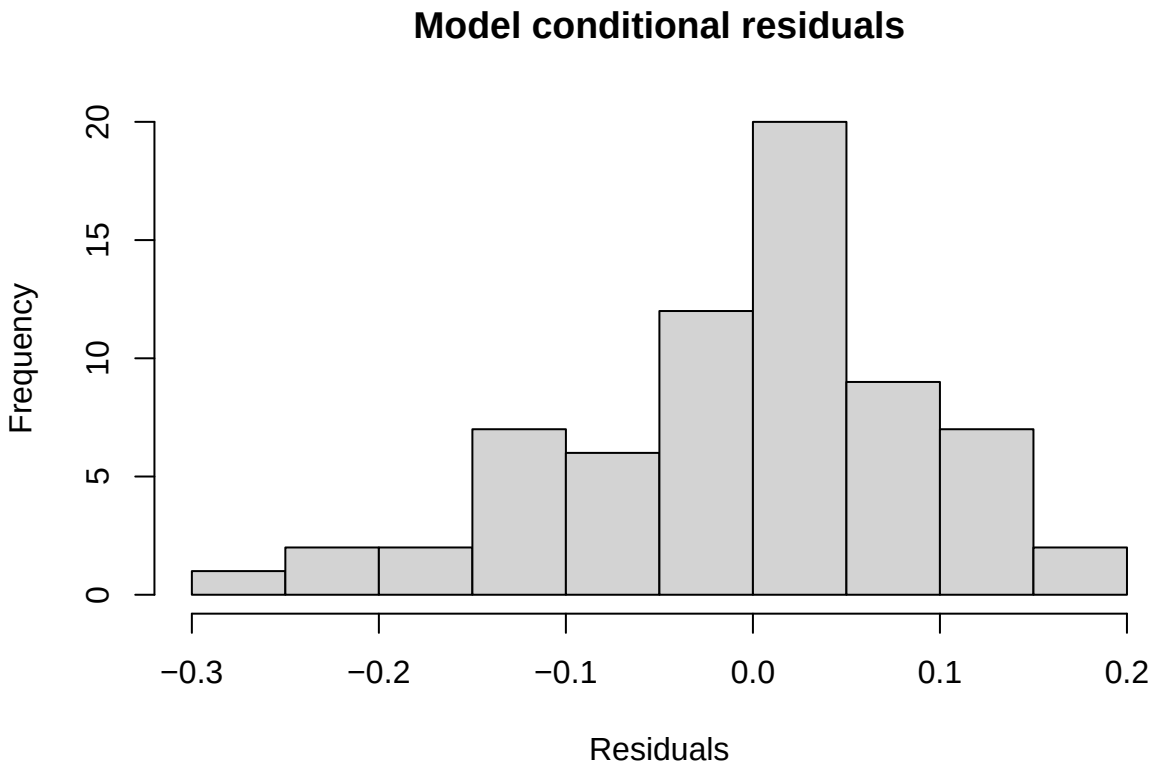
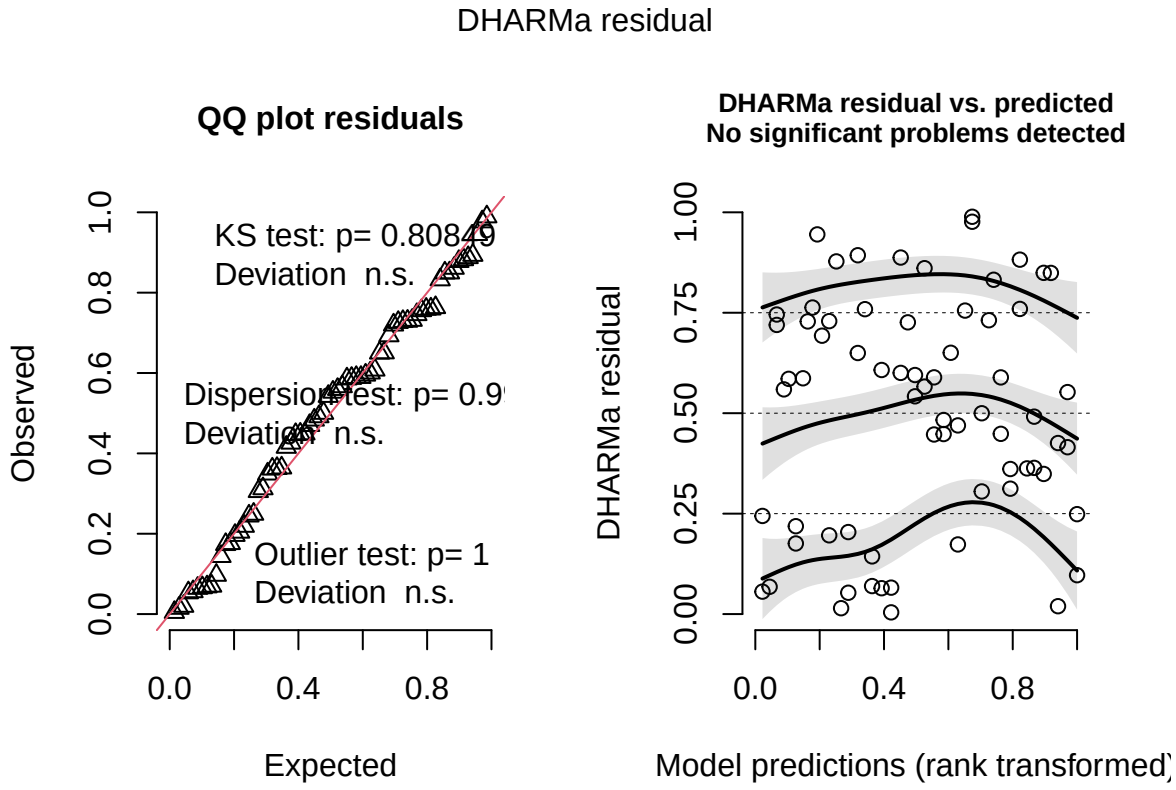


Figure S4 : Projection of the samples in onto two first discriminant analysis components after model tuning.

```
##  Groups             Name        Variance  Std.Dev.
##  colony:census_date (Intercept) 2.737e-02 1.654e-01
##  colony             (Intercept) 5.970e-18 2.443e-09
##  Residual                       1.630e-02 1.277e-01
## Number of obs: 68, groups:  colony:census_date, 44; colony, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  0.34822    0.05221 43.48300   6.670 3.69e-08 ***
## sang_prop    0.97864    0.10145 42.96746   9.646 2.57e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.822
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.96726, p-value = 0.07072
```





2.2.1.2 *F. fusca*

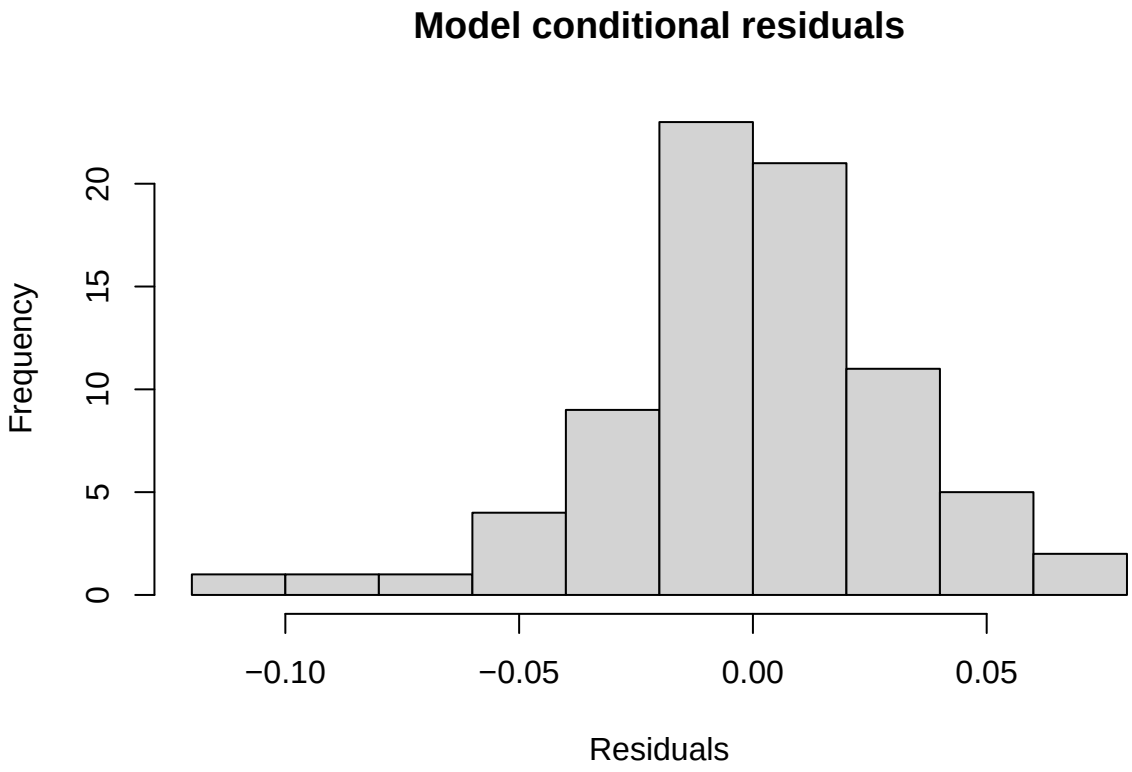
$\sqrt{\text{Species\_Identity\_Score+10}} = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon,$

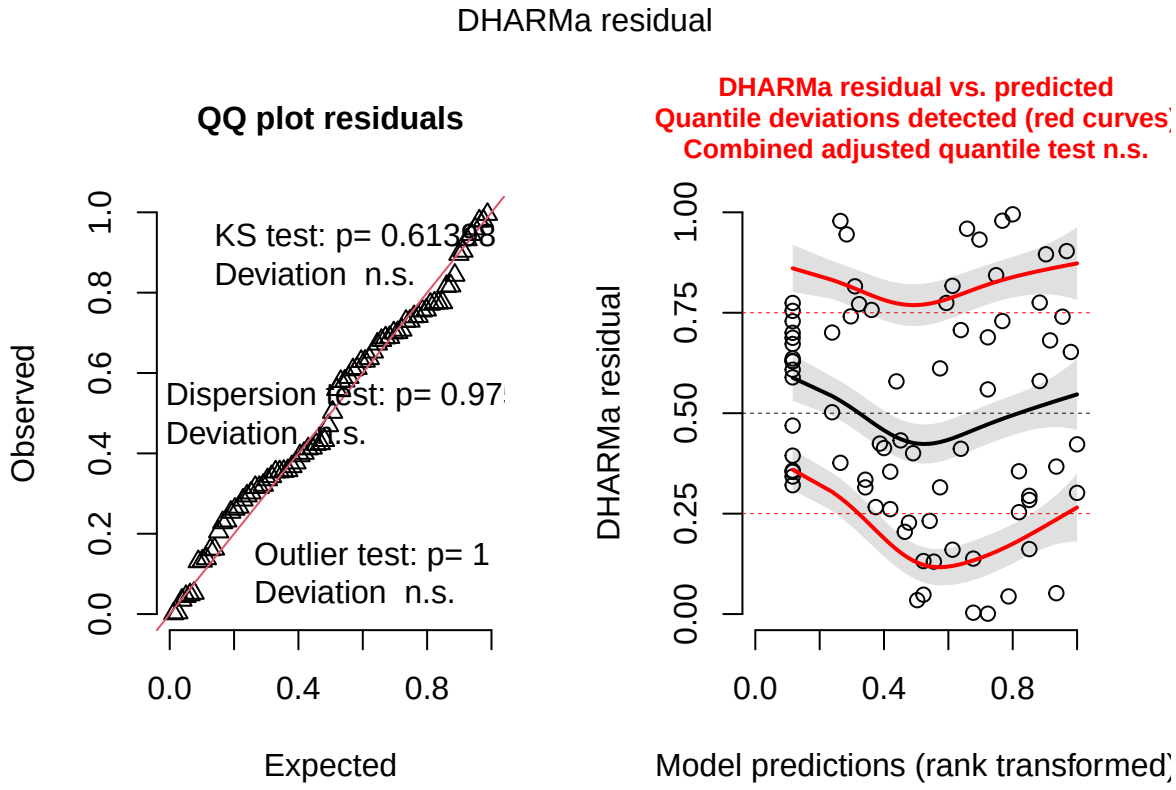
where Species_Identity_Score denotes the Species Identity Score of the mature *F. fusca* workers.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(sqrt(predicted_species + 10)) ~ sang_prop + (1 | colony) +      (1 | colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: -273.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.08306 -0.48706  0.01499  0.52343  2.30571
##
## Random effects:
##  Groups             Name        Variance  Std.Dev.
##  colony:census_date (Intercept) 0.0001468 0.01212
##  colony             (Intercept) 0.0001915 0.01384
##  Residual                       0.0011886 0.03448
## Number of obs: 78, groups:  colony:census_date, 55; colony, 20
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
```

```
## (Intercept)  3.178455   0.007102 41.996873 447.565  < 2e-16 ***
## sang_prop    0.137135   0.015292 35.973490   8.968 1.06e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.652
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.97662, p-value = 0.1611
```





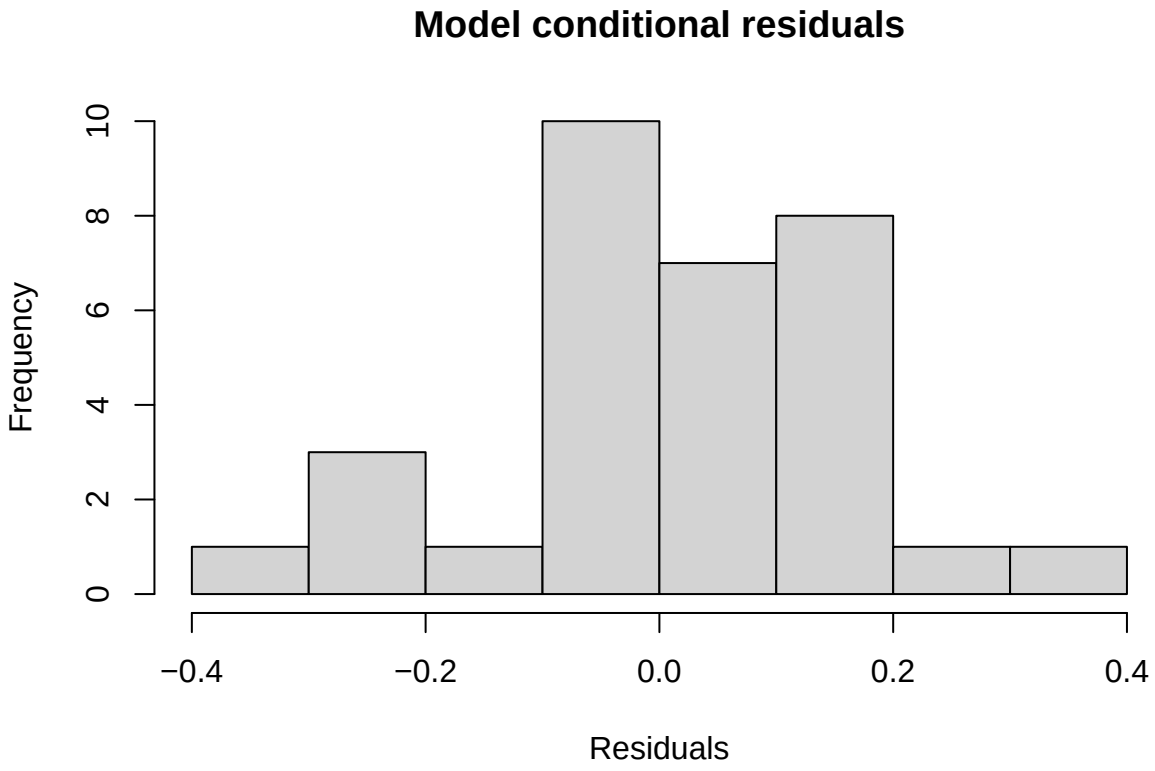
2.2.1.3 Callow *F. sanguinea*

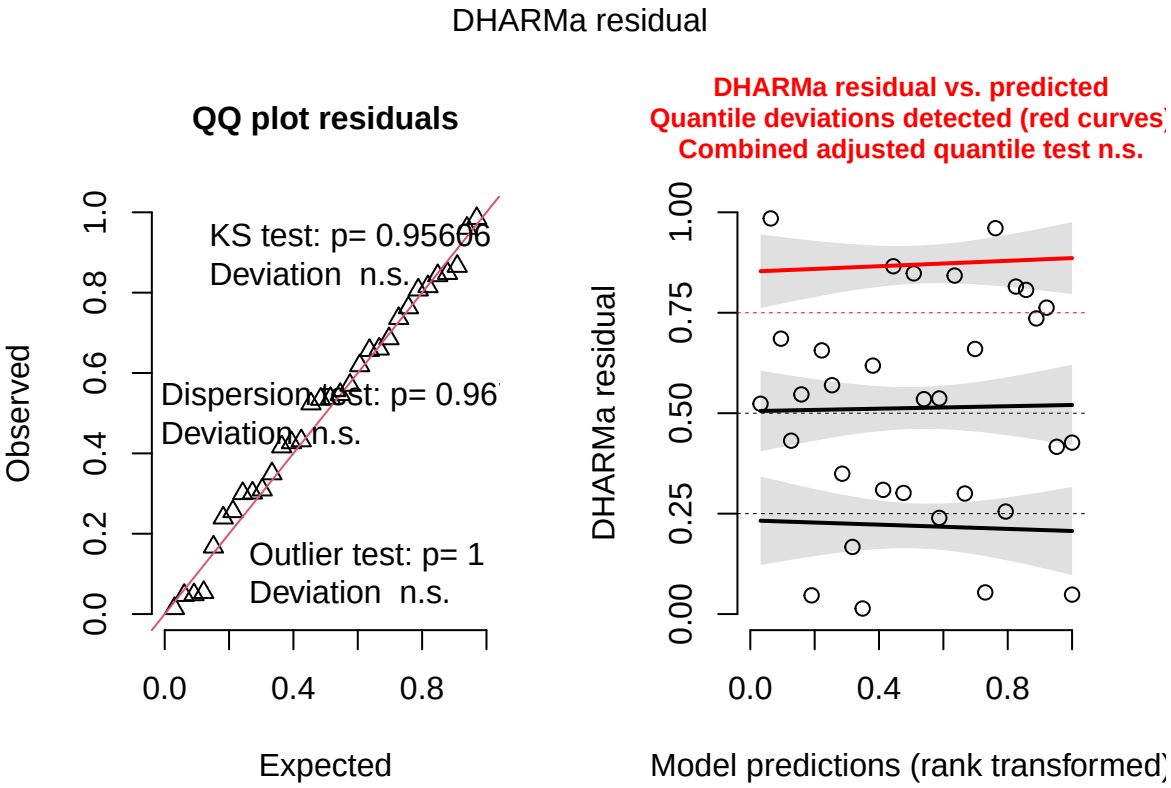
$$\text{Species\_Identity\_Score} = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony}) + \epsilon,$$

where Species_Identity_Score denotes the Species Identity Score of the callow *F. sanguinea* workers.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: predicted_species ~ sang_prop + (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: -15.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.18796 -0.51654  0.04526  0.70017  1.99467
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.003883 0.06231
##  Residual             0.026443 0.16261
## Number of obs: 32, groups:  colony, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  0.16153    0.05533 29.08619   2.919  0.00671 **
```

```
## sang_prop    0.81420    0.10109 28.64191   8.054 7.66e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.798

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98302, p-value = 0.8809
```

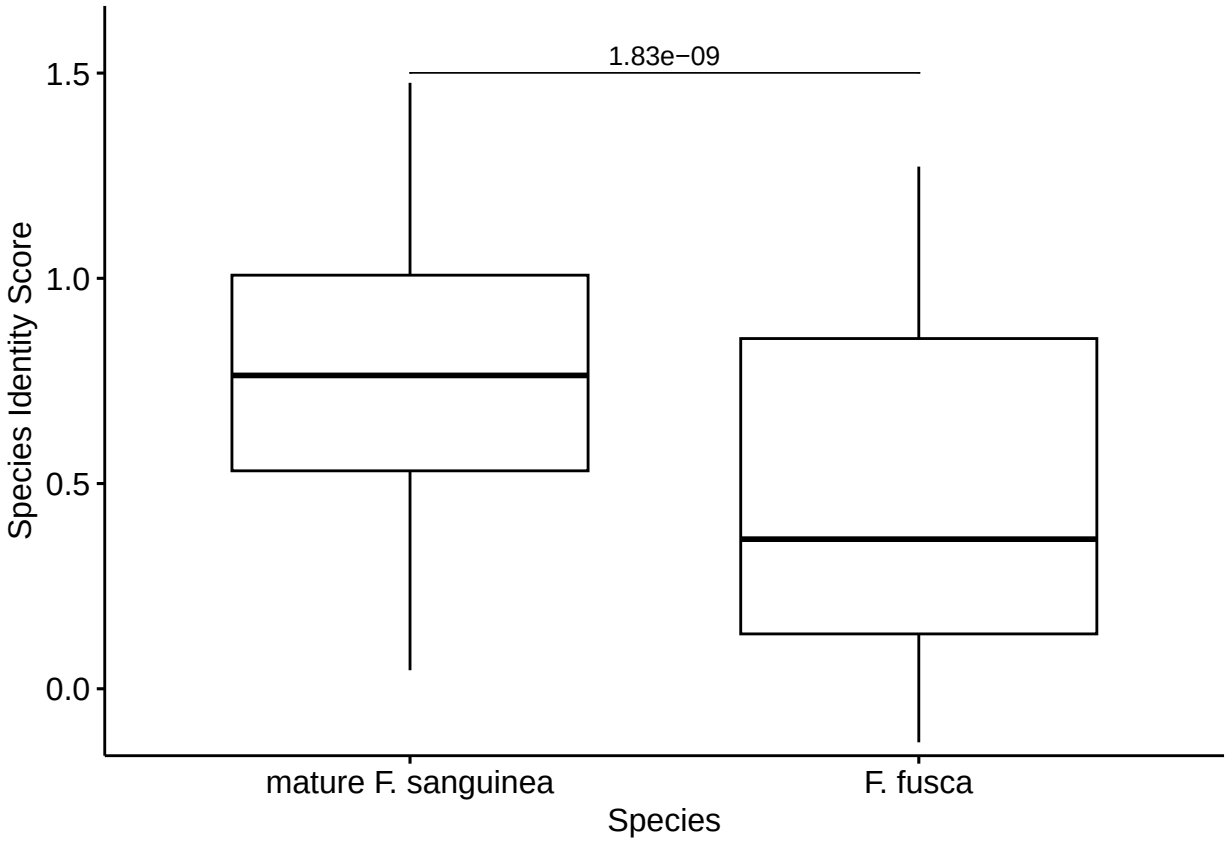




2.2.2 Difference in Species Identity Score between mature *F. sanguinea* ants and *F. fusca* slaves

Since the linear model does not meet one of the diagnostics criteria, Wilcoxon paired test was applied.

