HAL open science

Data analysis: Tumors alter life-history traits in the freshwater cnidarian, Hydra oligactis

Justine Boutry, Sophie Tissot

To cite this version:

Justine Boutry, Sophie Tissot. Data analysis: Tumors alter life-history traits in the freshwater cnidarian, Hydra oligactis. 2022. hal-03647721

HAL Id: hal-03647721

https://hal.science/hal-03647721v1

Preprint submitted on 1 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Table S1 : The formulas of the different generalized linear mixed models constructed and the associated value of AICc's weights, related to the analysis of the traits related to asexual reproduction (i.e. budding rate, budding type, and gap between two reproductive episodes)

First, there is the results for the random effect selection and then for the fixed effects selection. NC indicates that the model has not converged.

Model	Formulas	Budding rate	Budding type	Gap between two reproductive episode
1.1	Trait ~ Status * Phase + (1|individual)	0.553	0.489	0.507
1.2	Trait ~ Status * Phase + (1|birthday)	0.000	0.511	0.212
1.3	Trait ~ Status * Phase + (1|individual)+(1|birthday)	0.447	NC	0.281
1.4	Trait ~ Status * Phase	NC	NC	NC
2.1	Trait ~ Status * Phase + Random effect	0.994	0.158	0.397
2.2	Trait ~ Status + Phase + Random effect	0.006	0.332	0.603
2.3	Trait ~ Status + Random effect	0.000	0.000	0.000
2.4	Trait ~ Phase + Random effect	0.000	0.510	0.000
2.5	Trait ~1 + Random effect	0.000	0.000	0.000

Table S2 : The formulas of the different generalized linear mixed models constructed during the analysis of the tumoral transmission and the associated value of AICc's weights, related to the analysis of tumor transmission rate.

First, there is the results for the random effect selection and then for the fixed effects selection.

Model	Formulas	AICc's weight
1.1	Bud status ~ Production phase + Budding type + (1|Complet batch) + (1|Survey date)	0.125
1.2	Bud status ~ Production phase + Budding type+ (1|Complet batch)	0.375
1.3	Bud status ~ Production phase + Budding type + (1|Survey date)	0.375
1.4	Bud status ~ Production phase + Budding type (1|Survey date/Complet batch)	0.125
1.5	Bud status ~ Production phase + Budding type	NC
2.1	Bud status ~ Production phase * Budding type + (1|Complet batch)	0.034
2.2	Bud status ~ Production phase + Budding type + (1|Complet batch)	0.091
2.3	Bud status ~ Production phase + (1|Complet batch)	0.207
2.4	Bud status ~ Budding type + (1|Complet batch)	0.209
2.5	Bud status ~ 1 + (1|Complet batch)	0.459

Table S3 : The formulas of the different generalized linear mixed models constructed during the analysis of bud's survival and the associated value of AICc's weights, related to the analysis of bud survival.

First, there is the results for the random effect selection and then for the fixed effects selection.

Model	Formulas	AICc's weight
1.1	Survival ~ Status * Budding type * Production phase + (1|Survey date)+(1|Complet batch)	0.387
1.2	Survival ~ Status * Budding type * Production phase+(1|Complet batch)	0.226
1.3	Survival ~ Status * Budding type * Production phase + (1|Survey date)	0.000
1.4	Survival ~ Status * Budding type * Production phase + (1|Survey date/Complet batch)	0.387
1.5	Survival ~ Status * Budding type * Production phase	0.000
2.1	Survival ~ Status * Budding type * Production phase + (1|Survey date/Complet batch)	0.003
2.2	Survival ~ Budding type * Production phase + (1|Survey date/Complet batch)	0.074
2.3	Survival ~ Status * Budding type + (1|Survey date/Complet batch)	0.001
2.4	Survival ~ Status * Production phase + (1|Survey date/Complet batch)	0.055
2.5	Survival ~ Status + Budding type + Production phase + (1|Survey date/Complet batch)	0.160
2.6	Survival ~ Status + Production phase * Budding type + (1|Survey date/Complet batch)	0.062
2.7	Survival ~ Status * Production phase + Budding type + (1|Survey date/Complet batch)	0.063
2.8	Survival ~ Status * Budding type + Production phase + (1|Survey date/Complet batch)	0.057
2.9	Survival ~ Budding type + Production phase + (1|Survey date/Complet batch)	0.171
2.10	Survival ~ Status + Budding type + (1|Survey date/Complet batch)	0.002
2.11	Survival ~ Status + Production phase + (1|Survey date/Complet batch)	0.133
2.12	Survival ~ Production phase + (1|Survey date/Complet batch)	0.212
2.13	Survival ~ Budding type + (1|Survey date/Complet batch)	0.003
2.14	Survival ~ Status + (1|Survey date/Complet batch)	0.001
2.15	Survival ~ 1 + (1|Survey date/Complet batch)	0.003