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  model_input$predicted_species and model_input$slave_species_index
## V = 1993, p-value = 1.831e-09
## alternative hypothesis: true location shift is not equal to 0
```



2.2.3 Difference in Species Identity Score between callow *F. sanguinea* ants and *F. fusca* slaves

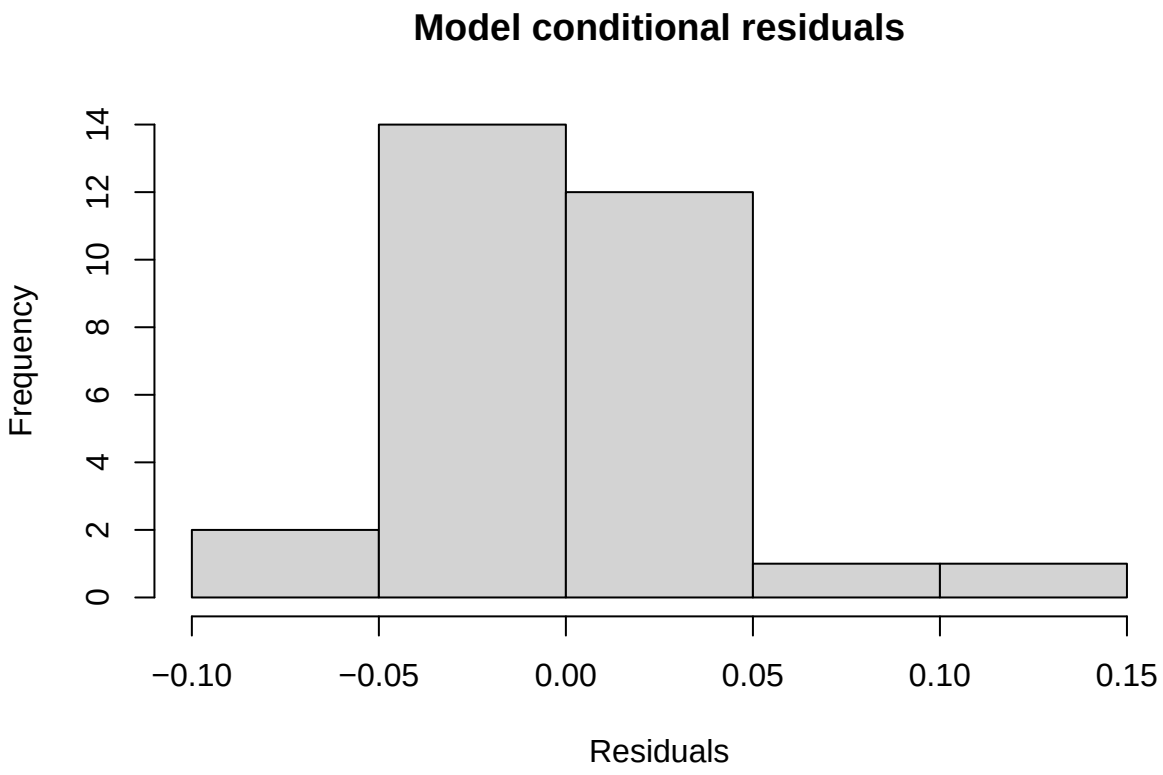
$$\text{SIS\_diff} = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony}) + \epsilon,$$

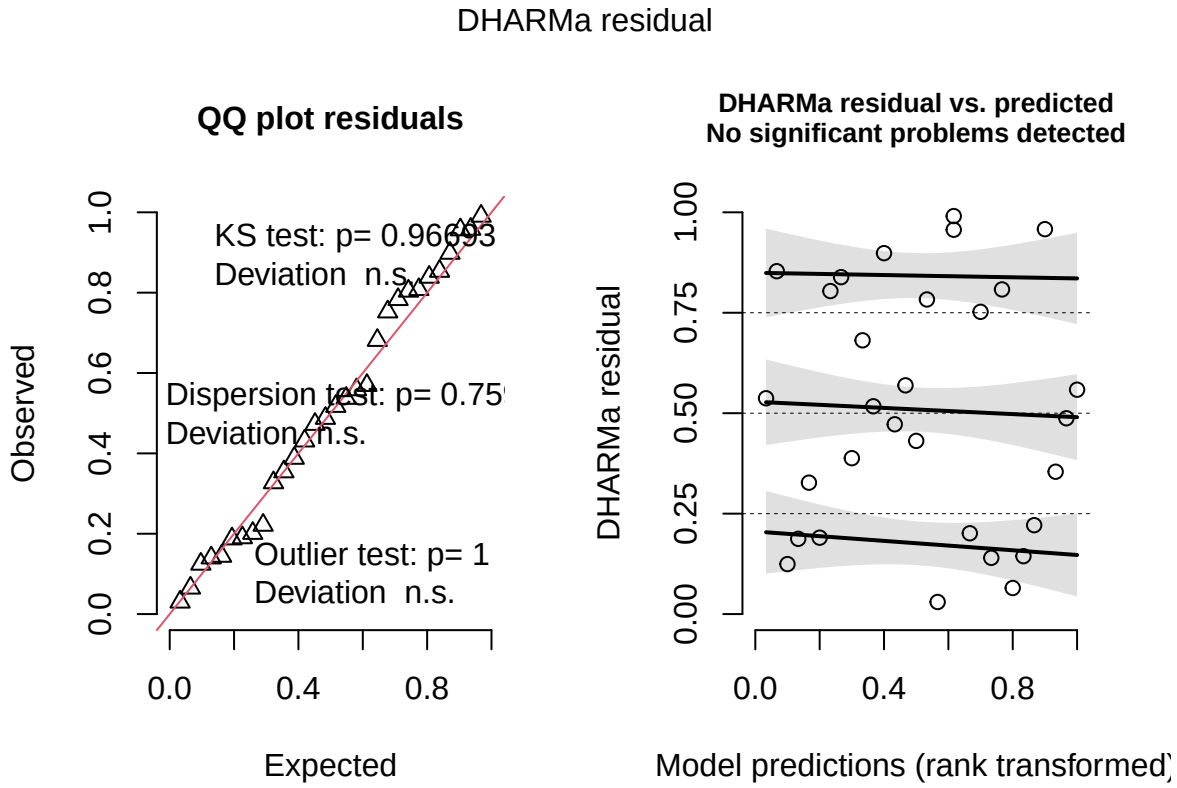
where SIS_diff denotes the difference in Species Identity Score between the mature *F. sanguinea* and *F. fusca* sampled at the same time and from the same colony. Replicated samples were averaged before the calculations.

Since the linear model does not meet one of diagnostics criterion we will apply Wilcoxon paired test.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: index_diff ~ sang_prop + (1 | colony) + (1 | colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: -3.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.73565 -0.34594 -0.01578  0.30506  1.25988
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.039383 0.19845
##  colony             (Intercept) 0.000000 0.00000
##  Residual                       0.007635 0.08738
```

```
## Number of obs: 30, groups:  colony:census_date, 29; colony, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) -0.01125    0.07139 27.47953  -0.158    0.876
## sang_prop    0.02462    0.14663 27.50589   0.168    0.868
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.826
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.95855, p-value = 0.2843
```

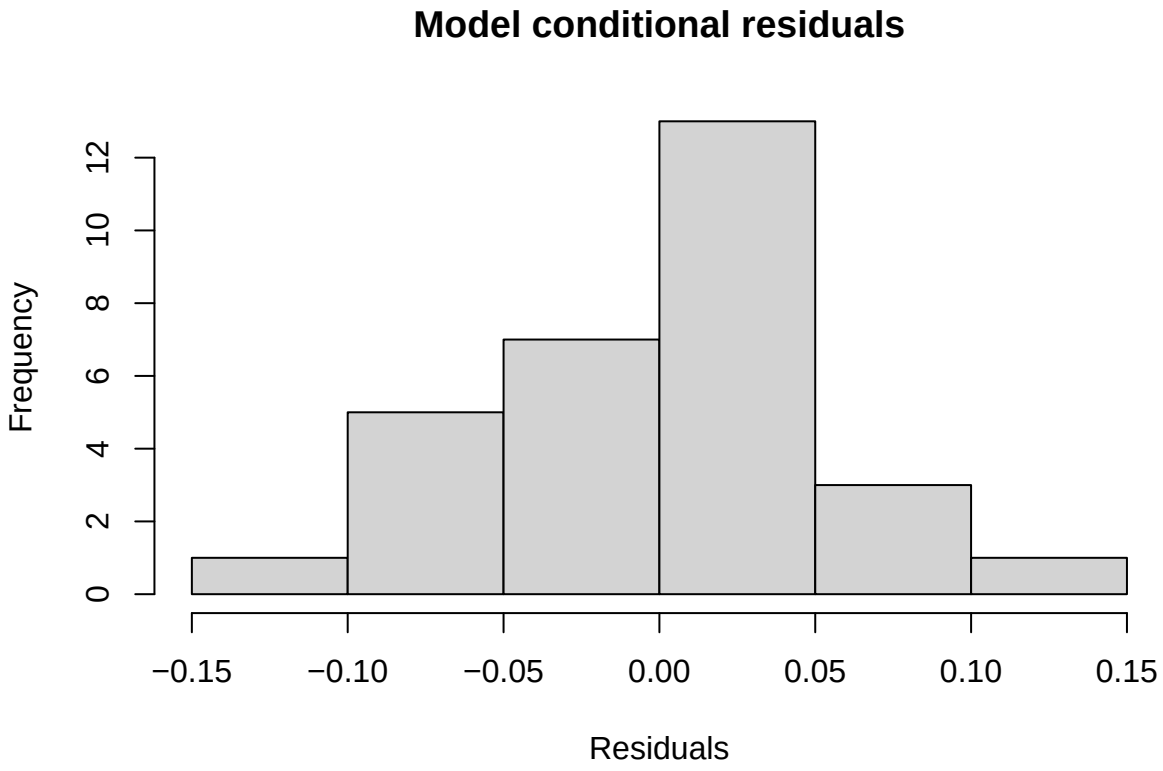


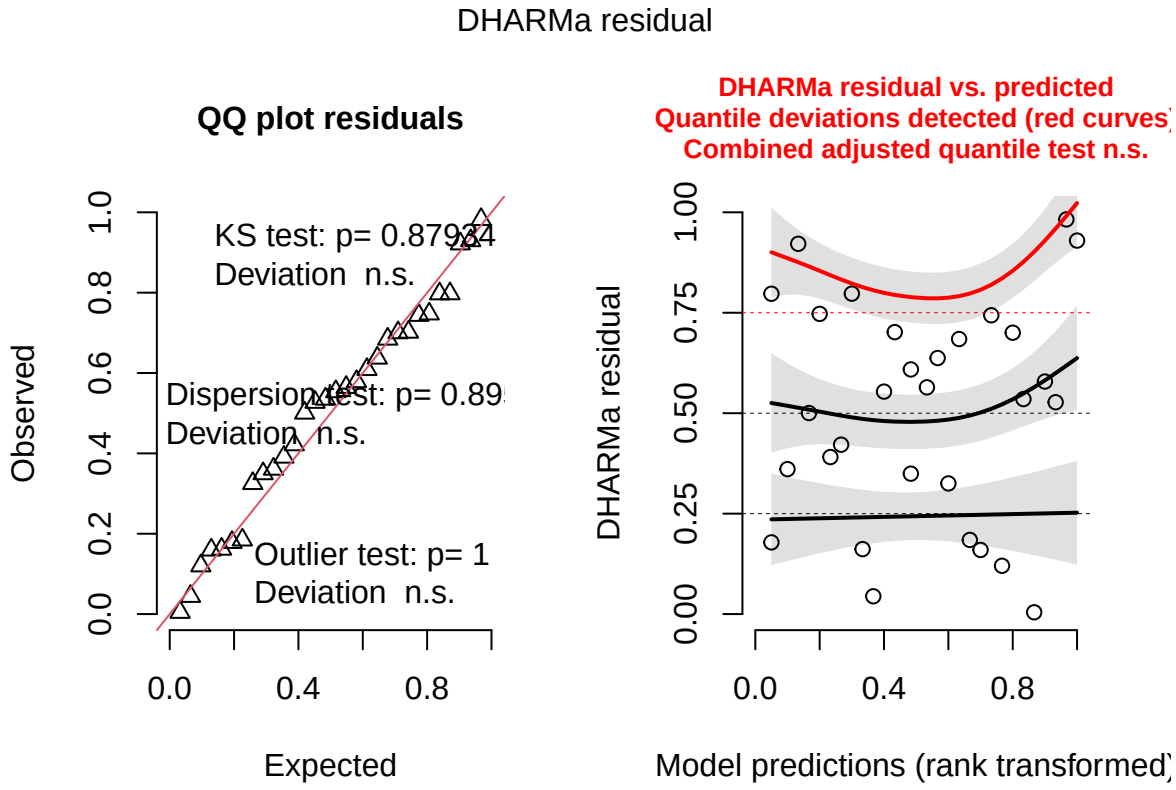


2.2.4 Difference in Species Identity Score between callow and mature *F. sanguinea* ants

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: index_diff ~ sang_prop + (1 | colony) + (1 | colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: -2.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.26576 -0.38931  0.05713  0.30812  1.05875
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.03633  0.1906
##  colony             (Intercept) 0.00000  0.0000
##  Residual                       0.01157  0.1076
## Number of obs: 30, groups:  colony:census_date, 29; colony, 15
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) -0.26885    0.07376 26.79462  -3.645  0.00113 **
## sang_prop   -0.02415    0.13412 26.94731  -0.180  0.85846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.835
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98512, p-value = 0.9393
```





3 Peaks differentiating callow and mature *F. sanguinea* workers

The procedure identifying species markers was also applied to determine peaks characteristic of *F. sanguinea* callow ants. In discriminant analysis, samples were classified into one of the three groups: callow *F. sanguinea*, adult *F. sanguinea*, or adult *F. fusca*.

3.1 Marker indentification

Table S2 : List of the peaks with the score indicating their importance as a marker of mature *F. fusca*, mature *F. sanguinea*, or callow *F. sanguinea* samples.

Peak ID	Compound	Callow score	Callow marker	Mature score	Mature marker
31	3,13-diMeC25	0.0417234	TRUE	-0.0178222	FALSE
24	7-MeC25	0.0386897	TRUE	-0.0327379	FALSE
28	5,13-diMeC25	0.0348832	TRUE	-0.0317316	FALSE
43	13-MeC27	0.0305553	TRUE	-0.0234924	FALSE
32	14-; 10-MeC26 + 3,7,11-triMeC25	0.0296244	TRUE	-0.0373913	FALSE
68	C31ene; 3-MeC30	0.0295610	TRUE	-0.0189614	FALSE
58	15-; 13-; 11-; 9-; 7-MeC29	0.0264429	TRUE	-0.0073115	FALSE
53	14-; 12-; 10-MeC28	0.0244541	TRUE	-0.0172616	FALSE
22	13-; 11-; 9-MeC25	0.0241983	TRUE	-0.0457768	FALSE
13	12-; 11-; 10-; 9-; 8-MeC24	0.0235522	TRUE	-0.0429399	FALSE

Table S2 : List of the peaks with the score indicating their importance as a marker of mature *F. fusca*, mature *F. sanguinea*, or callow *F. sanguinea* samples. *(continued)*

Peak ID	Compound	Callow score	Callow marker	Mature score	Mature marker
10	3-MeC23	0.0233639	TRUE	-0.0332244	FALSE
12	3,13-; 3,11-; 3,9-; 3,7-diMeC23	0.0216728	TRUE	-0.0257661	FALSE
52	3,9-; 3,7-diMeC27	0.0207321	FALSE	-0.0121735	FALSE
27	7,11-diMeC25 + 3-MeC25	0.0199631	FALSE	-0.0195566	FALSE
38	C27-ene + 6,12-diMeC26	0.0184351	FALSE	0.0157453	FALSE
19	4,12-diMeC24	0.0174060	FALSE	-0.0198299	FALSE
49	7,11-diMeC27	0.0161277	FALSE	-0.0162222	FALSE
26	11,15-; 9,13-diMeC25	0.0147779	FALSE	-0.0161658	FALSE
6	11-Me-C23 + 9-Me-C23	0.0145914	FALSE	-0.0286949	FALSE
23	9-MeC25	0.0127681	FALSE	0.0068808	FALSE
7	7-Me-C23	0.0126172	FALSE	-0.0154044	FALSE
47	9,17-; 9,15-diMeC27	0.0124392	FALSE	-0.0148644	FALSE
36	10,16-; 10,14-diMeC26	0.0112164	FALSE	-0.0087982	FALSE
17	C25-ene	0.0111402	FALSE	0.0257789	TRUE
54	8,12,16-triMeC28	0.0073349	FALSE	-0.0072594	FALSE
70	15-; 13-; 11-; 9-MeC31	0.0055168	FALSE	0.0233545	TRUE
3	C23-ene	0.0043691	FALSE	-0.0326940	FALSE
5	C23	0.0040449	FALSE	-0.0385045	FALSE
66	14-; 13-; 12-; 11-MeC30	0.0035414	FALSE	0.0179533	FALSE
45	5-MeC27	0.0006812	FALSE	0.0241601	TRUE
72	13,19-;13,17-diMeC31	0.0002150	FALSE	0.0134093	FALSE
60	13,17-diMeC29	-0.0004368	FALSE	0.0173010	FALSE
61	7,11-; 7,13-; 7,15-diMeC29	-0.0024130	FALSE	0.0053942	FALSE
44	7-MeC27	-0.0043482	FALSE	-0.0216863	FALSE
75	x-MeC32	-0.0048461	FALSE	0.0234901	TRUE
77	C33-ene	-0.0086015	FALSE	0.0217457	FALSE
50	C28-ene	-0.0108417	FALSE	0.0393204	TRUE
59	11,17-diMeC29	-0.0115498	FALSE	0.0183661	FALSE
67	(di)MeC30 (mix)	-0.0137230	FALSE	0.0377407	TRUE
76	13,17-diMeC32 + x,y-diMeC32	-0.0141744	FALSE	0.0414447	TRUE
37	7,x-; 5,x-; 10,14-diMeC26	-0.0143602	FALSE	-0.0206806	FALSE
51	C28 + 3,15-diMeC27	-0.0144155	FALSE	0.0029460	FALSE
79	x-MeC32	-0.0166912	FALSE	0.0323093	TRUE
48	9,13-diMeC27	-0.0187918	FALSE	0.0029403	FALSE
46	11,15-diMeC27	-0.0188050	FALSE	0.0345049	TRUE
11	C24	-0.0194113	FALSE	-0.0263170	FALSE
78	x-MeC32	-0.0200604	FALSE	0.0050884	FALSE
56	C29	-0.0272152	FALSE	-0.0143614	FALSE
64	C30	-0.0275233	FALSE	-0.0042033	FALSE
80	11,21-; 11,19-diMeC33	-0.0300799	FALSE	0.0357983	TRUE
71	11,19-; 11,17-; 11,15-diMeC31	-0.0324295	FALSE	0.0406445	TRUE
20	C25	-0.0341089	FALSE	-0.0184739	FALSE
30	C26	-0.0342660	FALSE	0.0023813	FALSE
25	5-MeC25	-0.0360878	FALSE	0.0417177	TRUE
69	C31	-0.0421958	FALSE	0.0033627	FALSE

Table S2 : List of the peaks with the score indicating their importance as a marker of mature *F. fusca*, mature *F. sanguinea*, or callow *F. sanguinea* samples. *(continued)*

Peak ID	Compound	Callow score	Callow marker	Mature score	Mature marker
41	C27	-0.0464376	FALSE	0.0009429	FALSE

3.2 Verifying the performance of the discrimination model

The performance of the discriminant model was evaluated by examining the accuracy of its predictions in a cross-validation test, measured as the area under the receiver operating characteristic curve (AUROC). This procedure was repeated 10^3 times to account for variance resulting from the random splitting of samples into training and validation sets. In a parallel analysis, the model was trained on data with permuted age status of *F. sanguinea* to assess whether the true assignment of samples influenced model performance. Results from both training approaches were paired, and the difference in AUROC was calculated each for pair. The empirical p-value was defined as the proportion of models in which the AUROC after permutation was equal to or greater than the AUROC based on the original data.

The p-value is calculated as proportion of differences with value equal to or less than zero.

All differences are positive, so the p-value is less than 0.001 since the null distribution consists of 1000 values.

4 Changes in CHC amount over time

This section contains a series of linear (mixed) models fitted to examine how the proportion of *F. sanguinea* workers in a colony is associated with the amount of CHC, analyzed as a whole or in subsets. Included are summary reports, results of the Shapiro-Wilk test of normality of conditional residuals, as well as diagnostic plots and tests generated with the use of `DHARMa` R package[Hartig, 2024]. The response variable often needed to be transformed to meet model assumptions. Random terms with no variance have been dropped.

In model specification (1|random_factor) denotes random intercept (separate for each level of the random factor) and ((1|random_factor_1:random_factor_2)) denotes random intercepts generated from the combinations of two factors. β_0 and ϵ denote intercept and error term, respectively.

4.1 Colony growth and body size

n the analyses below, normalized CHC amounts were used as the response variable. Because body shape in both species deviates from an isometric growth pattern (see Section 9), normalizing CHC amounts by the square of head width may introduce bias. To address this, we first examined the relationship between body size (head width) and the proportion of F. sanguinea workers in a colony, and then incorporated the square of head width as an explanatory variable in the linear model.

$$\text{head\_width} = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + \epsilon,$$

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: head_width ~ sang_prop + (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: -103.8
##
## Scaled residuals:
```

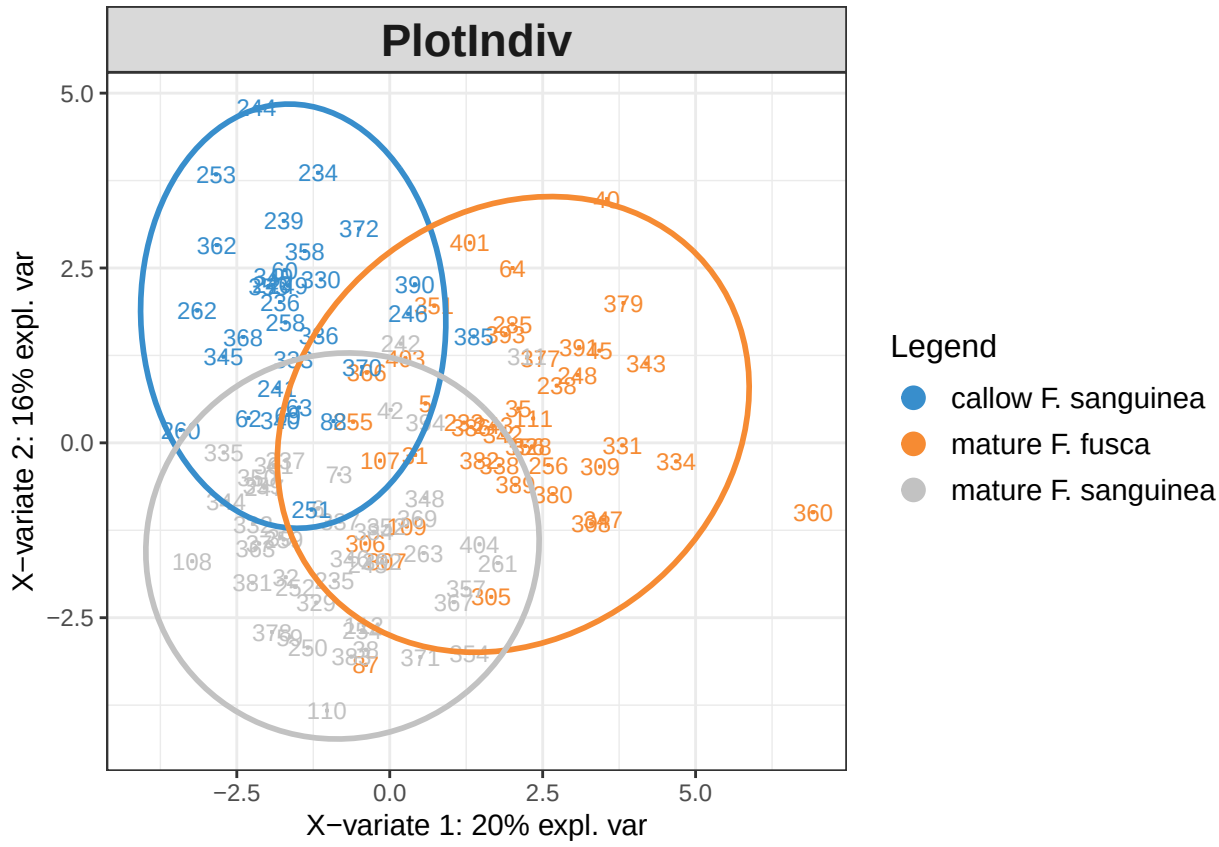


Figure S5 : Distribution of the samples projected onto two first latent components before optimizing the number of model paramaters.

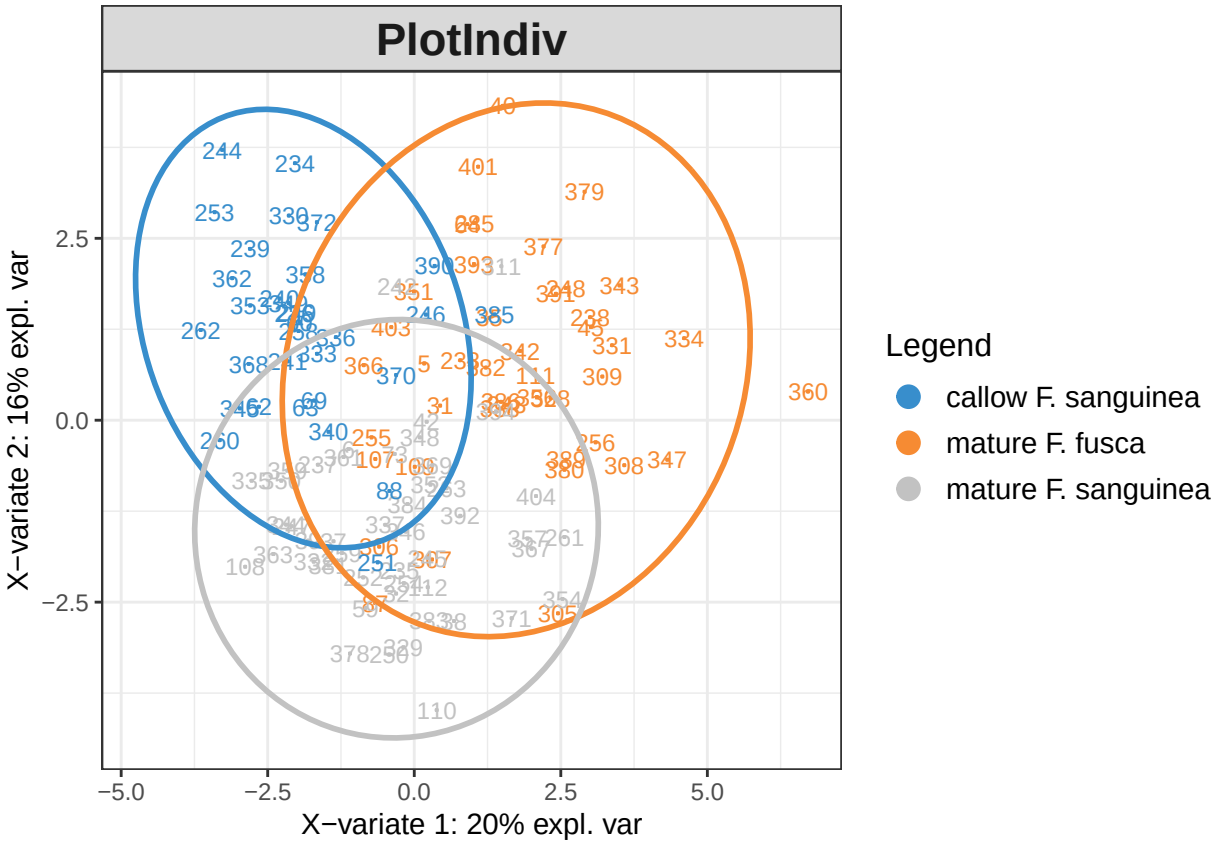


Figure S6 : Projection of the samples in onto two first discriminant analysis components after model tuning.

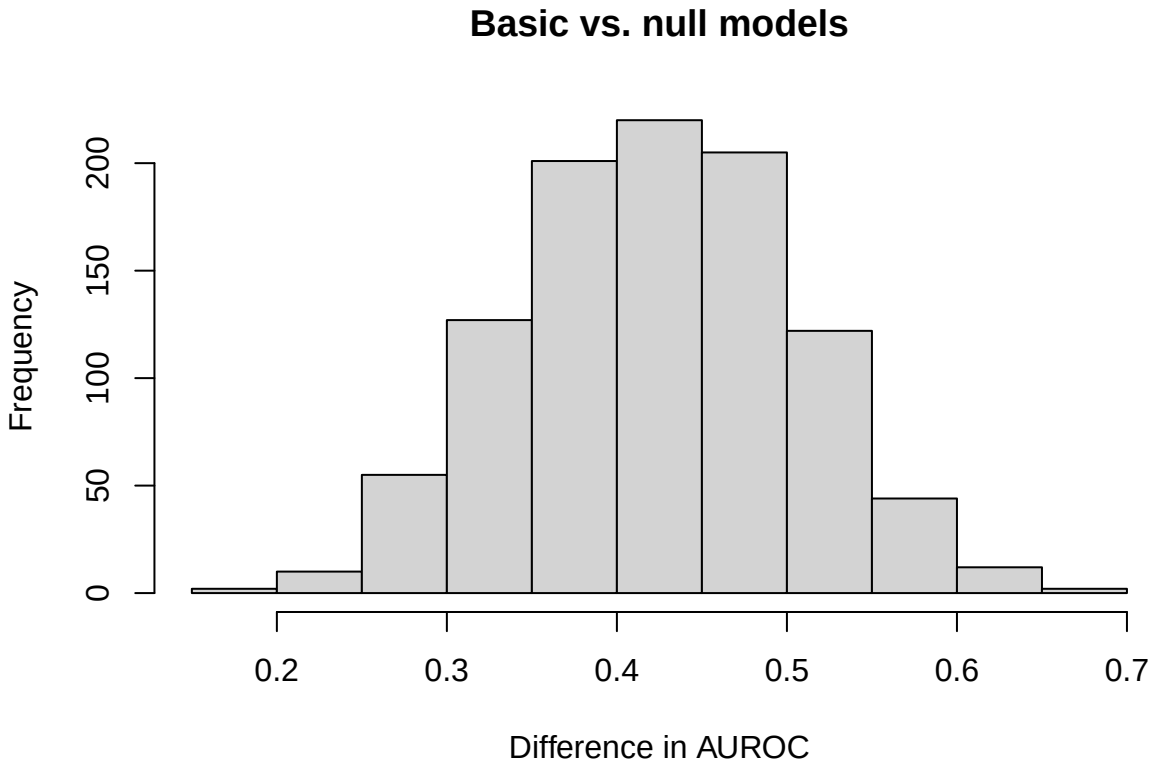
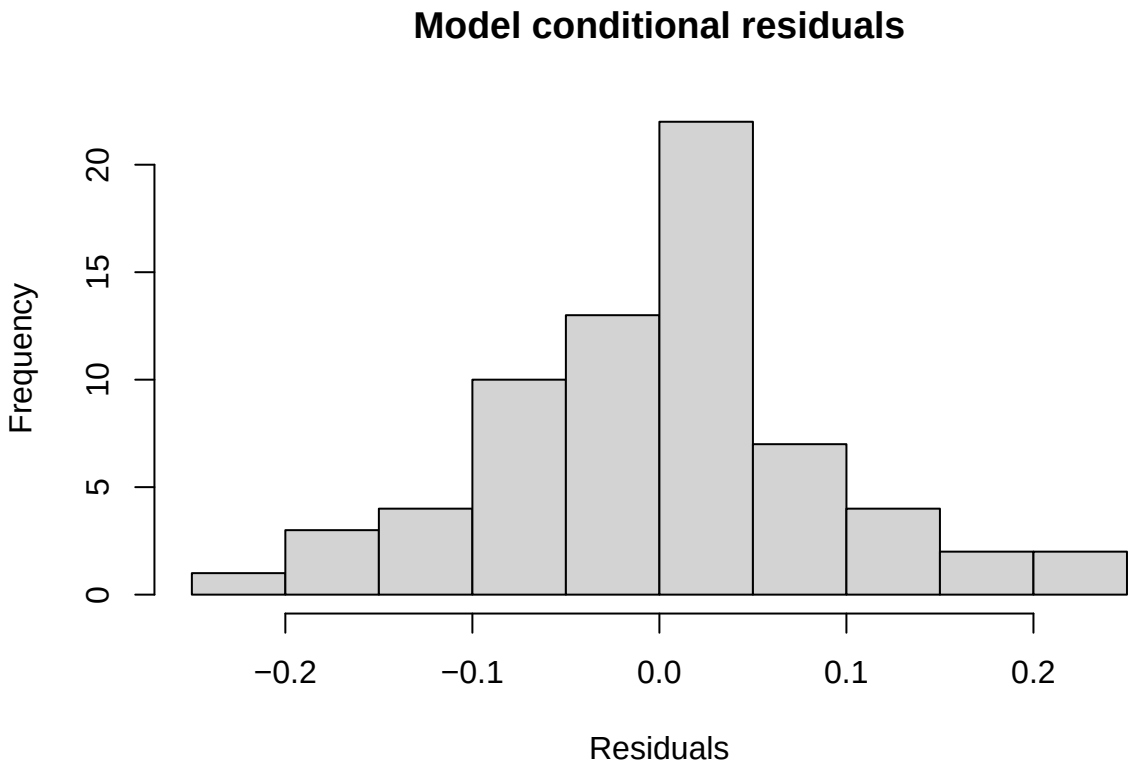
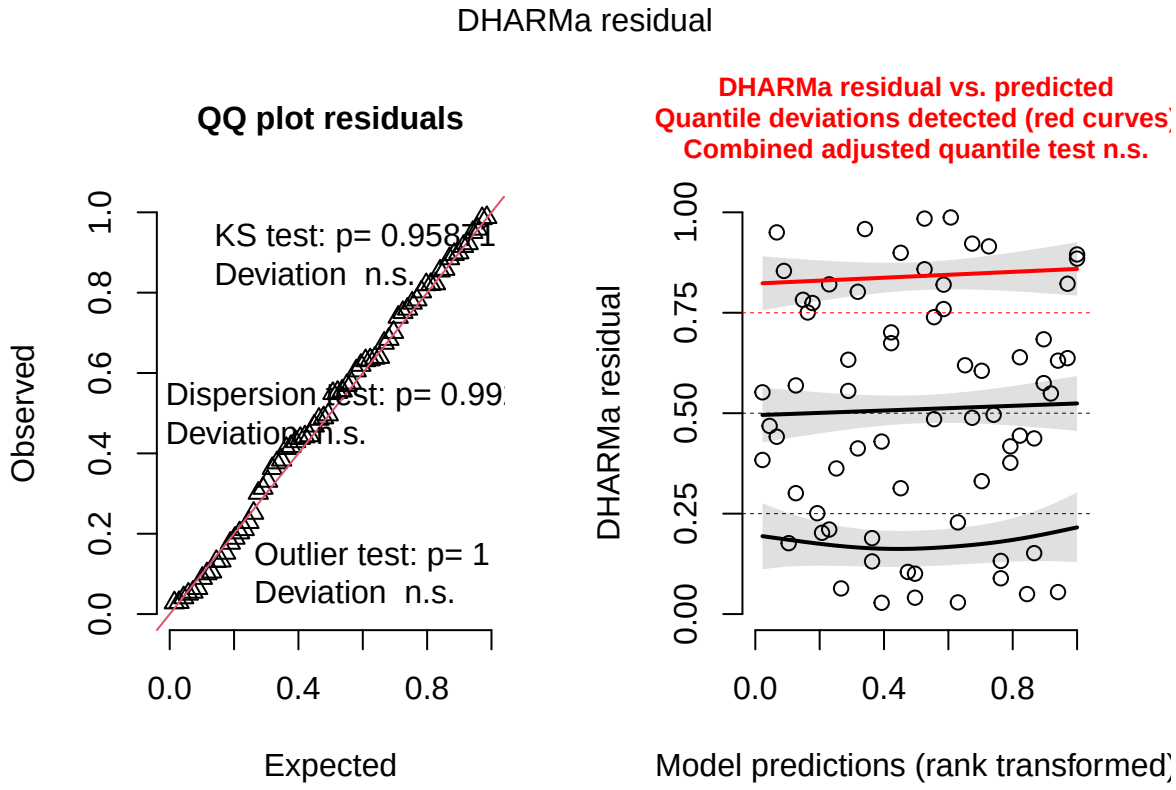


Figure S7 : Distribution of the differences in the performance of the discriminant models trained on true and permuted data. Age status of *F. sanguinea* ants were shuffled in the null variant.

```
##      Min       1Q   Median       3Q      Max
## -2.10138 -0.61076  0.06058  0.43106  2.28980
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.003155 0.05617
##  Residual             0.009110 0.09545
## Number of obs: 68, groups:  colony, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  1.21806    0.02641 38.92702  46.121  < 2e-16 ***
## sang_prop    0.12514    0.04344 62.09975   2.881  0.00544 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.715

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.9861, p-value = 0.6514
```





4.2 Change in total CHC mass

In this series of models the total CHC mass (normalized by assuming head width of 1.3 mm) was regressed on the proportion of *Formica sanguinea* workers in the colony.

4.2.1 Mature *F. sanguinea*

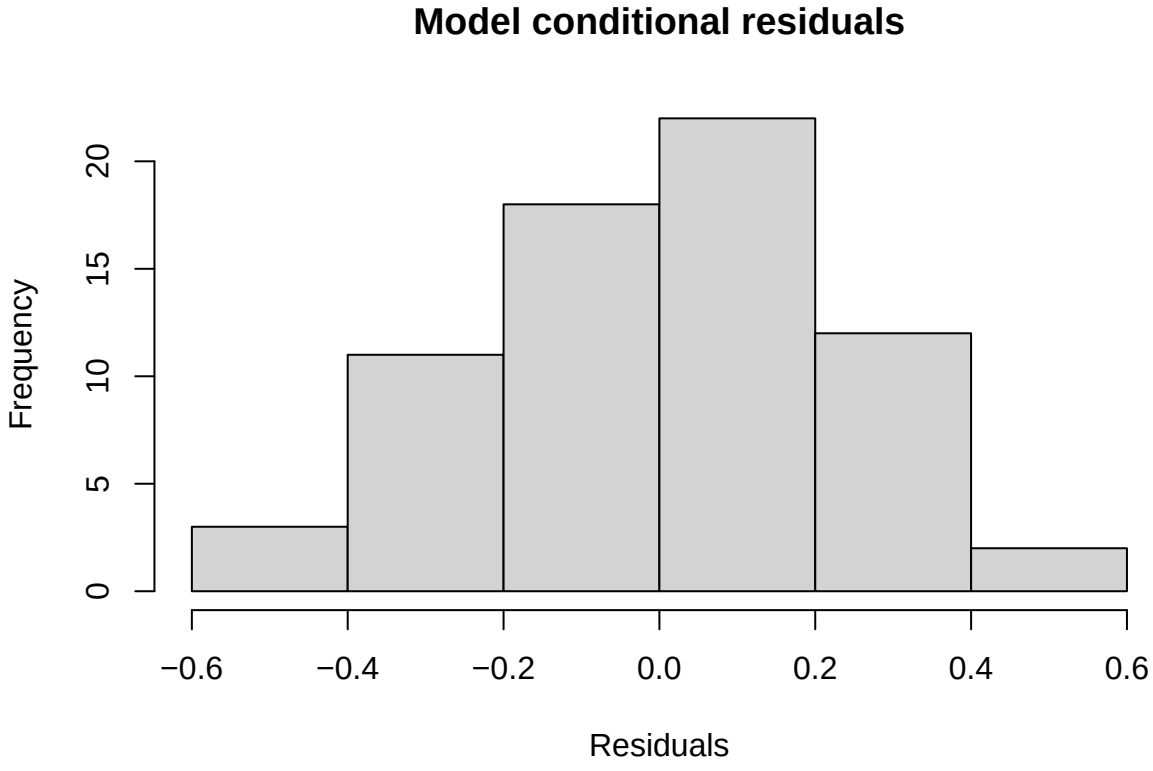
$\log(\text{CHC\_mass\_sanguinea}) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon$,

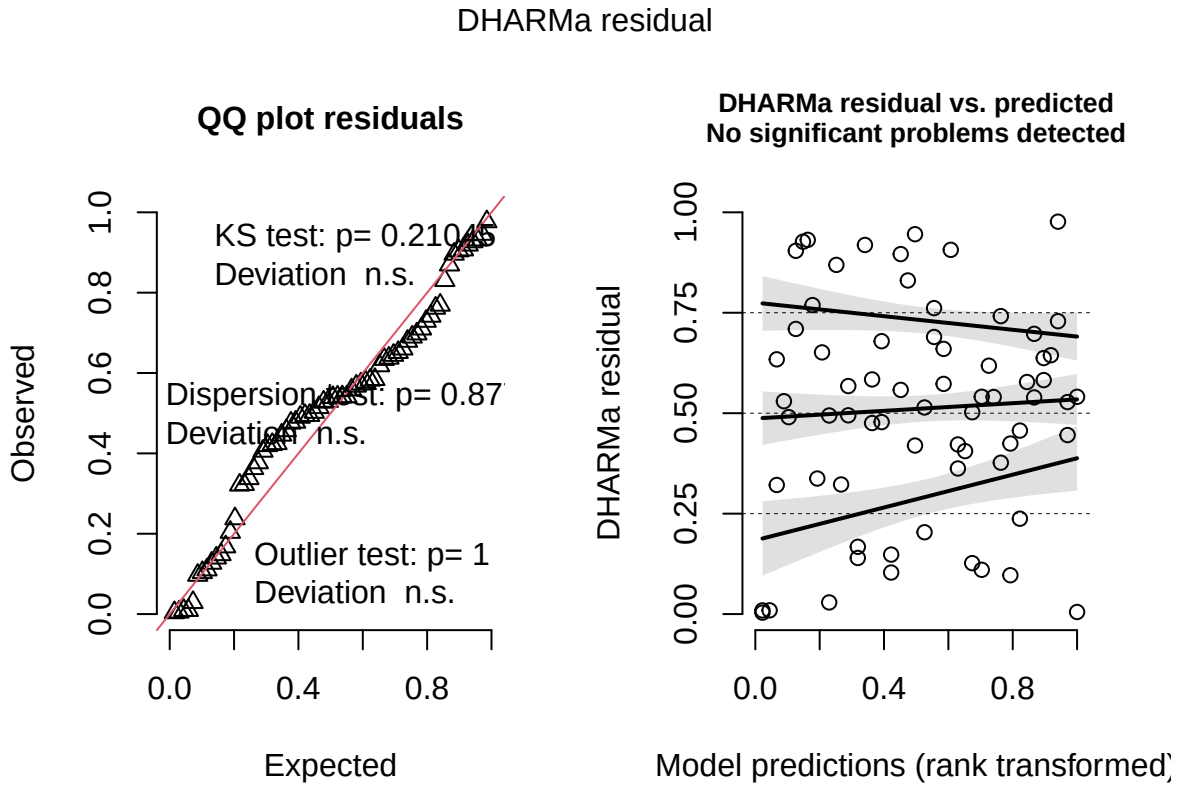
where CHC_mass_sanguinea denotes the total normalized CHC mass on the body of adult *F. sanguinea* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass)) ~ sang_prop + (1 | colony) + (1 | colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: 92.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.83723 -0.39249  0.01222  0.53571  1.73459
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.12975  0.3602
```

```
##  colony             (Intercept) 0.02626  0.1620
##  Residual                       0.10333  0.3215
## Number of obs: 68, groups:  colony:census_date, 44; colony, 16
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   1.2035     0.1295 36.0905   9.296 4.11e-11 ***
## sang_prop     0.7991     0.2392 36.9985   3.340  0.00192 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.786
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98623, p-value = 0.6591
```





The model shows no significant effect of body size on CHC amount in *F. sanguinea*, and this term will therefore be excluded from subsequent models for this species.

4.2.2 *F. fusca*

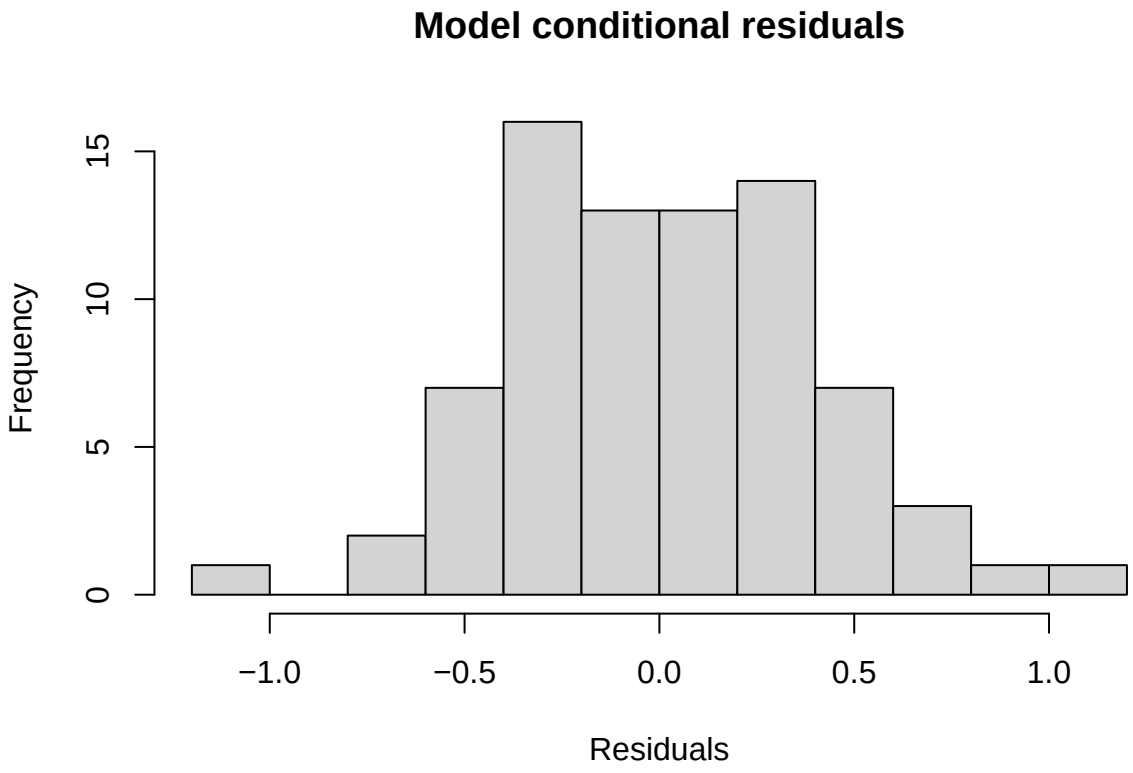
$$\log(\text{CHC\_mass\_fusca}) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + \epsilon,$$

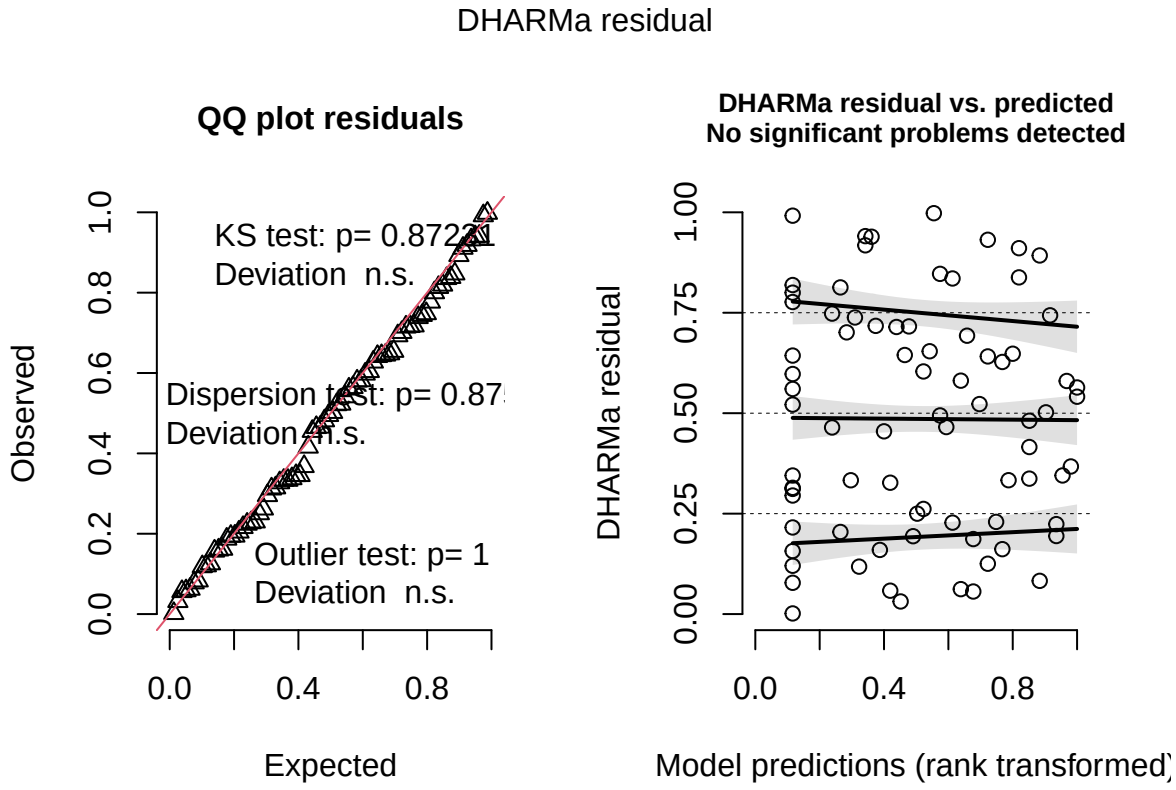
where CHC_mass_fusca denotes the total normalized CHC mass on the body of adult *F. fusca* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass)) ~ sang_prop + (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: 103.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.79685 -0.62513 -0.00685  0.63754  2.49220
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.05014  0.2239
##  Residual             0.17722  0.4210
## Number of obs: 78, groups:  colony, 20
##
```

```
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  0.81635    0.08805 31.60123   9.272 1.57e-10 ***
## sang_prop    0.17928    0.17517 70.00046   1.024     0.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.594
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.99196, p-value = 0.9107
```