Table S4 : The formulas of the different linear mixed models and generalized linear mixed models constructed during the first sexual reproduction analysis and the analysis of the number of eggs produced, and the associated value of AICc's weight, related to the analysis of the timing of sexual reproduction.

First, there is the results for the random effect selection and then for the fixed effects selection.

Model	Formulas	First reproduction	Eggs number
1.1	Trait ~ Status * Phase + (1|Cold placement date)	0.996	0.998
1.2	Trait ~ Status * Phase	0.004	0.002
2.1	Trait ~ Status * Phase + (1|Cold placement date)	0.734	0.418
2.2	Trait ~ Status + Phase + (1|Cold placement date)	0.258	0.043
2.3	Trait ~ Status + (1|Cold placement date)	0.000	0.114
2.4	Trait ~ Phase + (1|Cold placement date)	0.008	0.116
2.5	Trait ~ 1 + (1|Cold placement date)	0.000	0.308

Code S1 : Code of the analysis of the traits related to asexual reproduction, tumor transmission rate, and bud survival, related to STAR Methods.

Tumors alter life-history traits in the freshwater cnidarian, Hydra oligactis

Justine BOUTRY and Sophie TISSOT

R Packages loading

```
library(ade4)
library(car)
library(DHARMa)
library(factoextra)
library(ggplot2)
library(ggridges)
library(glmmTMB)
library(lme4)
library(MASS)
library(MuMIn)
library(plyr)
library(readr)
library(RVAideMemoire)
library(sjPlot)
library(survival)
library(survminer)
library(vegan)
```