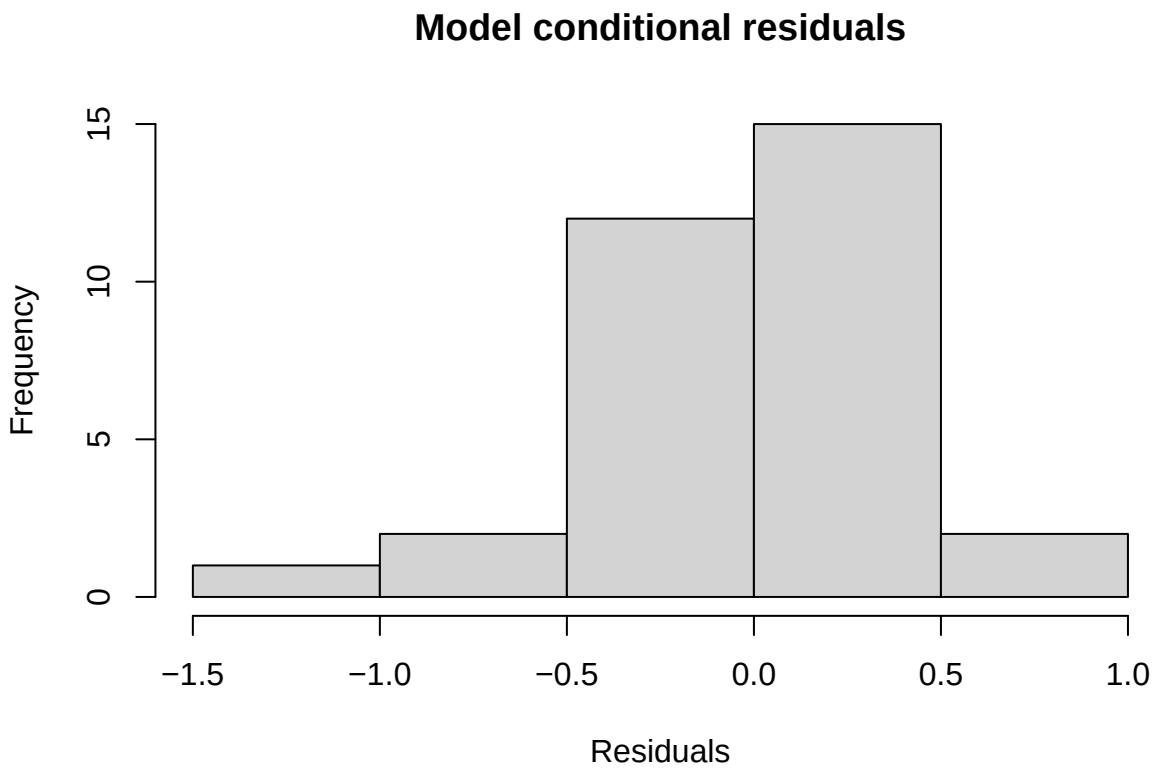
4.2.3 Callow *F. sanguinea*

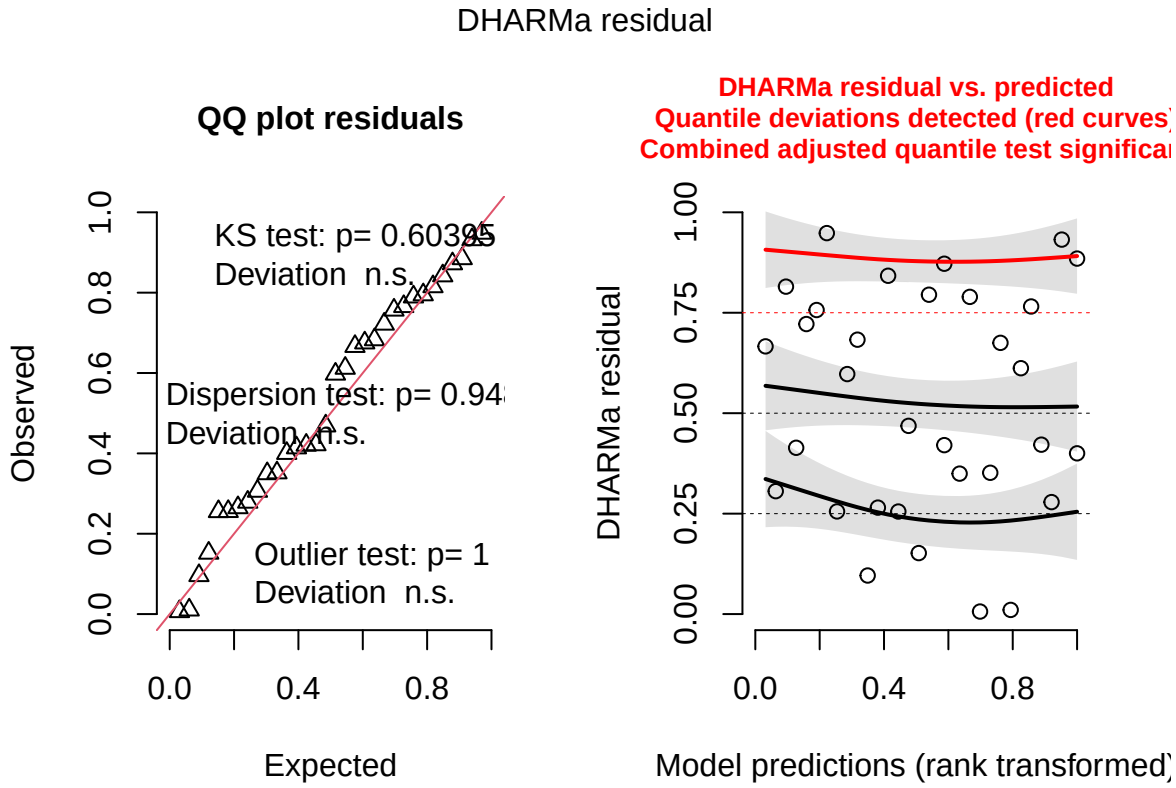
$\log(\text{CHC\_mass\_callow}) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon$,

where CHC_mass_callow denotes the total normalized CHC mass on the body of callow *F. sanguinea* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass)) ~ sang_prop + (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: 51.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.16539 -0.61246  0.07593  0.62418  1.49264
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.06331  0.2516
##  Residual             0.22509  0.4744
## Number of obs: 32, groups:  colony, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
```

```
## (Intercept)  0.03373    0.17027 28.88675   0.198  0.84435
## sang_prop    1.05723    0.30345 25.91578   3.484  0.00177 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.775

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.95692, p-value = 0.2258
```





Similarly to mature *F. sanguinea* ants, in callow workers there is negative but not significant relationship between body size and normalized CHC amount.

4.2.4 Predictions from the linear model

The linear models explaining the change in CHC amounts in relation to the proportion of *F. sanguinea* workers in a colony were used to predict the normalized CHC amounts of workers of both species in pure (monospecific) colonies. The predictions were obtained by sampling the distribution of the response variable with random effects integrated out. The prediction distributions for both species were matched element-wise, and for each pair of values the CHC mass of *F. sanguinea* was divided by the CHC mass of *F. fusca* to generate the distribution of ratios of the predicted normalized mass.

4.3 Change in CHC characteristic of *F. sanguinea*

In this section the subset of peaks were used to calculate normalized CHC mass. These were peaks identified as *F. sanguinea* markers.

4.3.1 Mature *F. sanguinea* ants

$\log(\text{CHC\_mass\_mature}) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon,$

where CHC_mass_mature denotes the normalized mass of *F. sanguinea* markers on the body of adult *F. sanguinea* worker.

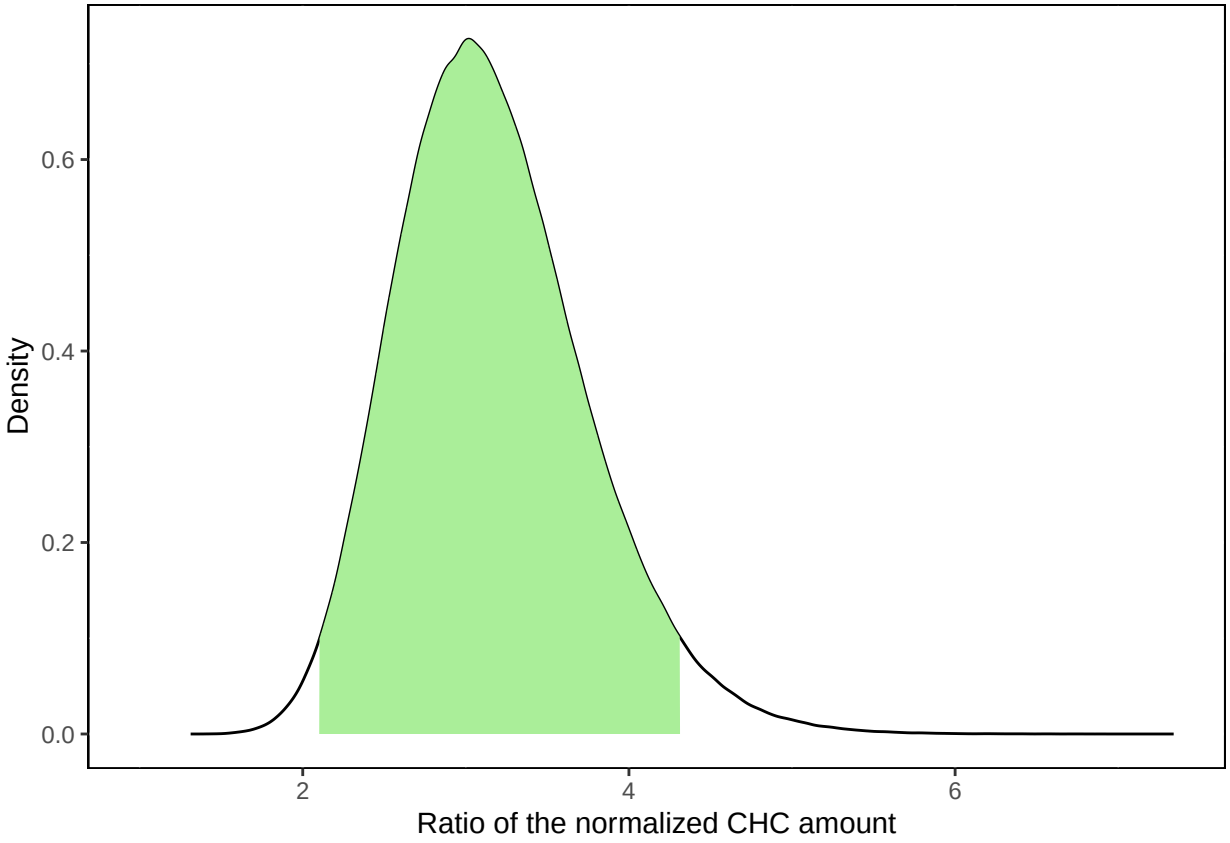
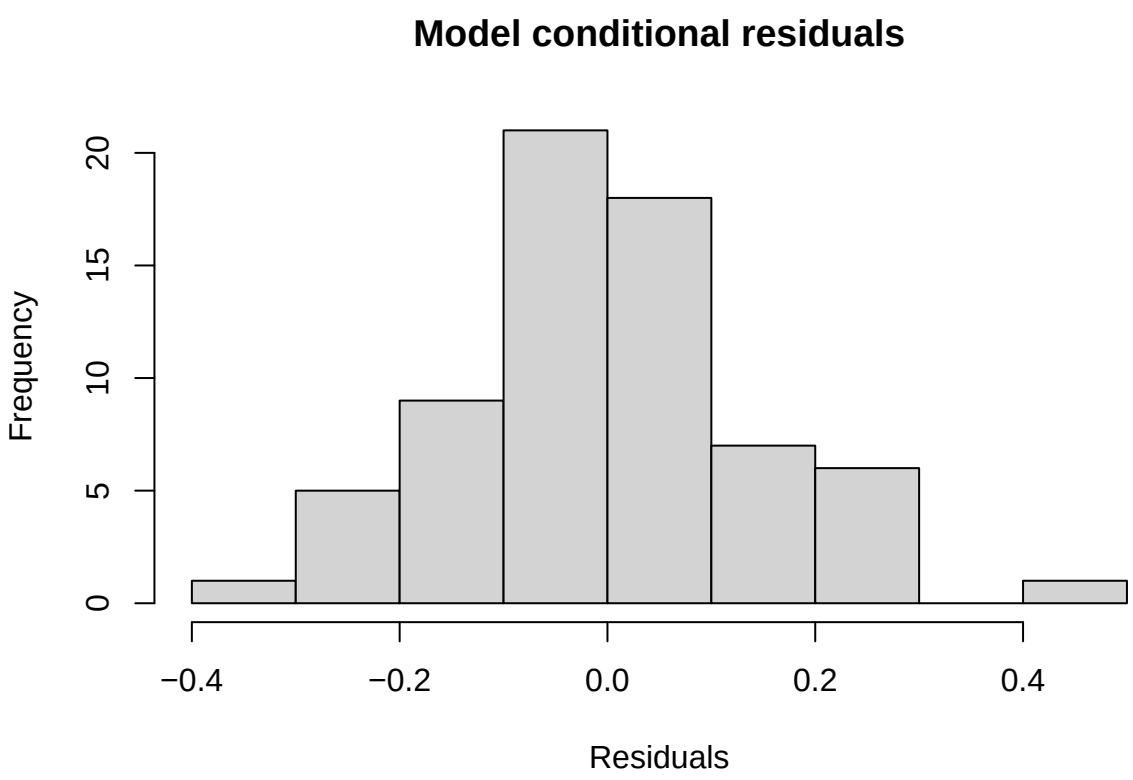
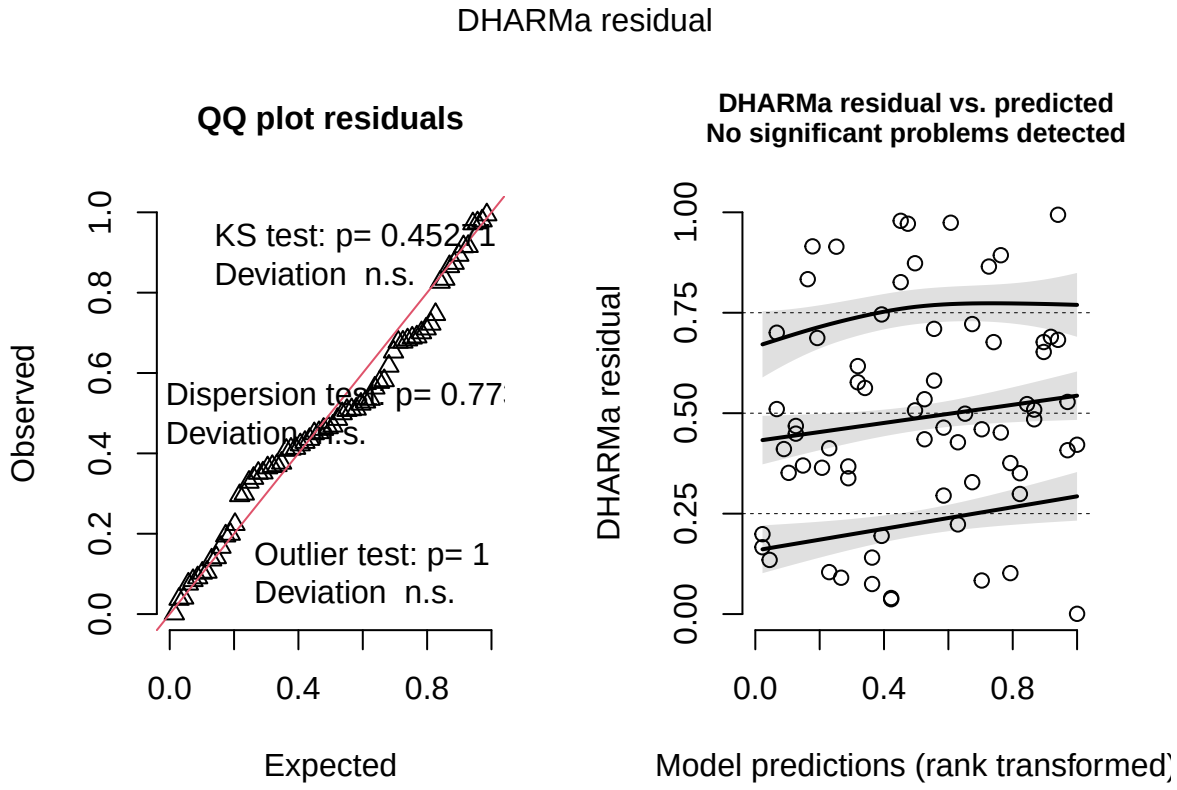


Figure S8 : Ditribution of the ratios (*F. sanguinea*/*F. fusca*) of the predicted CHC mass of workers from pure colonies. The shaded area corresponds to the 95% highest density interval.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass + 1)) ~ sang_prop + (1 | colony:census_date) +      (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: 30.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8821 -0.3658 -0.0362  0.3439  2.1883
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.06558  0.2561
##  colony             (Intercept) 0.00000  0.0000
##  Residual                       0.03804  0.1950
## Number of obs: 68, groups:  colony:census_date, 44; colony, 16
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   0.2833     0.0805 40.1068   3.519  0.00109 **
## sang_prop     1.3131     0.1564 39.5702   8.394 2.54e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.822
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')


##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98702, p-value = 0.7046
```


Model conditional residuals
Frequency
0
5
10
15
20
−0.4
−0.2
0.0
0.2
0.4
Residuals



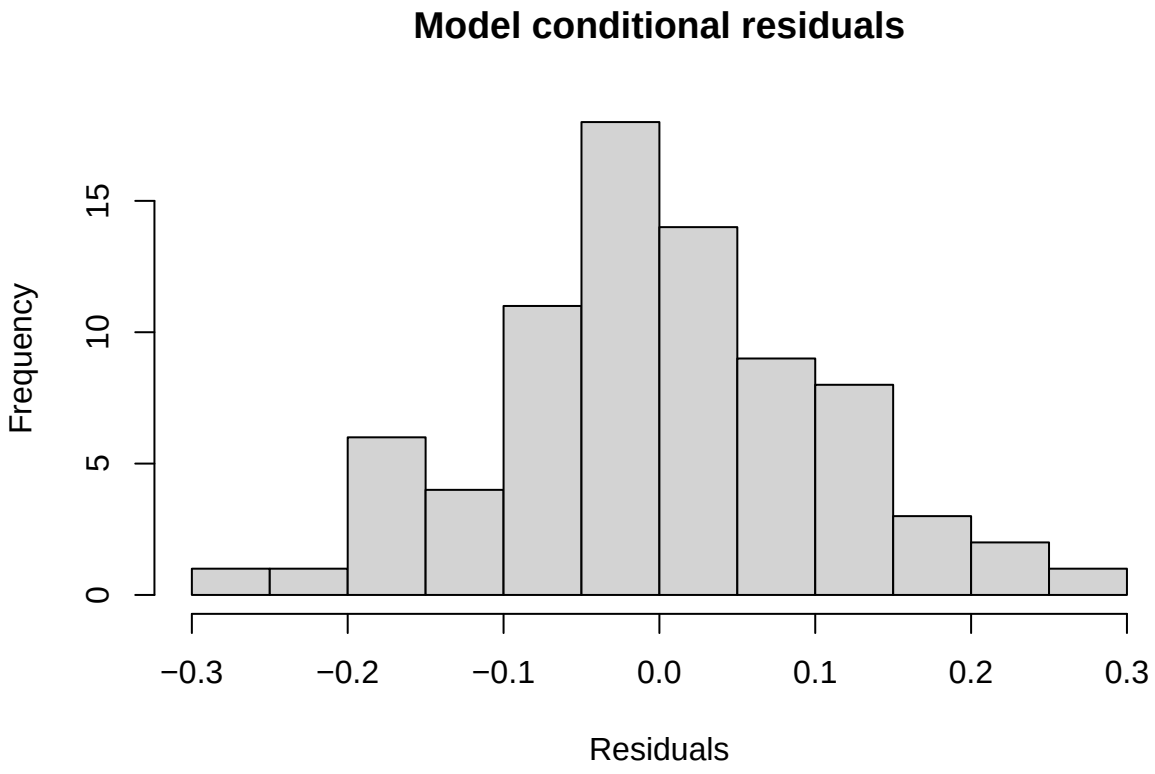
4.3.2 *F. fusca* ants

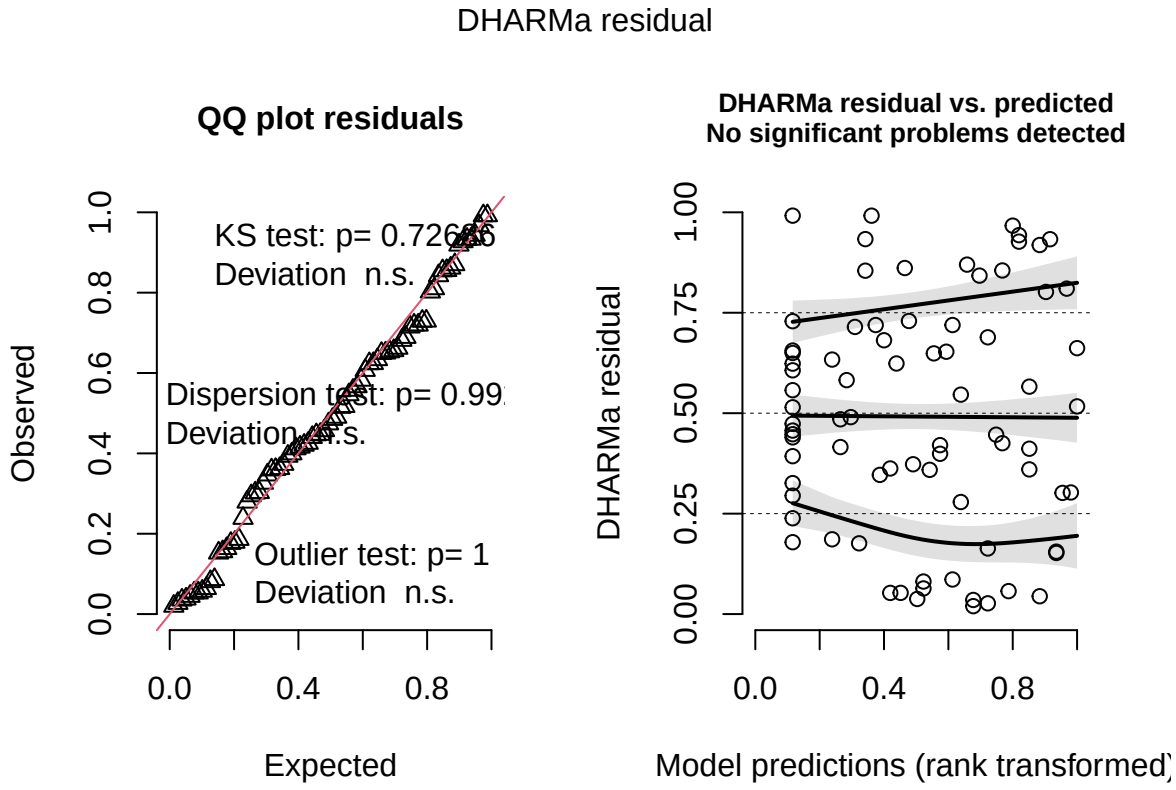
$\sqrt{(\text{CHC\_mass\_fusca})} = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon$,

where CHC_mass_fusca denotes the normalized mass of *F. sanguinea* markers on the body of adult *F. sanguinea* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(sqrt(normalized_mass)) ~ sang_prop + (1 | colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: -58.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.06253 -0.51814 -0.07238  0.45433  1.95590
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.009805 0.09902
##  Residual                       0.016560 0.12869
## Number of obs: 78, groups:  colony:census_date, 55
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
```

```
## (Intercept)  0.41024    0.02938 54.47670  13.965  < 2e-16 ***
## sang_prop    0.61852    0.06935 51.99602   8.919 4.63e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.729

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.99067, p-value = 0.8465
```





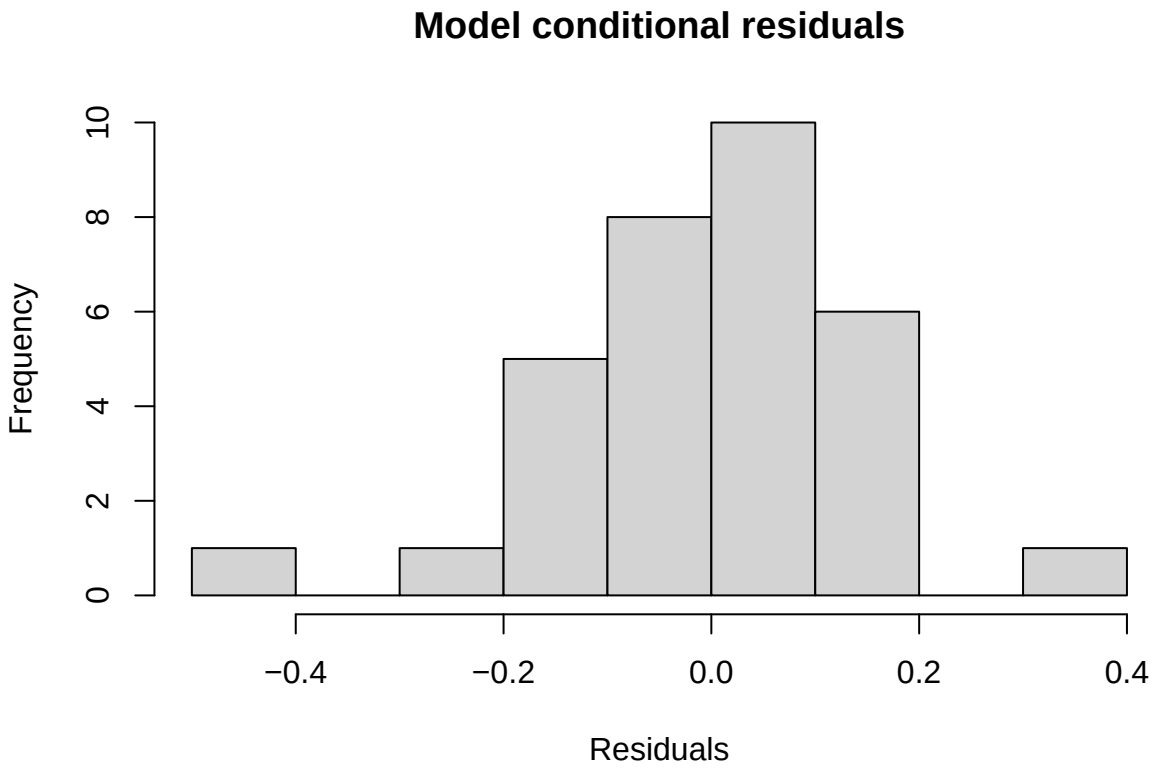
4.3.3 Callow *F. sanguinea* ants

$$\sqrt[3]{(\text{CHC\_mass\_mature})} = \beta_0 + \text{sanguinea\_proportion} + \epsilon,$$

where CHC_mass_mature denotes the normalized mass of *F. sanguinea* markers on the body of callow *F. sanguinea* worker.

```
##
## Call:
## lm(formula = I((normalized_mass)^(1/3)) ~ sang_prop, data = model_input)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41432 -0.06803  0.02430  0.07590  0.35296
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.31592    0.04638   6.812 1.48e-07 ***
## sang_prop    0.71797    0.08719   8.235 3.42e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1473 on 30 degrees of freedom
## Multiple R-squared:  0.6933, Adjusted R-squared:  0.6831
## F-statistic: 67.81 on 1 and 30 DF,  p-value: 3.423e-09
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.96738, p-value = 0.4305
```



4.4 Change in CHC characteristic of *F. fusca*

Similarly to the analysis with the use of markers of *F. sanguinea* workers, the procedure was repeated using the mass of *F. fusca* markers as a response variable.

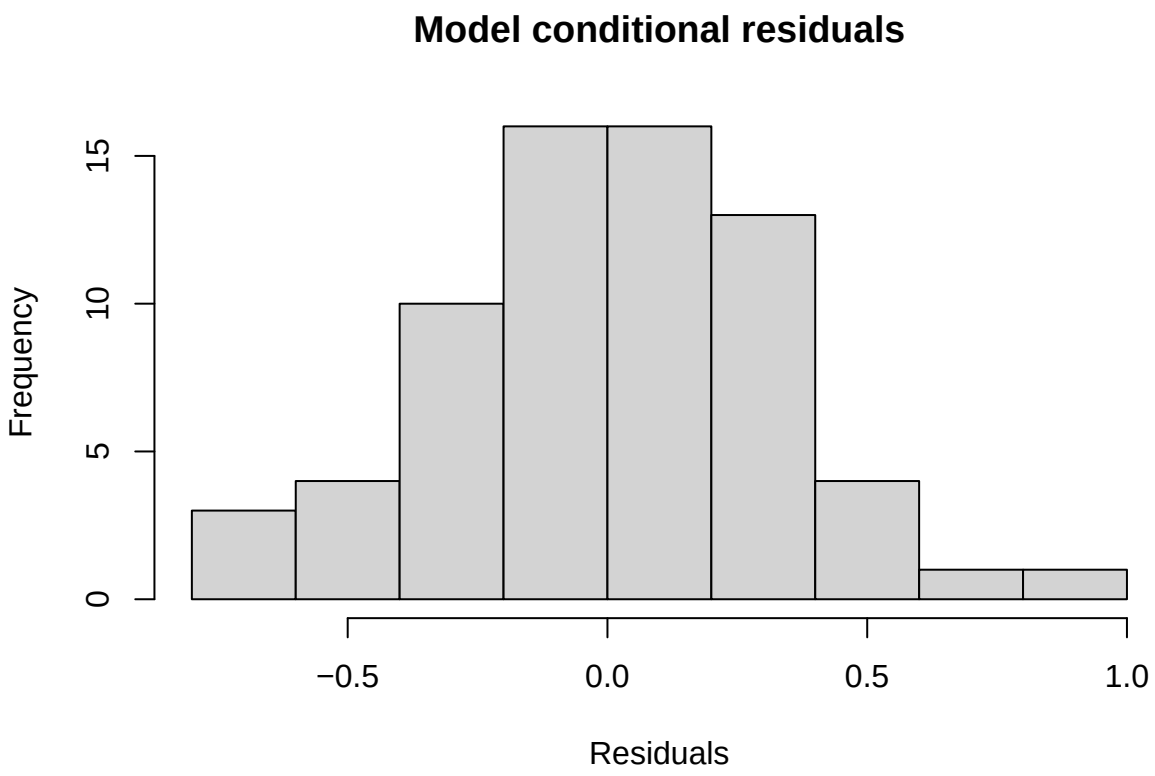
4.4.1 Mature *F. sanguinea* ants

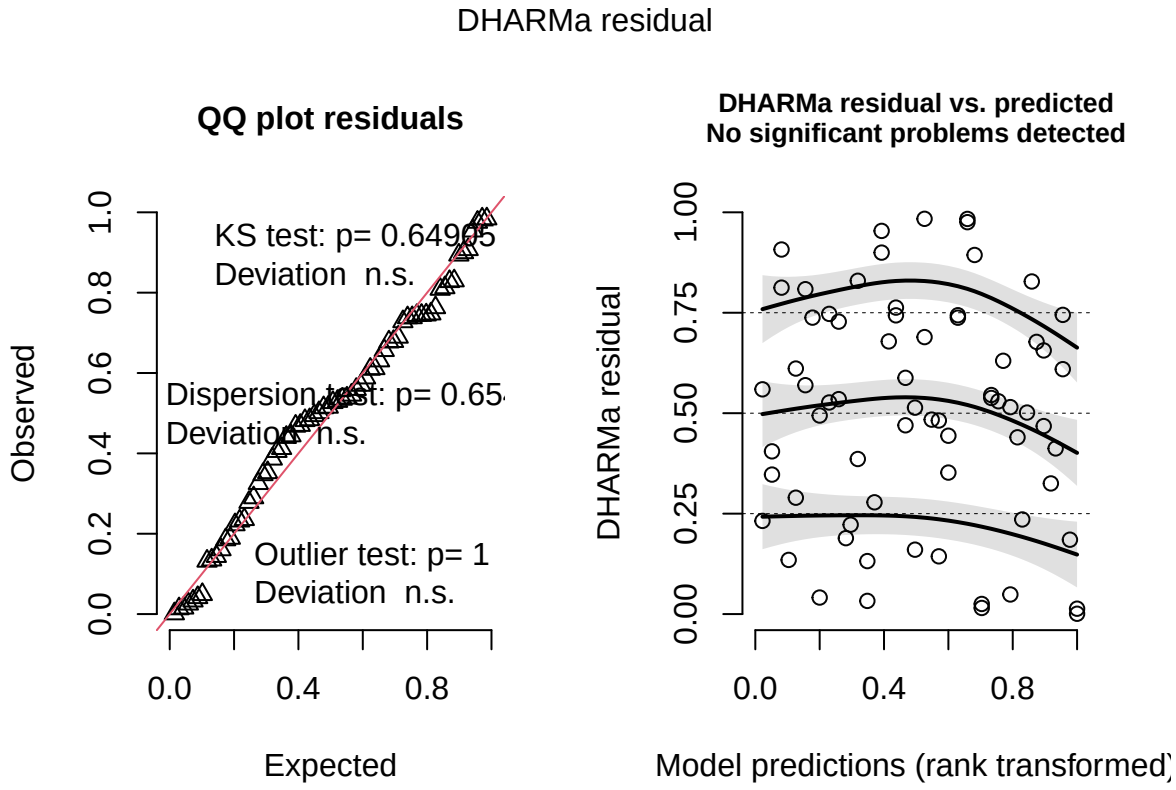
$\log(\text{CHC\_mass\_mature}) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon$,

where CHC_mass_mature denotes the normalized mass of *F. fusca* markers on the body of adult *F. sanguinea* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass)) ~ sang_prop + (1 | colony:census_date) +      (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: 131.2
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82877 -0.45873  0.05316  0.55616  1.98511
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.1471   0.3835
##  colony             (Intercept) 0.2359   0.4857
##  Residual                       0.1804   0.4247
## Number of obs: 68, groups:  colony:census_date, 44; colony, 16
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  -0.4566     0.1947 33.8202  -2.345 0.025027 *
## sang_prop    -1.1016     0.2951 33.2726  -3.733 0.000707 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.656


##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98971, p-value = 0.8512
```


Model conditional residuals
Frequency
0
5
10
15
−0.5
0.0
0.5
1.0
Residuals

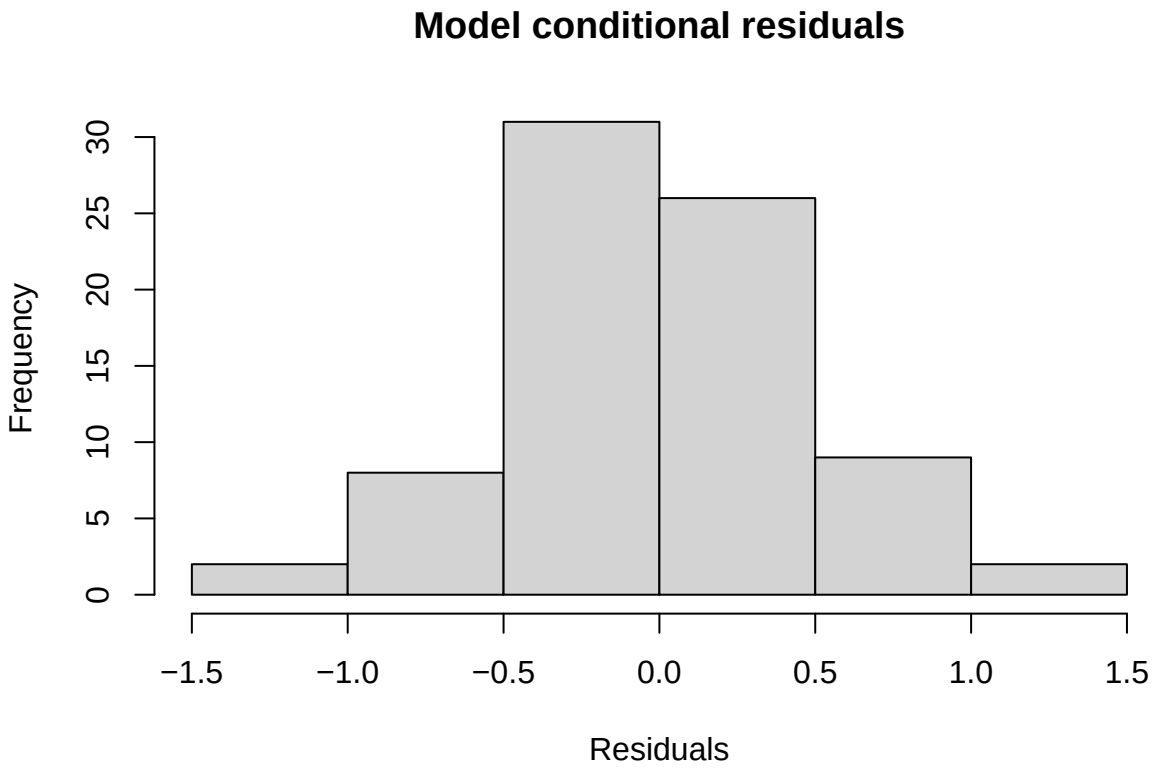


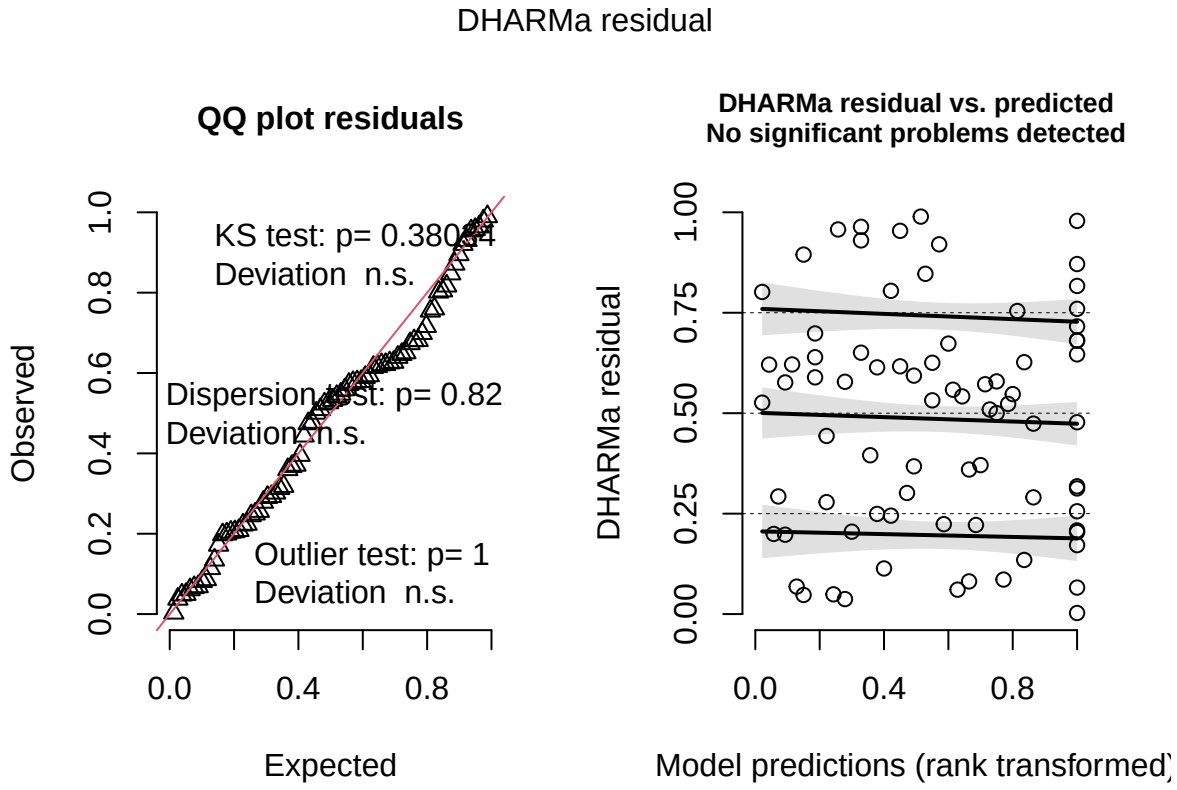
4.4.2 *F. fusca* ants

$\log(\text{CHC\_mass\_mature}+0.01) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + (1|\text{colony\_ID:sampling\_occasion\_ID}) + \epsilon$, where CHC_mass_mature denotes the normalized mass of *F. fusca* markers on the body of *F. fusca* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass + 0.01)) ~ sang_prop + (1 | colony) + (1 |      colony:census_date)
##    Data: model_input
##
## REML criterion at convergence: 159.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.34120 -0.60468 -0.05522  0.59876  2.14890
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  colony:census_date (Intercept) 0.0000   0.0000
##  colony             (Intercept) 0.2325   0.4821
##  Residual                       0.3213   0.5668
## Number of obs: 78, groups:  colony:census_date, 55; colony, 20
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
```

```
## (Intercept)  -0.8547     0.1462 27.2550  -5.844 3.08e-06 ***
## sang_prop    -0.8184     0.2436 65.7807  -3.360   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.479
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.99138, p-value = 0.8835
```





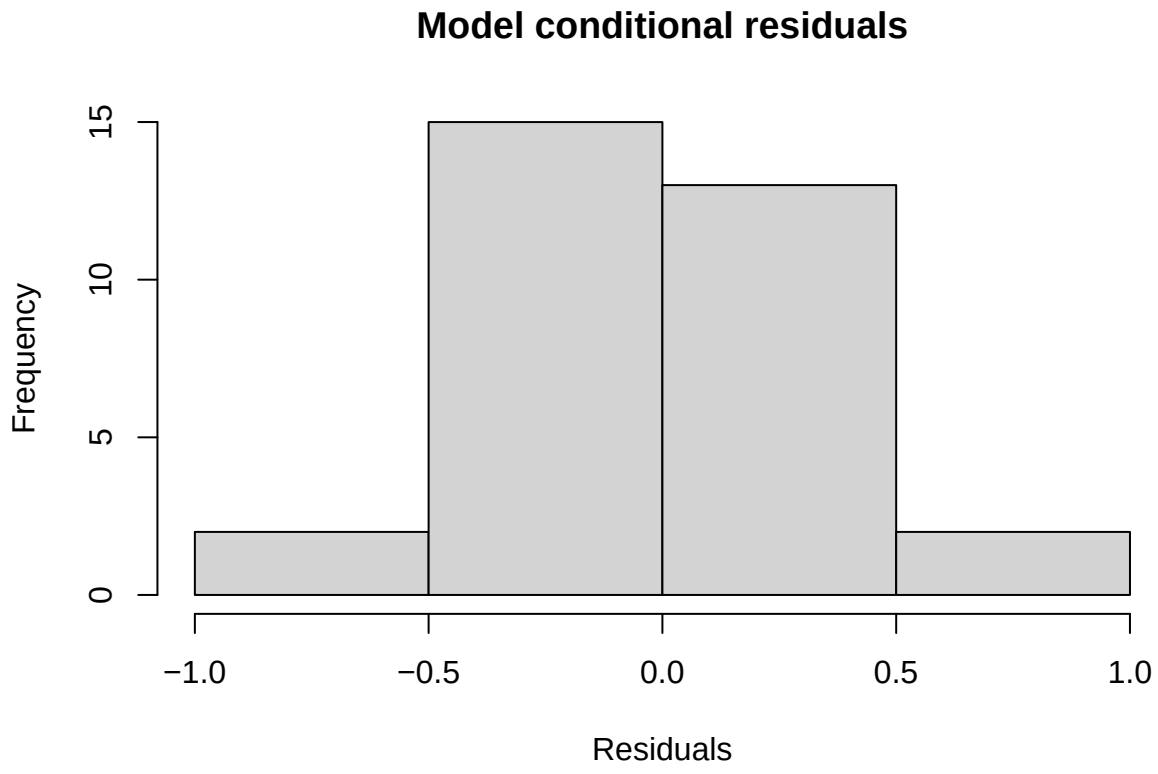
4.4.3 Callow *F. sanguinea* ants

$$\log(\text{CHC\_mass\_mature}) = \beta_0 + \text{sanguinea\_proportion} + (1|\text{colony\_ID}) + \epsilon,$$

where log(CHC_mass_mature) denotes the normalized mass of *F. fusca* markers on the body of callow *F. sanguinea* worker.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(log(normalized_mass)) ~ sang_prop + (1 | colony)
##    Data: model_input
##
## REML criterion at convergence: 59.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.88270 -0.59925 -0.01403  0.65154  1.88156
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.2208   0.4699
##  Residual             0.2225   0.4717
## Number of obs: 32, groups:  colony, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
```

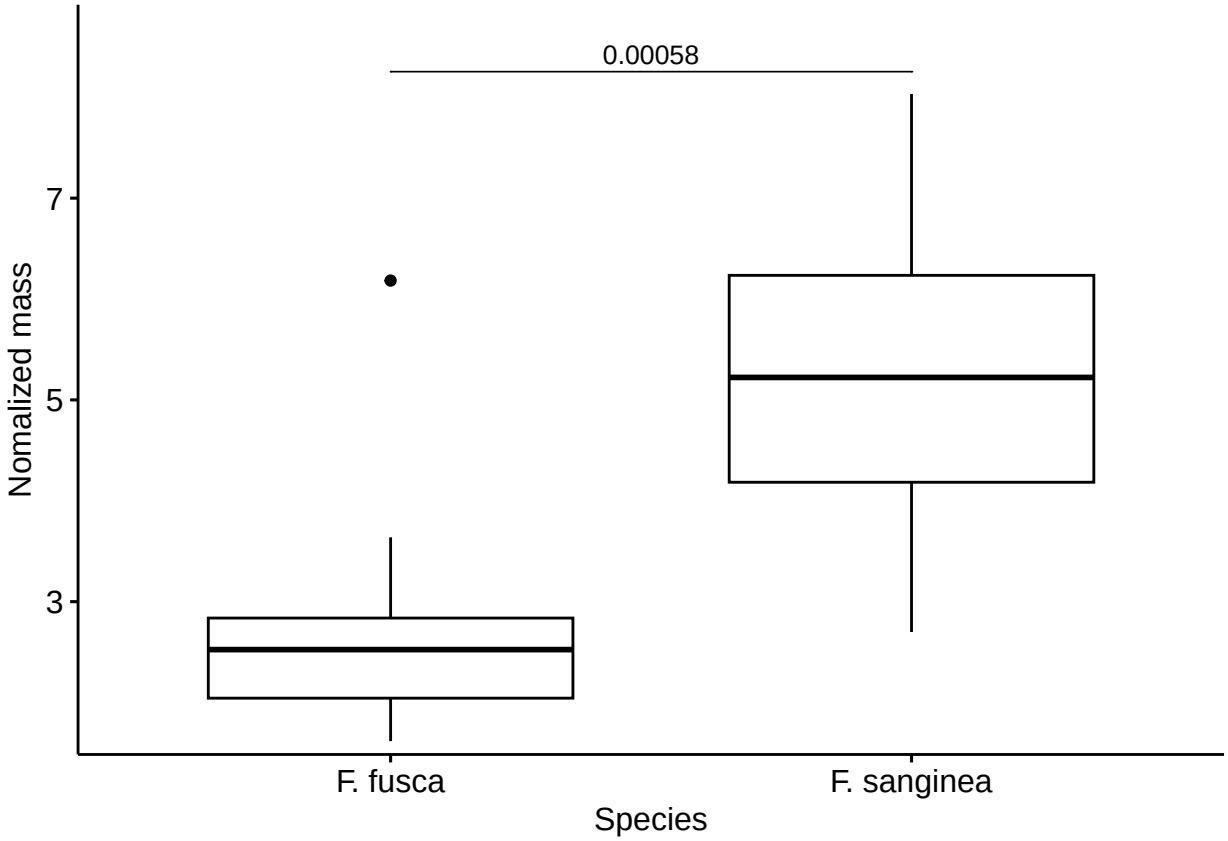
```
## (Intercept) -1.66659    0.20293 28.05257  -8.213 6.04e-09 ***
## sang_prop   -0.07903    0.32299 19.69963  -0.245    0.809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## sang_prop -0.685

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98634, p-value = 0.9487
```