Data formatting

The dataframe *all* (corresponding to *asexual_reproduction_indivudual_following.csv*) contains the detail of asexual reproduction's following for each individual. *data_sum* (corresponding to *asexual_reproduction_summary.csv*) is the summary for each variable per individual of the asexual reproduction's following. *surv* (corresponding to *asexual_reproduction_survival.csv*) contains the survival's data. *transmission* (corresponding to *asexual_reproduction_tumoral_transmission.csv*) gather the data about the tumoral's transmission during the budding.

```
##      Status       ID_parent      Parent_birthday     Survey_date
##  Control: 993   HC01   :  28   07/09/2020:840    02/10/2020:  72
##  Tumor  :1008   HC02   :  28   09/09/2020:784    02/11/2020:  72
##                 HC03   :  28   14/09/2020:377    04/11/2020:  72
##                 HC04   :  28                     05/10/2020:  72
##                 HC05   :  28                     06/11/2020:  72
##                 HC06   :  28                     07/10/2020:  72
##                 (Other):1833                     (Other)   :1569
##  ID_reproductive_episode     Delta        Nb_simultanous_buds  Cum_Nb_buds
##  3      :  73            Min.   : 2.000   Min.   :1.000       Min.   : 0.000
##  1      :  72            1st Qu.: 2.000   1st Qu.:1.000       1st Qu.: 2.000
##  2      :  72            Median : 3.000   Median :1.000       Median : 7.000
##  4      :  72            Mean   : 4.016   Mean   :1.274       Mean   : 7.292
##  5      :  69            3rd Qu.: 5.000   3rd Qu.:1.000       3rd Qu.:11.000
##  (Other): 392            Max.   :18.000   Max.   :5.000       Max.   :23.000
##  NA's   :1251            NA's   :1323     NA's   :1250
##       Time           Week             Phase
##  Min.   : 0.0   2      :216   Week_0_to_4: 863
##  1st Qu.:14.0   3      :216   Week_5_to_9:1138
##  Median :30.0   4      :216
##  Mean   :31.1   5      :216
##  3rd Qu.:47.0   6      :216
##  Max.   :63.0   7      :216
##                 (Other):705
```

```
##    ID_parent           Phase        Status     Parent_birthday    Nb_buds
##  HC01   :  2   Week_0_to_4:72   Control:72   07/09/2020:59     Min.   : 2.000
##  HC02   :  2   Week_5_to_9:69   Tumor  :69   09/09/2020:54     1st Qu.: 5.000
##  HC03   :  2                                 14/09/2020:28     Median : 6.000
##  HC04   :  2                                                   Mean   : 6.787
##  HC05   :  2                                                   3rd Qu.: 8.000
##  HC06   :  2                                                   Max.   :16.000
##  (Other):129
##      Delay      Mean_simult_buds   Mean_Delta      Stand_nb_bud
##  Min.   : 5.0   Min.   :1.000    Min.   : 2.250   Min.   : 6.00
##  1st Qu.:16.0   1st Qu.:1.000    1st Qu.: 3.200   1st Qu.:21.00
##  Median :23.0   Median :1.222    Median : 3.833   Median :31.00
##  Mean   :24.4   Mean   :1.264    Mean   : 4.325   Mean   :32.09
##  3rd Qu.:35.0   3rd Qu.:1.400    3rd Qu.: 4.667   3rd Qu.:44.00
##  Max.   :35.0   Max.   :2.000    Max.   :14.000   Max.   :73.00
##
```

```
##   ID_parent            Status          Parent_birthday
##  Length:62          Length:62          Min.   :2020-09-07
##  Class :character   Class :character   1st Qu.:2020-09-07
##  Mode  :character   Mode  :character   Median :2020-09-09
##                                        Mean   :2020-09-09
##                                        3rd Qu.:2020-09-12
##                                        Max.   :2020-09-14
##
##    Death_date         Spont_Tum_date       Spont_tum_event    Censored
##  Min.   :2020-10-20   Min.   :2021-01-12   Min.   :0.00    Min.   :0.0000
##  1st Qu.:2020-11-25   1st Qu.:2021-01-12   1st Qu.:0.00    1st Qu.:0.0000
##  Median :2020-12-24   Median :2021-01-22   Median :0.00    Median :1.0000
##  Mean   :2020-12-24   Mean   :2021-01-23   Mean   :0.25    Mean   :0.5645
##  3rd Qu.:2021-01-25   3rd Qu.:2021-01-27   3rd Qu.:0.25    3rd Qu.:1.0000
##  Max.   :2021-02-24   Max.   :2021-02-18   Max.   :1.00    Max.   :1.0000
##  NA's   :27           NA's   :54           NA's   :30
##  Last_check_date
##  Min.   :2021-03-10
##  1st Qu.:2021-03-10
##  Median :2021-03-10
##  Mean   :2021-03-10
##  3rd Qu.:2021-03-10
##  Max.   :2021-03-10
##
```

```
##      Status       Production_phase     Batch        Bud_type       Bud_birthday
##  Control:168   Week_0_to_4:204     7      :77   Multiple:222   04/03/2021: 13
##  Tumor  :205   Week_5_to_9:169     8      :56   Simple  :151   11/03/2021: 10
##                                    6      :54                  18/03/2021: 31
##                                    3      :47                  NA's      :319
##                                    4      :43
##                                    1      :38
##                                    (Other):58
##      Survey_date    Bud_status                Complet_batch  Survival
##  13/07/2021:208   Control:128   Tumor_Week_0_to_4_7  : 49   Alive:216
##  08/07/2021: 43   Dead   :157   Tumor_Week_0_to_4_6  : 19   Dead :157
##  22/04/2021: 31   Tumor  : 88   Tumor_Week_5_to_9_4  : 18
##  05/08/2021: 23                 Tumor_Week_0_to_4_8  : 17
##  29/07/2021: 21                 Control_Week_0_to_4_5: 16
##  (Other)   : 46                 Control_Week_0_to_4_7: 16
##  NA's      :  1                 (Other)              :238
```

Lexical

- *Status* : the strain from which the individual comes from (tumorous or healthy)
- *ID_parent* : the name of each individual
- *Parent_birthday* : the date of birth of the individual
- *Survey_date* : the date when the survey has been done
- *ID_reproductive_episode* : the number attributed to each reproductive episode
- *Delta* : the number of day between two reproductive episode
- *Nb_simultanous_buds* : the number of bud produce by each individual at each reproductive episode