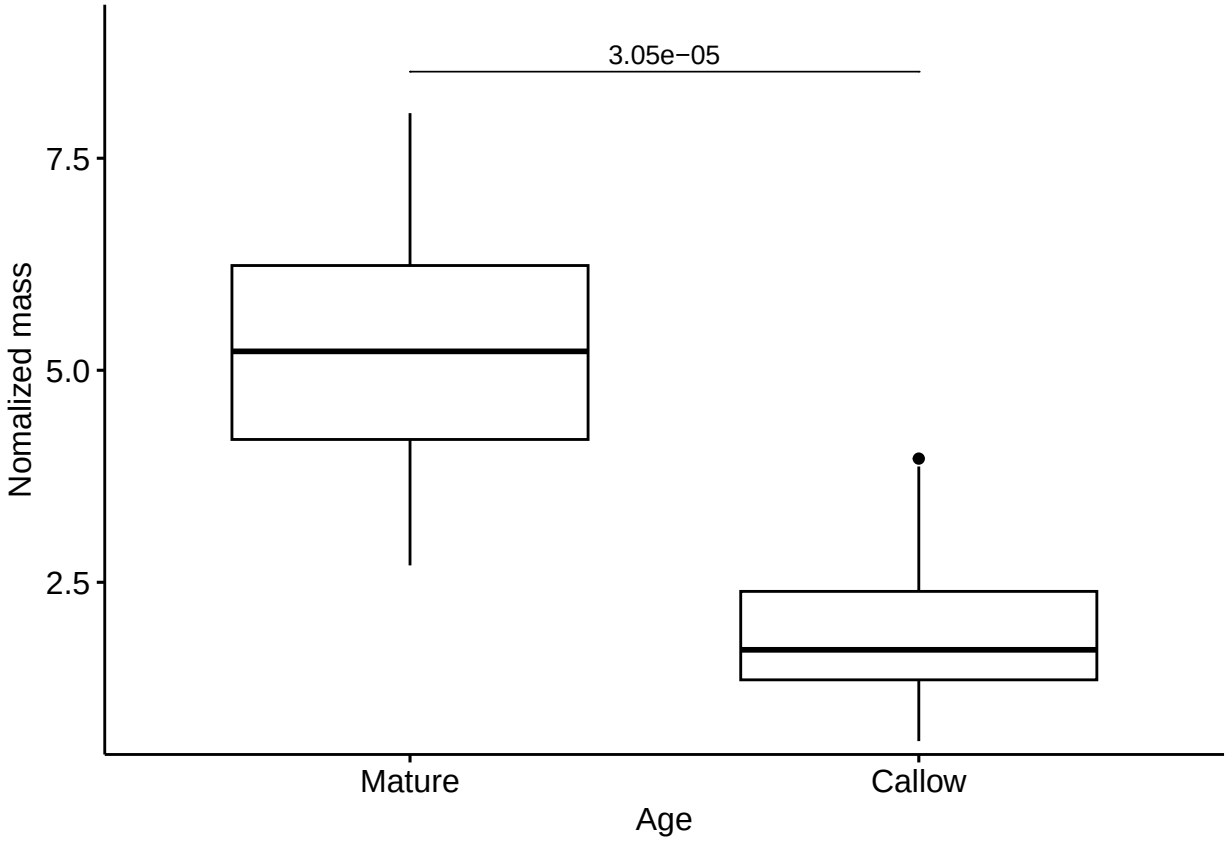
5 Comparisons of CHC amounts and proportions between species and age categories

This section presents the results of Wilcoxon tests comparing features of CHC profiles between callow and mature *F. sanguinea* ants as well as *F. fusca* slaves. Samples from the same colony were averaged before calculation of the final statistics to account for their non-independence.

5.1 Difference in total CHC amount between *F. sanguinea* and *F. fusca*

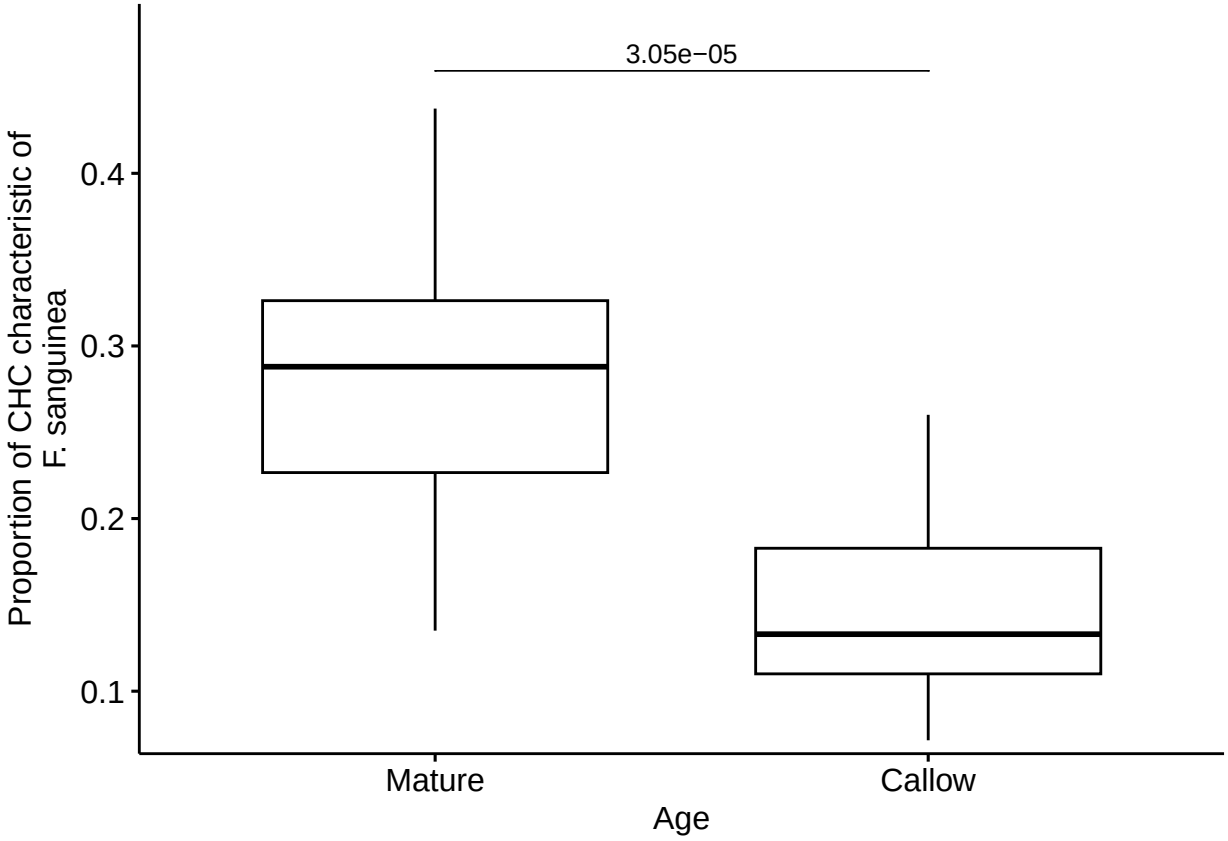


5.2 Difference in total CHC amount between callow and mature *F. sanguinea*

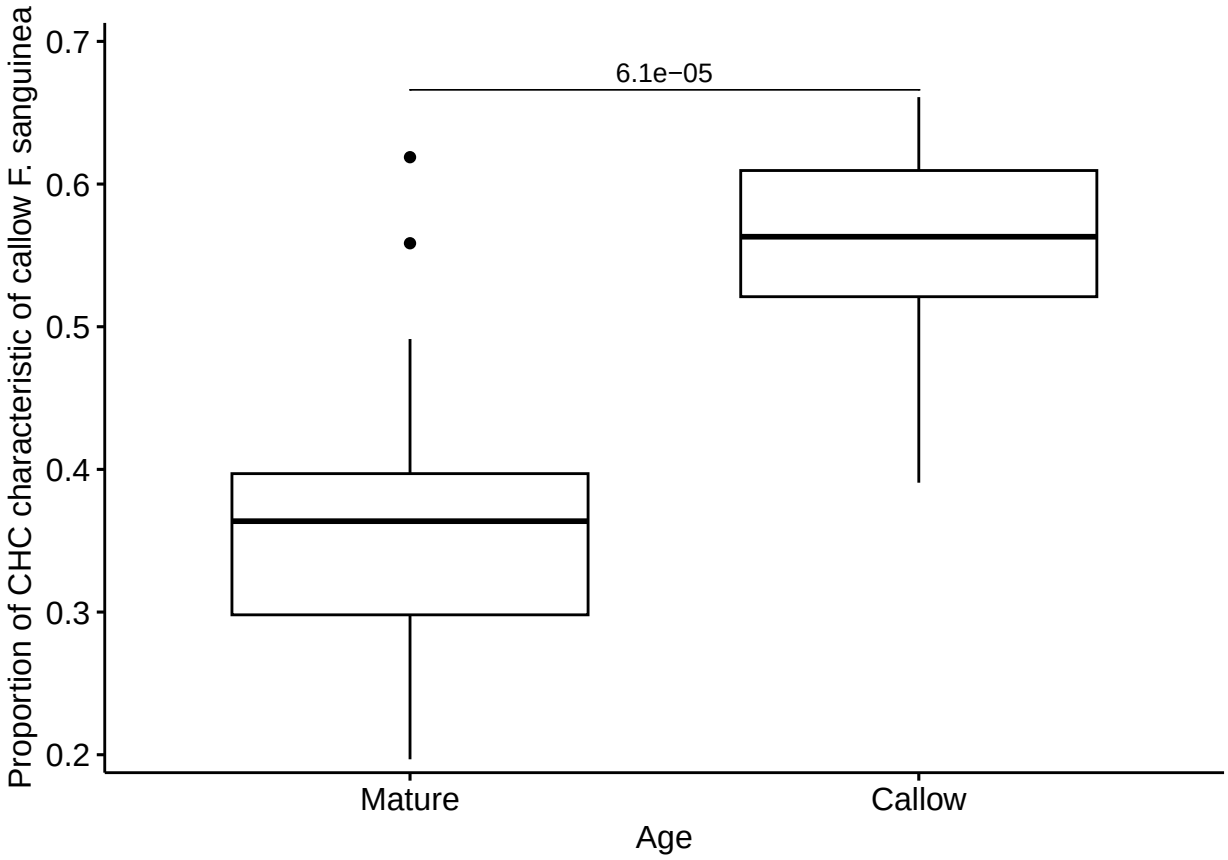


5.3 Difference in the proportion of CHC characteristic of *F. sanguinea* between mature *F. sanguinea* and *F. fusca*

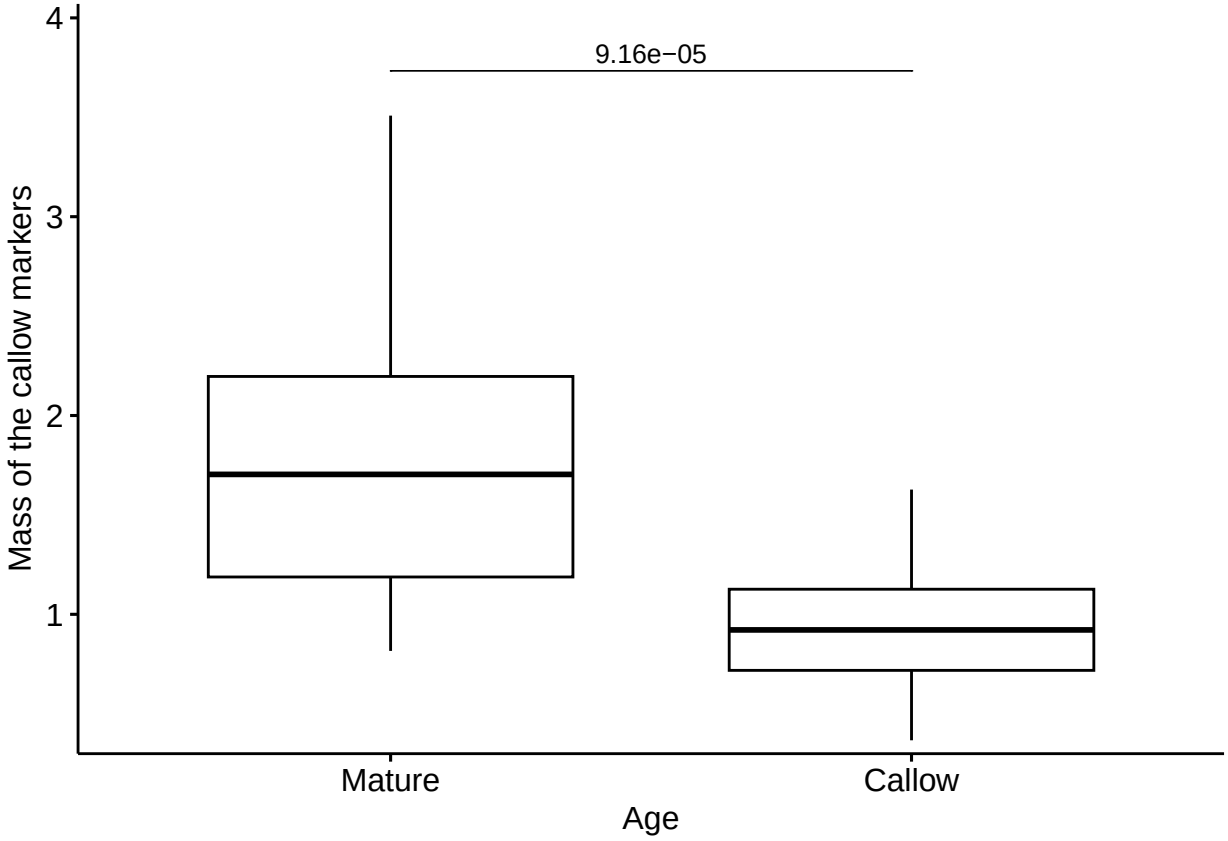
5.4 Difference in the proportion of CHC characteristic of *F. sanguinea* between mature and callow *F. sanguinea*



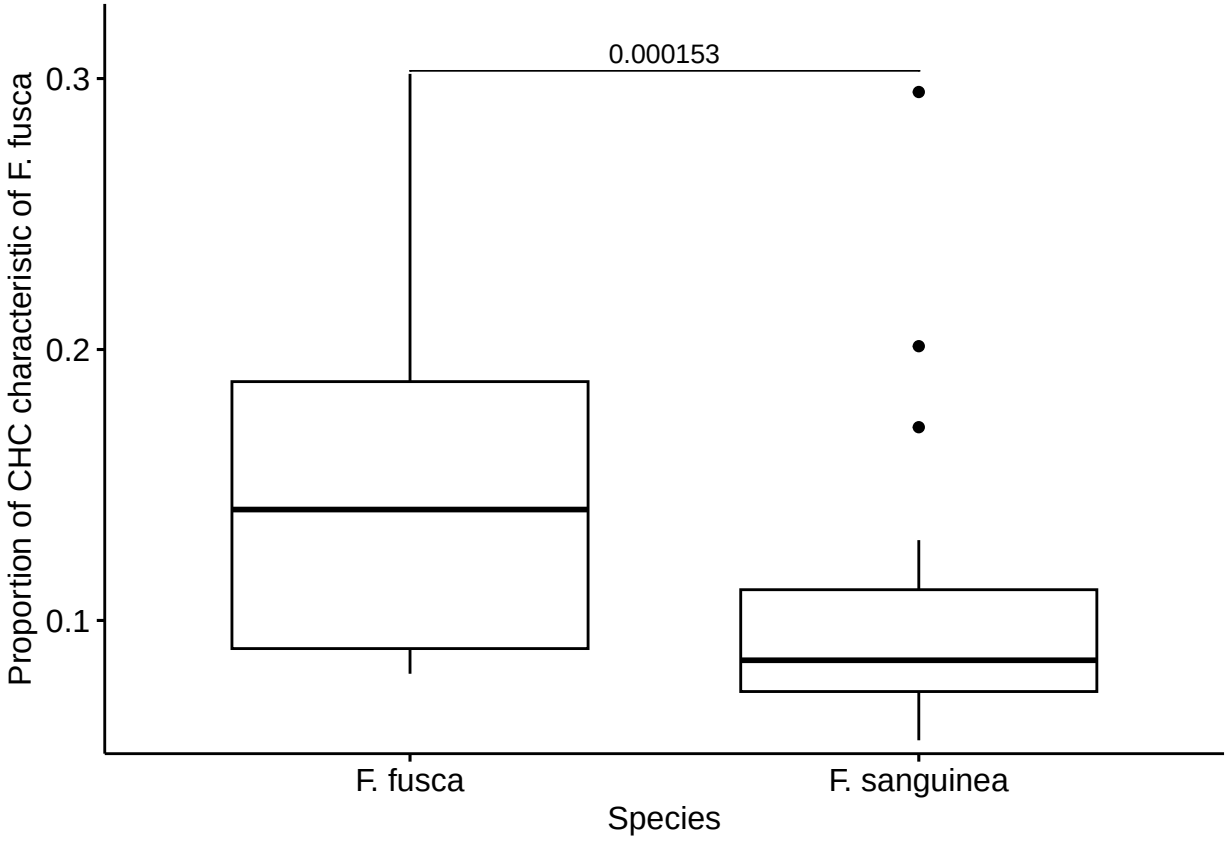
5.5 Difference in the proportion of CHC characteristic of callow *F. sanguinea* between mature and callow *F. sanguinea*



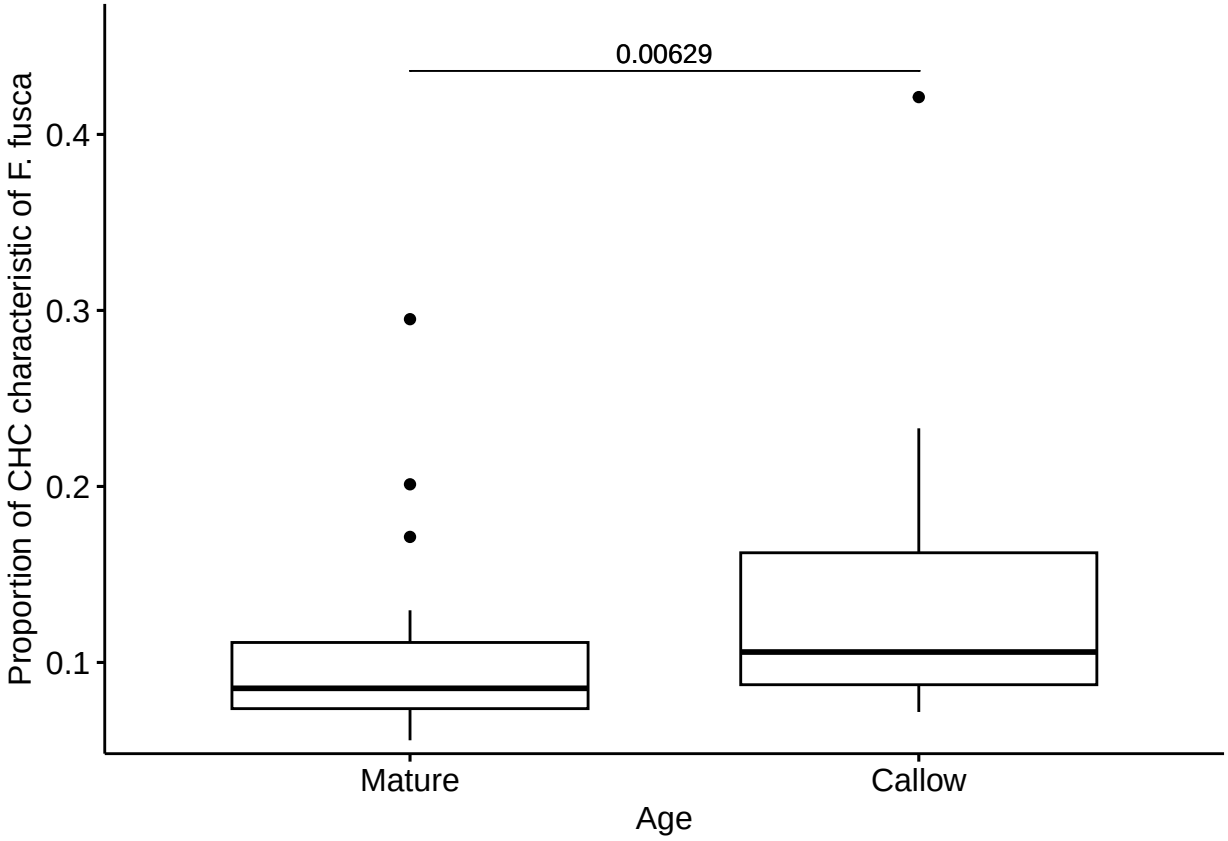
5.6 Difference in the amount of CHC characteristic of callow *F. sanguinea* between mature and callow *F. sanguinea*



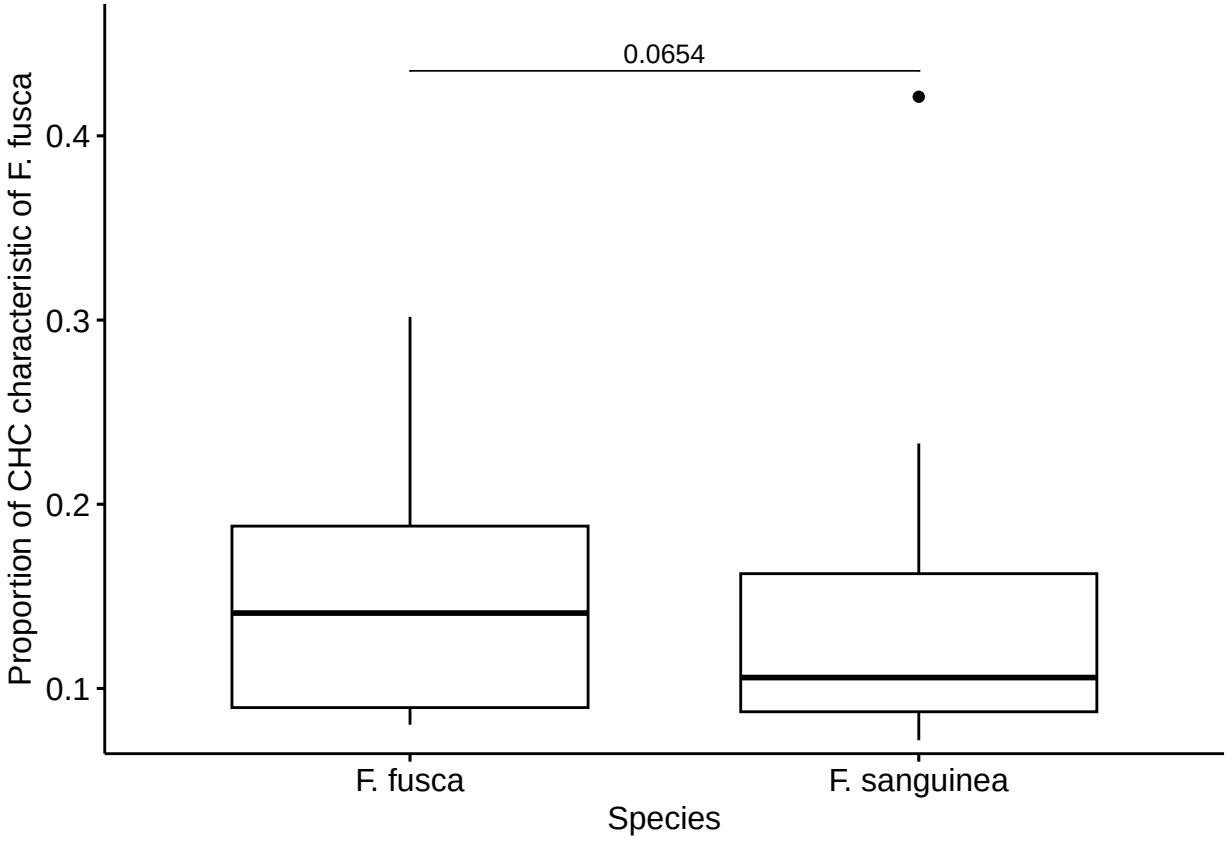
5.7 Difference in the proportion of CHC characteristic of *F. fusca* between mature *F. sanguinea* and *F. fusca*



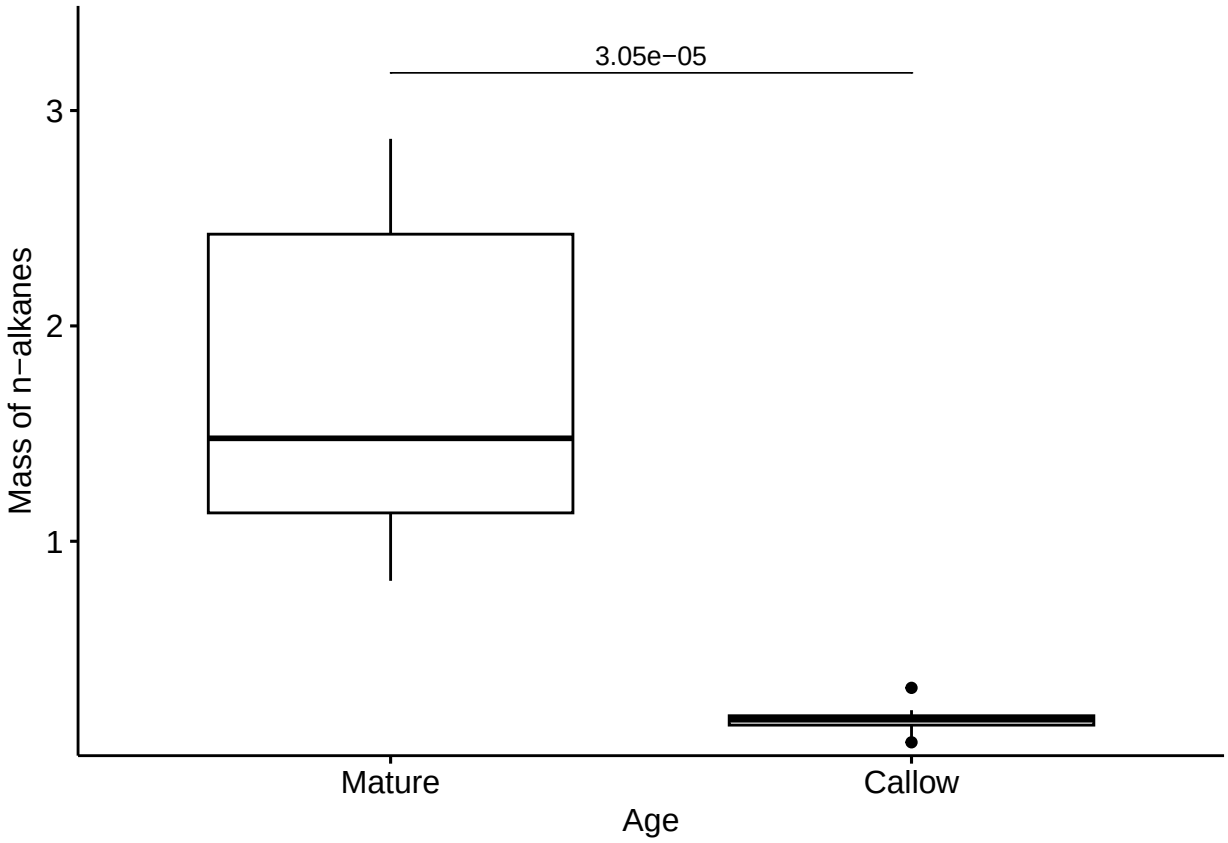
5.8 Difference in the proportion of CHC characteristic of *F. fusca* between mature and callow *F. sanguinea*



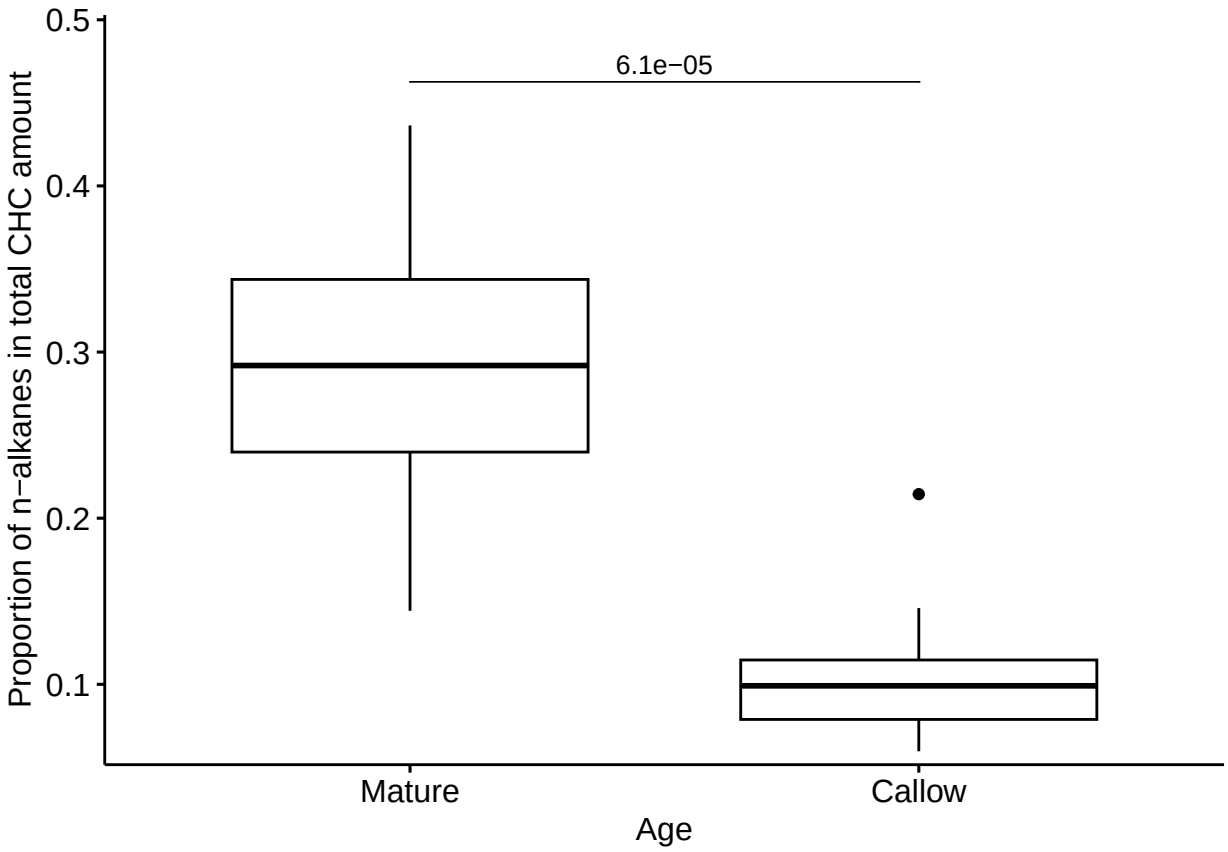
5.9 Difference in the proportion of CHC characteristic of *F. fusca* between callow *F. sanguinea* and *F. fusca*



5.10 Difference in the mass of *n*-alkanes between callow and mature *F. sanguinea*



5.11 Difference in the proportion of *n*-alkanes between callow and mature *F. sanguinea*



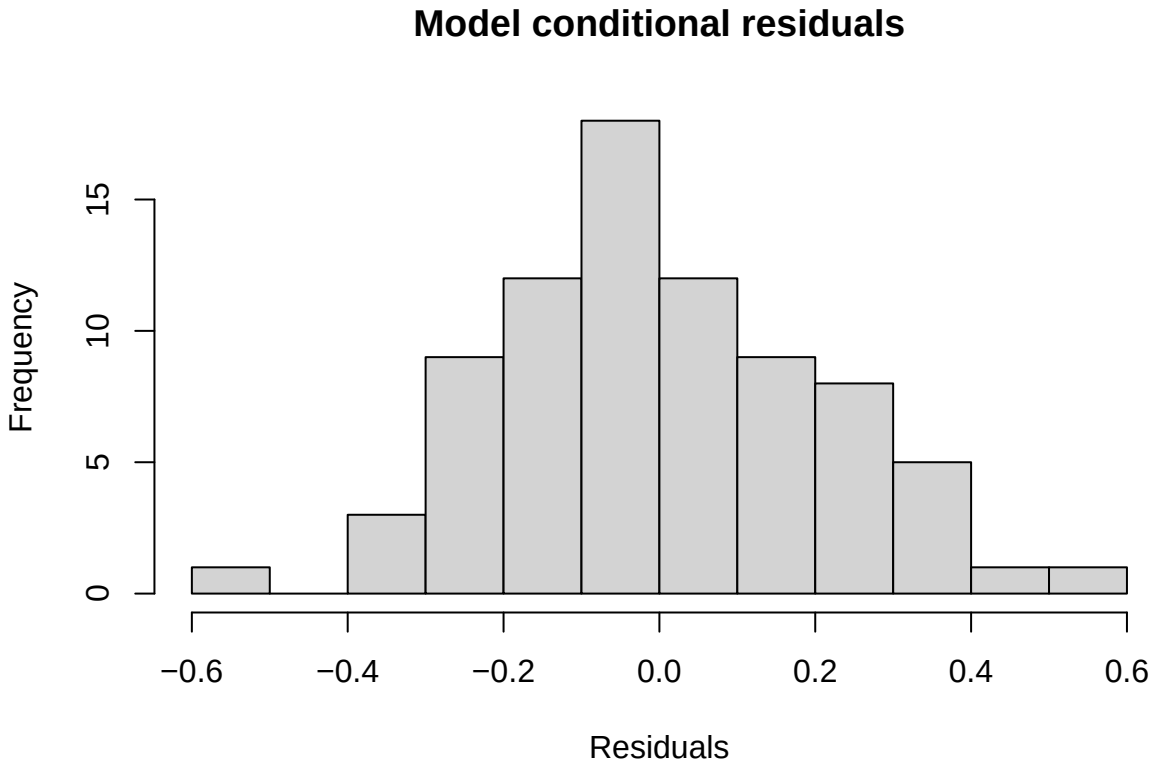
6 Change in the CHC profile of separated callow *F. sanguinea* ants

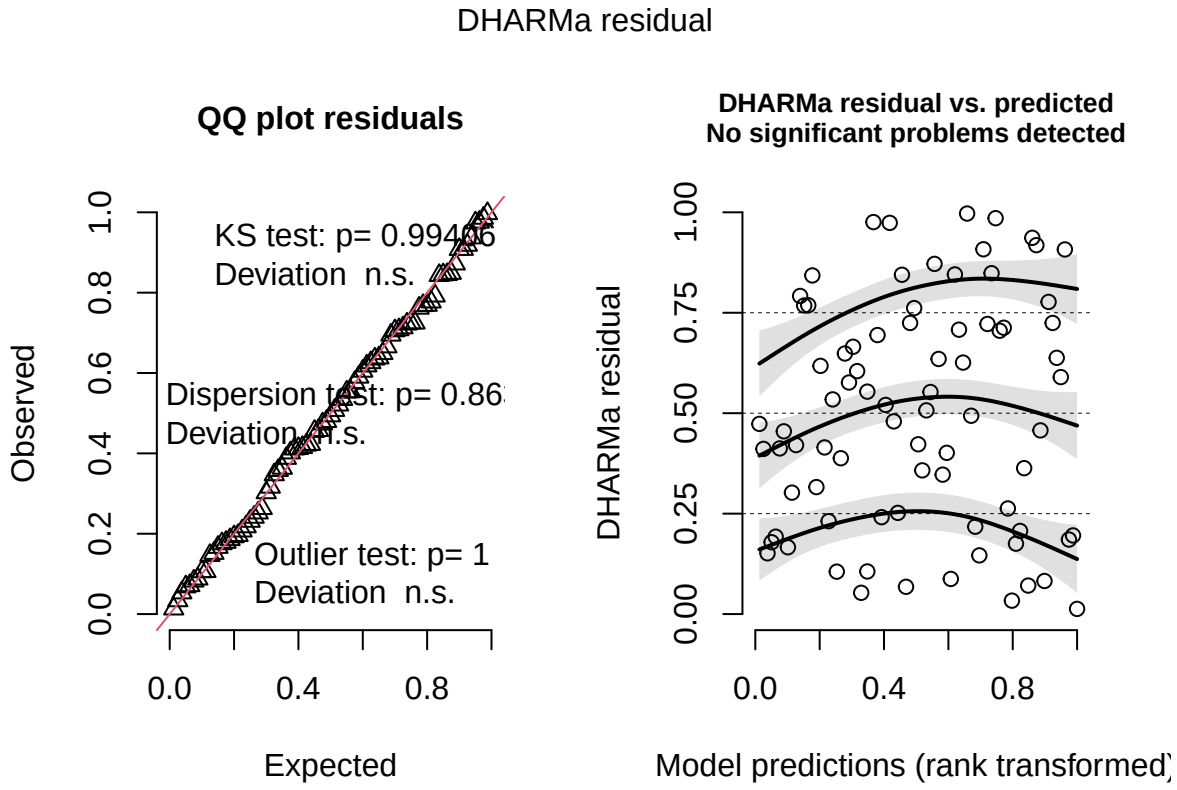
6.1 Change in total CHC amount over time

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(sqrt(normalized_mass)) ~ mean_delta + (1 | colony)
##    Data: separation_data[-74, ]
##
## REML criterion at convergence: 9.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.55998 -0.69772 -0.07088  0.67280  2.51792
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.01765  0.1329
##  Residual             0.04716  0.2172
## Number of obs: 79, groups:  colony, 11
```

```
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  1.367210   0.055393 16.298104  24.682 2.43e-14 ***
## mean_delta   0.002934   0.001933 68.480649   1.518    0.134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## mean_delta -0.515
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.9908, p-value = 0.8482
```

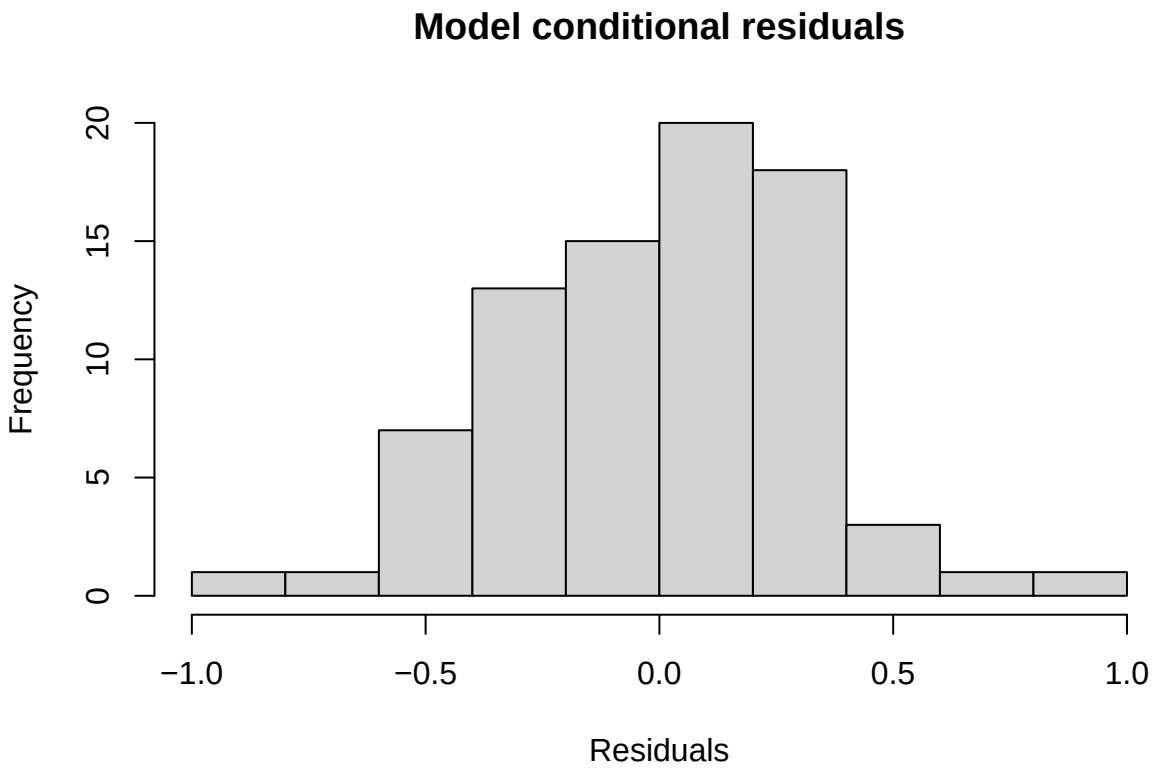


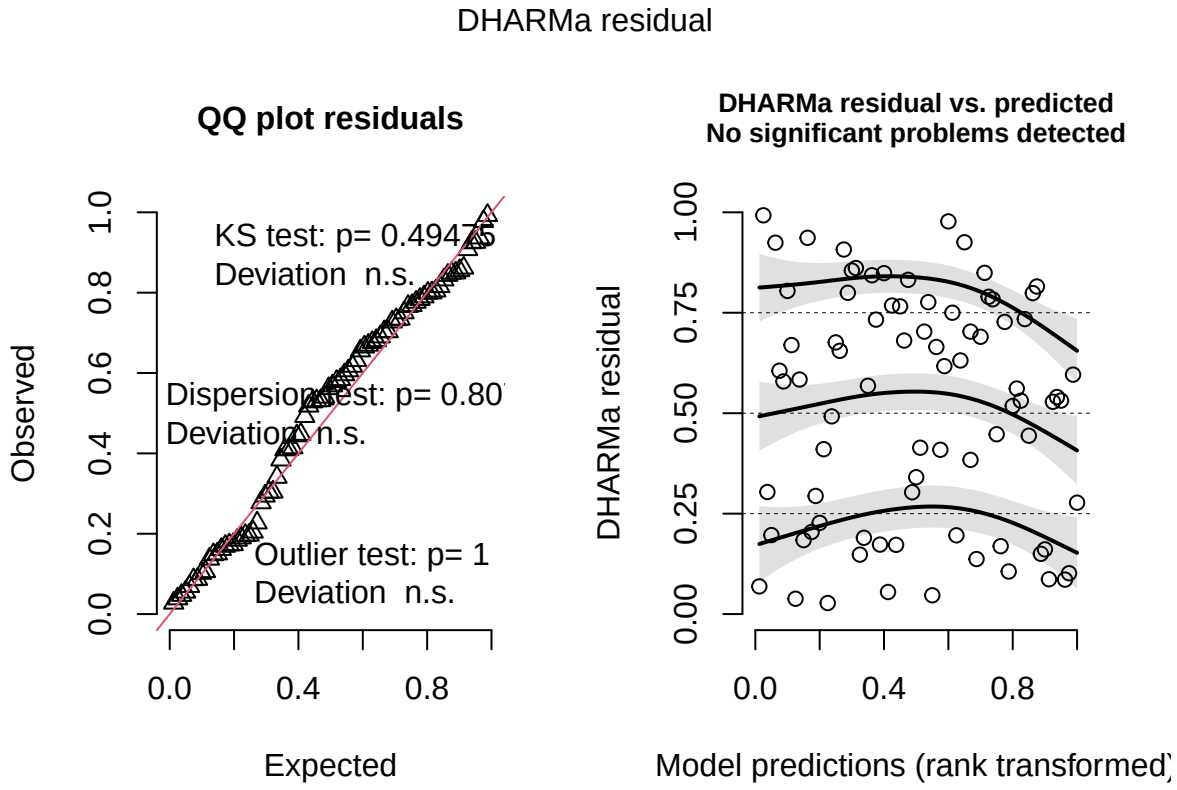


6.2 Change in the amount of compounds characteristic of callow *F. sanguinea*

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: log(normalized_mass_part) ~ mean_delta + (1 | colony)
##    Data: separation_data
##
## REML criterion at convergence: 92.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3303 -0.7669  0.1012  0.6201  2.5706
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.1129   0.3360
##  Residual             0.1238   0.3518
## Number of obs: 80, groups:  colony, 11
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) -0.003285   0.118687 13.016273  -0.028    0.978
## mean_delta  -0.004592   0.003054 68.334089  -1.503    0.137
##
## Correlation of Fixed Effects:
##            (Intr)
```

```
## mean_delta -0.389

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98886, p-value = 0.7216
```