- *Cum_nb_buds* : the cumulative sum of the number of buds for each individual
- *Time* : the number of day between the birth and the survey date
- *Week* : the number of week past since the birthday of the individual
- *Phase* : the age of the individual when it has been placed in the sexual reproduction conditions
- *Nb_buds* : the number of bud produce by each individual for each phase
- *Delay* : the delay (in days) of each reproductive phase for each individual (considering that the first phase start at the production of the first bud)
- *Mean_simult_buds* : the mean number of bud produce by each individual at a reproductive episode per each phase
- *Mean_Delta* : the mean number of day between two reproductive episode per each phase
- *Stand_nb_bud* : the mean number of bud produce by each individual for each phase on 100 days (calculated by correcting the total number of bud per phase by its delay and multiplied by 100)
- *Time* : the number of day between the birth and the survey date
- *Week* : the number of week past since the birthday of the individual
- *Death_date* : the date of each individual death
- *Spont_Tum_date* : the date of apparition of spontaneous tumors on control individuals
- *Spont_tum_event* : the observation of spontaneous tumor on control hydra during the experiment
- *Censored* : the observation the individual death during the experiment
- *Last_check_date* : the date when the last survival measurement had take place
- *Production_phase* : the parent's age-class when the bud was produced
- *Batch* : the batch number of the parent
- *Bud_type* : the type (simple or multiple) of budding episod which from the bud is originated
- *Bud_birthday* : the day when the bud was detach
- *Bud_status* : the status of the buds one month after their birth
- *Complet_batch* : the name of batch composed by the status, the phase and the batch number
- *Survival* : the individual's condition (alive or dead)

Analysis of the age at the first asexual reproduction

We start by isolating all the data concerning the first reproductive episode, then a Wilcoxon's and a Kolmorogov-Smirnov test are applied on the number of day past before the first reproduction between the heathly and tumorous hydras.

```
first_repro<-as.data.frame(all[all$ID_reproductive_episode=="1",] )
wilcox.test(first_repro[first_repro$Status=="Control","Time"],first_repro[first_repro$Status=
="Tumor","Time"])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  first_repro[first_repro$Status == "Control", "Time"] and first_repro[first_repro$St
atus == "Tumor", "Time"]
## W = 525, p-value = 0.1543
## alternative hypothesis: true location shift is not equal to 0
```

```
ks.test(first_repro[first_repro$Status=="Control","Time"],first_repro[first_repro$Status=="Tu
mor","Time"])
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  first_repro[first_repro$Status == "Control", "Time"] and first_repro[first_repro$St
atus == "Tumor", "Time"]
## D = 0.22222, p-value = 0.3364
## alternative hypothesis: two-sided
```

Theses tests are not significant, we standardized the variable by dividing by the maximal number of days before the first reproduction for each status. Then we reapply the same tests.

```
fr_ctrl<-subset(first_repro,Status=="Control")
fr_tum<-subset(first_repro,Status=="Tumor")

nctrl<-(fr_ctrl$Time/max(fr_ctrl$Time))
ntum<-(fr_tum$Time/max(fr_tum$Time))

ks.test(nctrl,ntum)
```