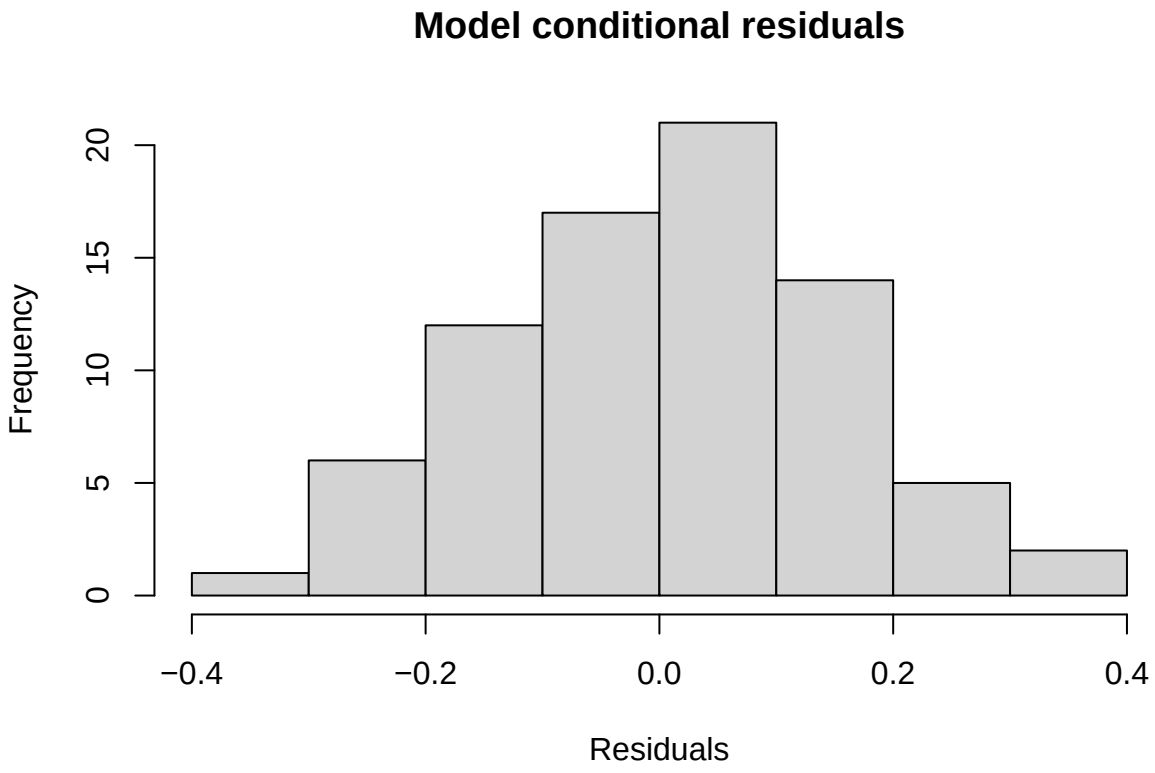


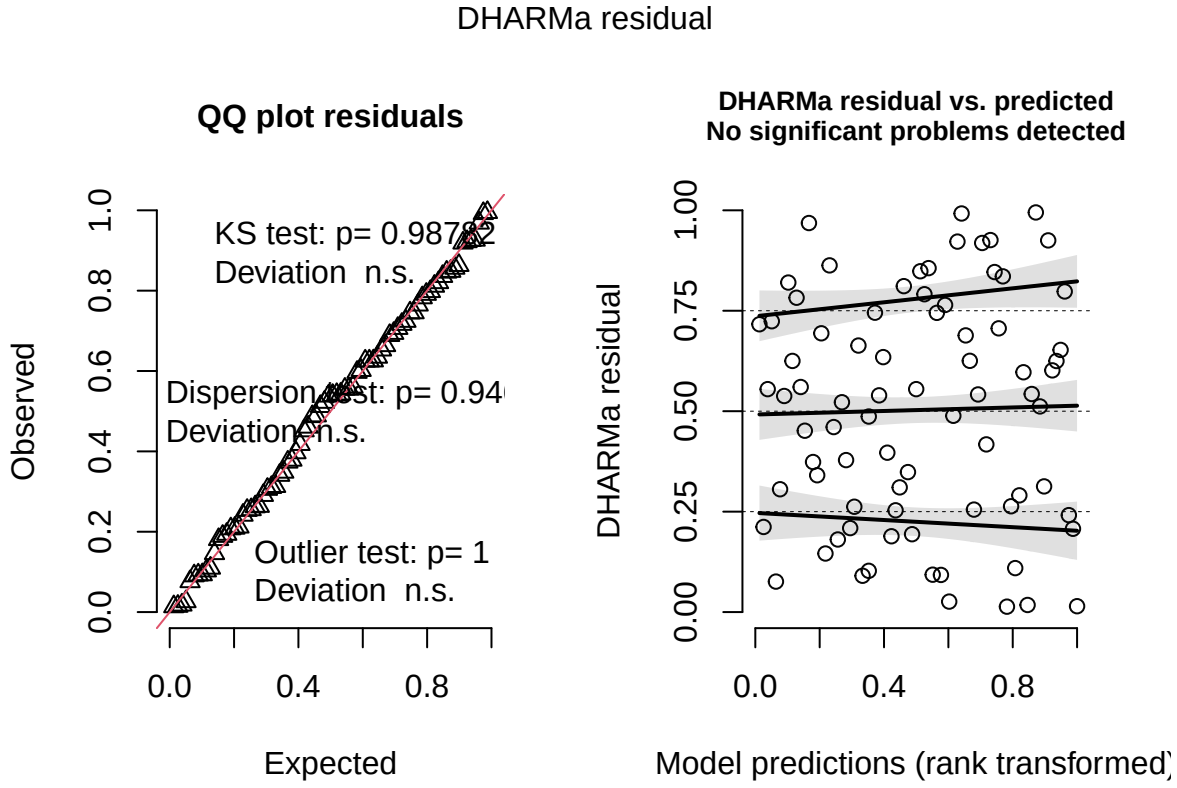
6.3 Change in the amount of compounds characteristic of *F. sanguinea*

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: I(sqrt(normalized_mass_part)) ~ poly(mean_delta, 1) + (1 | colony)
##    Data: separation_data[-c(69, 74), ]
##
## REML criterion at convergence: -67.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.27083 -0.64402  0.06067  0.73259  2.42696
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.002444 0.04944
##  Residual             0.021031 0.14502
## Number of obs: 78, groups:  colony, 11
##
## Fixed effects:
##                     Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)          0.60656    0.02252  7.88363  26.930 4.81e-09 ***
## poly(mean_delta, 1)  1.08605    0.14693 69.21997   7.392 2.54e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##             (Intr)
## ply(mn_d,1) 0.020
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.98936, p-value = 0.7684
```



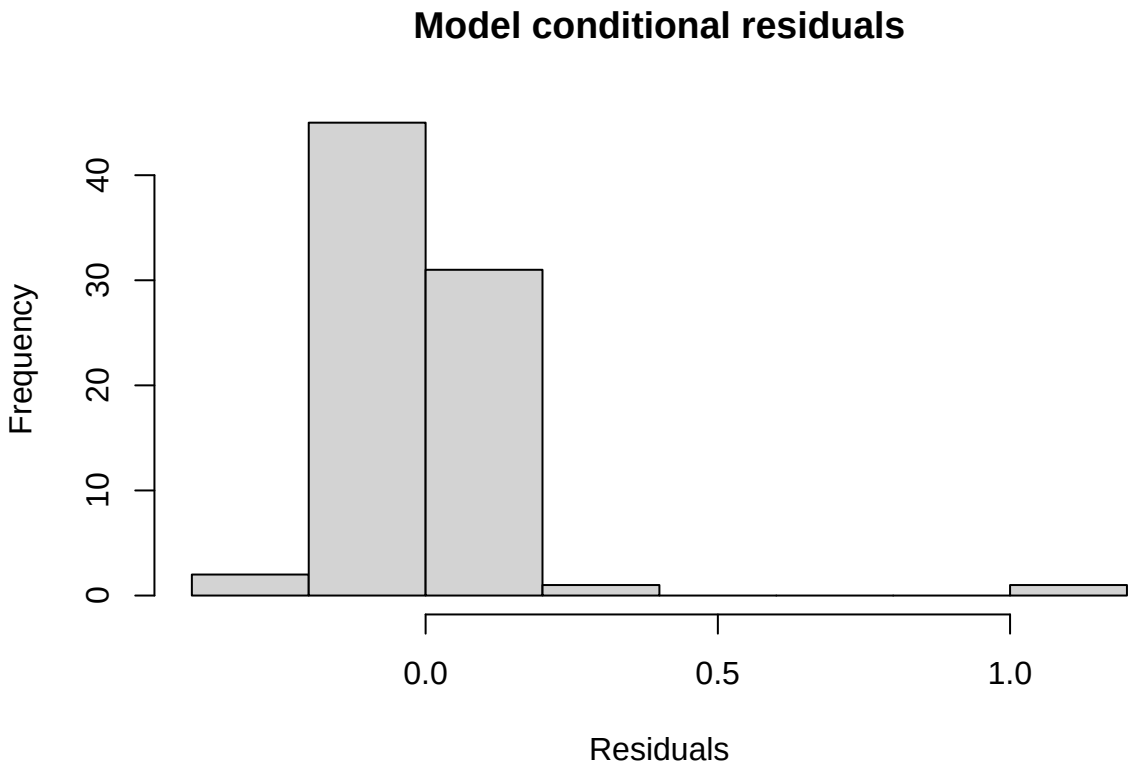


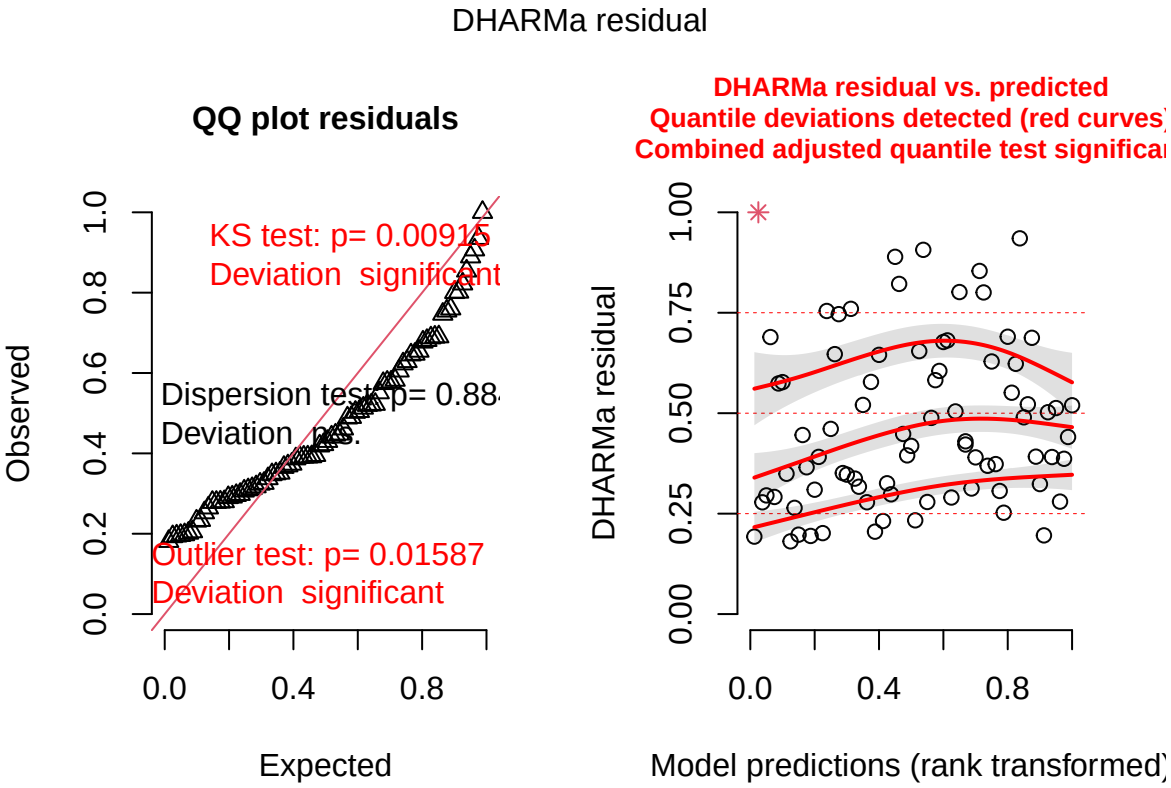
6.4 Change in the amount of compounds characteristic of *F. fusca*

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: normalized_mass_part ~ mean_delta + (1 | colony)
##    Data: separation_data
##
## REML criterion at convergence: -25.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5589 -0.4676 -0.0943  0.2358  6.8831
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  colony   (Intercept) 0.02205  0.1485
##  Residual             0.02765  0.1663
## Number of obs: 80, groups:  colony, 11
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  2.240e-01  5.346e-02  1.383e+01    4.19 0.000932 ***
## mean_delta  -4.298e-05  1.443e-03  6.870e+01   -0.03 0.976326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##            (Intr)
## mean_delta -0.408
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm_model)
## W = 0.59858, p-value = 1.852e-13
```





7 Impact of slaves on the development of the *F. sanguinea* CHC profile

We analyzed the impact of *F. fusca* slaves on the CHC profile of callow *F. sanguinea* ants. In this experiment, *F. sanguinea* ants were isolated in pairs before eclosion, and their CHC profiles were analyzed at one of four time points marking their adult age. We then tested whether callow *F. sanguinea* ants were chemically more similar to their slave relatives from free-living colonies. Unrelated *F. fusca* ants served as a background group."

7.1 Analysis based on all samples

Distribution of p-values for the separation period of 1-3 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 1.907e-06 1.907e-06 3.624e-05 1.011e-03 3.948e-04 8.255e-02
```

Distribution of p-values for the separation period of 8-10 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 3.815e-06 3.815e-06 2.670e-05 3.637e-04 1.259e-04 2.299e-02
```

Distribution of p-values for the separation period of 17-20 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 3.052e-05 6.104e-05 1.312e-03 1.199e-02 5.157e-03 7.820e-01
```

Distribution of p-values for the separation period of 35-40 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 6.104e-05 1.807e-02 1.514e-01 2.786e-01 4.887e-01 1.000e+00
```

7.2 Analysis restricted to individuals hatched from cocoons

Distribution of p-values for the separation period of 1-3 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 1.907e-06 1.907e-06 3.624e-05 1.010e-03 3.948e-04 8.255e-02
```

Distribution of p-values for the separation period of 8-10 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 3.815e-06 3.815e-06 2.670e-05 3.638e-04 1.259e-04 2.299e-02
```

Distribution of p-values for the separation period of 17-20 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 3.052e-05 6.104e-05 1.312e-03 1.198e-02 5.157e-03 7.820e-01
```

Distribution of p-values for the separation period of 35-40 days:

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 6.104e-05 1.807e-02 1.514e-01 2.784e-01 4.887e-01 1.000e+00
```

8 Effect of dummy ants

In these experiments, *F. sanguinea* ants released from their cocoon envelopes were maintained in Petri dishes along with glass beads coated with CHC extracted from either *F. fusca* or *F. sanguinea* ants. We examined whether the contact with dummy nestmates influenced the development of CHC profile of young *F. sanguinea* workers. In particular, we aimed to determine there there is evidence for active chemical mimicry in *F. sanguinea*.

8.1 Distance to the CHC profile of the dummy ants treated with *F. fusca* CHC

We computed the chemical distance to the CHC profile of ants whose CHC were used to cover glass beads serving as dummy ants. In the control variant, we used the ants subjected to clean glass beads.

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  to_test$mean_dist.x and to_test$mean_dist.y
## V = 18.5, p-value = 0.6781
## alternative hypothesis: true location shift is not equal to 0
```

Table S3 : Chemical distance of separated *F. sanguinea* ants to the CHC profile of *F. fusca* ants from colonies that served as a source of CHC to caot the glass beads. In control variant, glass bead were left clean.

Colony ID	Treamtent	Averaged distance	Contrast	Averaged distance
SD18-3	12-15 days + F. fusca hydrocarbons	0.396	12-15 days (control)	0.509
SD18-5	12-15 days + F. fusca hydrocarbons	0.616	12-15 days (control)	0.513
SD18-5	12-15 days + F. fusca hydrocarbons	0.842	12-15 days (control)	0.832
SD19-11	12-15 days + F. fusca hydrocarbons	0.467	12-15 days (control)	0.469
SD19-3	12-15 days + F. fusca hydrocarbons	0.500	12-15 days (control)	0.436
SD19-4	12-15 days + F. fusca hydrocarbons	0.305	12-15 days (control)	0.424
SD19-6	12-15 days + F. fusca hydrocarbons	0.674	12-15 days (control)	0.699
SD19-8	12-15 days + F. fusca hydrocarbons	0.595	12-15 days (control)	0.587
SD20-2	12-15 days + F. fusca hydrocarbons	0.699	12-15 days (control)	0.709

Table S4 : Chemical distance of separated *F. sanguinea* ants to the CHC profile of *F. sanguinea* ants from colonies that served as a source of CHC to caot the glass beads. In control variant, glass bead were left clean.

Colony ID	Treamtent	Averaged distance	Contrast	Averaged distance
SD18-2	12-15 days + F. sanguinea hydrocarbons	0.571	12-15 days (control)	0.373
SD18-3	12-15 days + F. sanguinea hydrocarbons	0.457	12-15 days (control)	0.599
SD18-5	12-15 days + F. sanguinea hydrocarbons	0.525	12-15 days (control)	0.580
SD18-5	12-15 days + F. sanguinea hydrocarbons	0.250	12-15 days (control)	0.232
SD19-11	12-15 days + F. sanguinea hydrocarbons	0.595	12-15 days (control)	0.607
SD19-3	12-15 days + F. sanguinea hydrocarbons	0.419	12-15 days (control)	0.511
SD19-4	12-15 days + F. sanguinea hydrocarbons	0.648	12-15 days (control)	0.668
SD19-6	12-15 days + F. sanguinea hydrocarbons	0.553	12-15 days (control)	0.591
SD19-8	12-15 days + F. sanguinea hydrocarbons	0.850	12-15 days (control)	0.850
SD20-2	12-15 days + F. sanguinea hydrocarbons	0.646	12-15 days (control)	0.654

8.2 Distance to the CHC profile of the dummy ants treated with *F. sanguinea* CHC

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  to_test$mean_dist.x and to_test$mean_dist.y
## V = 12, p-value = 0.2361
## alternative hypothesis: true location shift is not equal to 0
```

8.3 Fraction of *n*-docosane in the CHC profile

For the control treatment, glass beads were left uncoated with CHC. In the experimental treatments, the CHC used to coat the glass beads were supplemented with *n*-docosane, which served as an internal standard. We examined differences in the proportion of *n*-docosane in the CHC profiles of the tested ants, as these could indicate the acquisition of chemicals from the dummy ants.

Table S5 : Proportion of *n*-docosane in CHC extracted from *F. sanguinea* ants maintained with the glass beads coated with the CHC of *F. fusca* ants and contaminated with *n*-docosane. In control variant, glass bead were left clean.

Colony ID	Treamtent	Averaged C22 fraction	Contrast	Averaged C22 fraction
SD18-3	12-15 days + F. fusca hydrocarbons	0.018	12-15 days (control)	0.000
SD18-5	12-15 days + F. fusca hydrocarbons	0.001	12-15 days (control)	0.000
SD18-5	12-15 days + F. fusca hydrocarbons	0.002	12-15 days (control)	0.000
SD19-11	12-15 days + F. fusca hydrocarbons	0.011	12-15 days (control)	0.002
SD19-3	12-15 days + F. fusca hydrocarbons	0.009	12-15 days (control)	0.001
SD19-4	12-15 days + F. fusca hydrocarbons	0.017	12-15 days (control)	0.000
SD19-6	12-15 days + F. fusca hydrocarbons	0.009	12-15 days (control)	0.001
SD19-8	12-15 days + F. fusca hydrocarbons	0.003	12-15 days (control)	0.002
SD20-2	12-15 days + F. fusca hydrocarbons	0.009	12-15 days (control)	0.005
W17-1	12-15 days + F. fusca hydrocarbons	0.021	12-15 days (control)	0.001

8.3.1 *F. fusca*

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  to_test$C22_prop.x and to_test$C22_prop.y
## V = 55, p-value = 0.005857
## alternative hypothesis: true location shift is not equal to 0
```

8.3.2 *F. sanguinea*

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  to_test$C22_prop.x and to_test$C22_prop.y
## V = 54, p-value = 0.007093
## alternative hypothesis: true location shift is not equal to 0
```

Table S6 : Proportion of *n*-docosane in CHC extracted from *F. sanguinea* ants maintained with the glass beads coated with the CHC of *F. sanguinea* ants and contaminated with *n*-docosane. In control variant, glass bead were left clean.

Colony ID	Treamtent	Averaged C22 fraction	Contrast	Averaged C22 fraction
SD18-2	12-15 days + F. sanguinea hydrocarbons	0.005	12-15 days (control)	0.002
SD18-3	12-15 days + F. sanguinea hydrocarbons	0.014	12-15 days (control)	0.000
SD18-5	12-15 days + F. sanguinea hydrocarbons	0.003	12-15 days (control)	0.000
SD18-5	12-15 days + F. sanguinea hydrocarbons	0.002	12-15 days (control)	0.000
SD19-11	12-15 days + F. sanguinea hydrocarbons	0.004	12-15 days (control)	0.002
SD19-3	12-15 days + F. sanguinea hydrocarbons	0.003	12-15 days (control)	0.001
SD19-4	12-15 days + F. sanguinea hydrocarbons	0.002	12-15 days (control)	0.000
SD19-6	12-15 days + F. sanguinea hydrocarbons	0.004	12-15 days (control)	0.001

SD19-8	12-15 days + F. sanguinea hydrocarbons	0.002	12-15 days (control)	0.002
SD20-2	12-15 days + F. sanguinea hydrocarbons	0.004	12-15 days (control)	0.005
W17-1	12-15 days + F. sanguinea hydrocarbons	0.003	12-15 days (control)	0.001

9 Body surface area of *F. sanguinea* ants and their slaves

The areas of planar projection of different body parts were regressed on the head width to obtain a model which was subsequently used to approximate the area for each sample. The regression was performed separately for each species. The following formula was applied:

$$\hat{A} = \sum_{i \in I} \beta_{0,i} + \beta_{1,i} w + \beta_{2,i} w^2,$$

where
$\hat{A}$ denotes the estimated body area,
I denotes the set of body parts for which the planar projection area was measured
w denotes head width
$\beta_{j,i}$ are coefficients from the polynomial regression.

These calculations do not reflect the absolute value of body surface area but were converted into scaling factors by using them a divisors of a standard area. In this way, the unit of measurement is canceled out, and absolute values are converted into dimensionless ratios. The standard area was calculated by substituting the $w = 1.15$ mm into the formula above and using coefficients obtained from measurements of *F. sanguinea*.

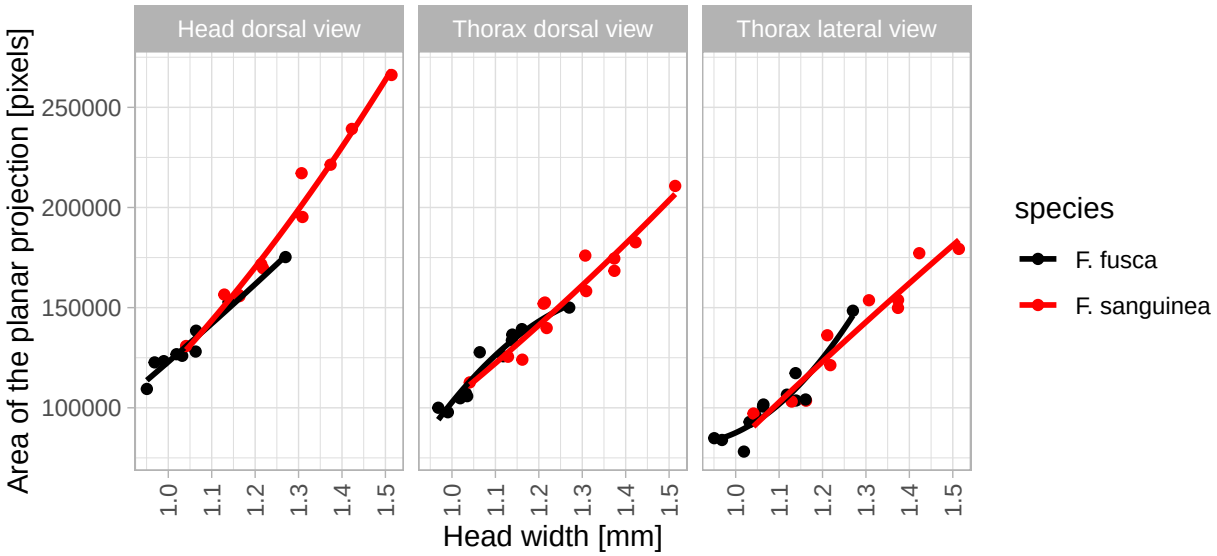


Figure S9 : Relation between head width and area of the planar projections of ant body parts.

References

Florian Hartig. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*, 2024. URL http://florianhartig.github.io/DHARMa/. R package version 0.4.7.

Kim-Anh Le Cao, Florian Rohart, Ignacio Gonzalez, and Sebastien Dejean. *mixOmics: Omics Data Integration Project*, 2025. URL https://bioconductor.org/packages/mixOmics. R package version 6.32.0.

Florian Rohart, Benoît Gautier, Amrit Singh, and Kim-Anh Lê Cao. mixomics: An r package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, 13(11):1–19, 11 2017. doi: 10.1371/journal.pcbi.1005752. URL https://doi.org/10.1371/journal.pcbi.1005752.

Michel Tenenhaus. *La régression PLS: théorie et pratique*. Editions TECHNIP, Paris, 1998.

Tomasz Włodarczyk and Lech Szczepaniak. Facultative slave-making ants formica sanguinea label their slaves with own recognition cues instead of employing the strategy of chemical mimicry. *Journal of Insect Physiology*, 96:98–107, 2017. ISSN 0022-1910. doi: https://doi.org/10.1016/j.jinsphys.2016.10.016. URL https://www.sciencedirect.com/science/article/pii/S0022191016301640.