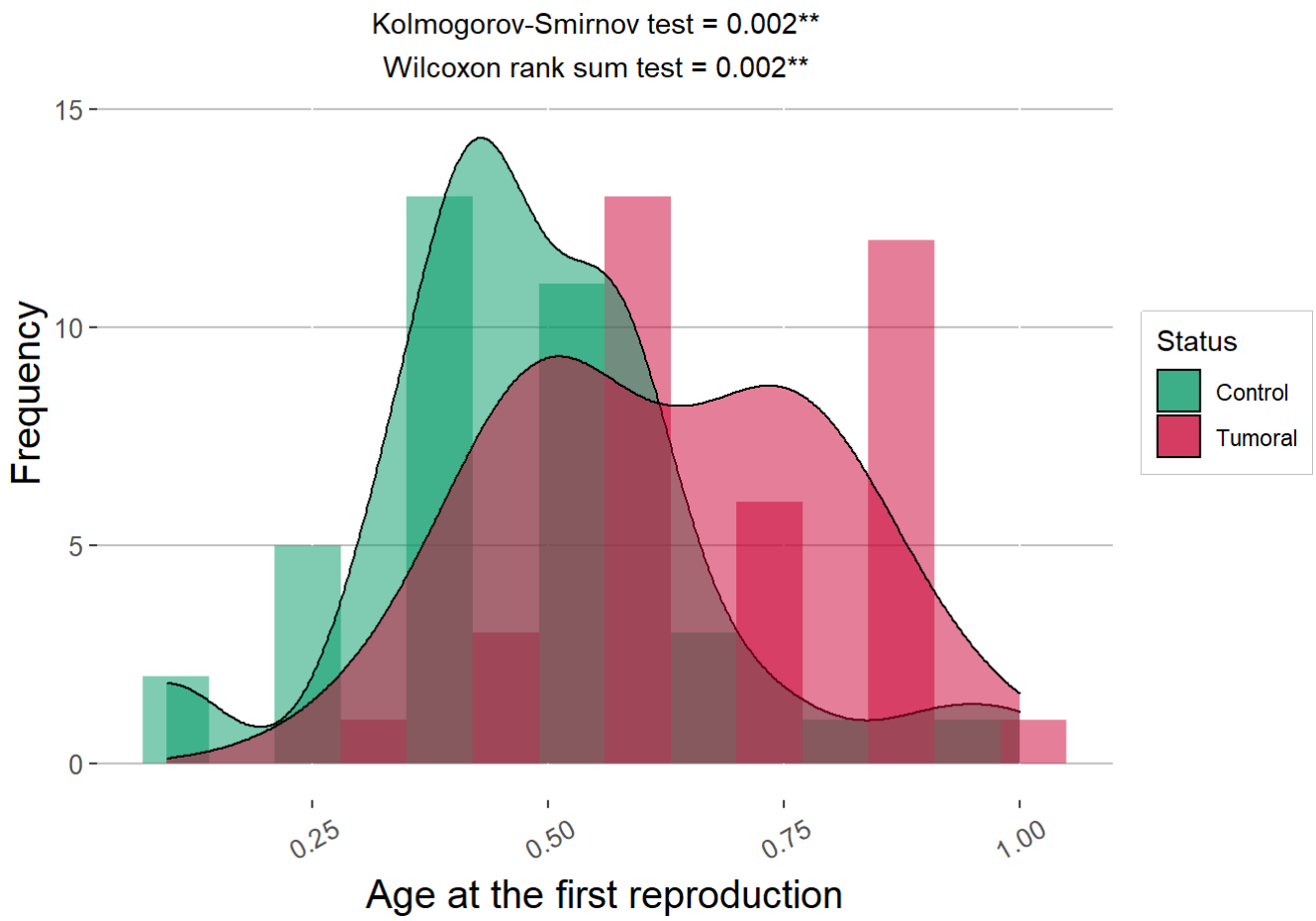
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  nctrl and ntum
## D = 0.44444, p-value = 0.001632
## alternative hypothesis: two-sided
```

```
wilcox.test(nctrl,ntum)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  nctrl and ntum
## W = 377.5, p-value = 0.00215
## alternative hypothesis: true location shift is not equal to 0
```

The standardized distributions of the first day of reproduction are a significantly different in both groups.

Graphical representation



Budding rate analysis

We start by testing if the random effects of the birthday of the individual and/or the individual are useful to explain the variable. Given that the b4 model do not converge, it has been excluded.

```
b1<-glmmTMB(Stand_nb_bud~Phase*Status+(1|ID_parent),REML = TRUE, family=poisson(log),data = d
ata_sum)
b2<-glmmTMB(Stand_nb_bud~Phase*Status+(1|Parent_birthday),REML = TRUE, family=poisson(log),da
ta = data_sum)
b3<-glmmTMB(Stand_nb_bud~Phase*Status+(1|ID_parent)+(1|Parent_birthday),REML = TRUE, family=p
oisson(log),data = data_sum)
b4<-glmmTMB(Stand_nb_bud~Phase*Status,REML = TRUE, family=poisson(log),data = data_sum)
```

```
##  model weights
## [1] 0.553 0.000 0.447
```

The weights of AICc suggest that the first model with the random effect of the individual best fitted the data than the other ones.

Then, we selected now the fixed effects to include.

```
b1_1<-glmmTMB(Stand_nb_bud~Phase*Status+(1|ID_parent),REML = FALSE,family=poisson(log),data =
data_sum)
b1_2<-glmmTMB(Stand_nb_bud~Phase+Status+(1|ID_parent),REML = FALSE,family=poisson(log),data =
data_sum)
b1_3<-glmmTMB(Stand_nb_bud~Phase+(1|ID_parent),REML = FALSE,family=poisson(log),data = data_s
um)
b1_4<-glmmTMB(Stand_nb_bud~Status+(1|ID_parent),REML = FALSE,family=poisson(log),data = data_
sum)
b1_5<-glmmTMB(Stand_nb_bud~1+(1|ID_parent),REML = FALSE,family=poisson(log),data = data_sum)
```

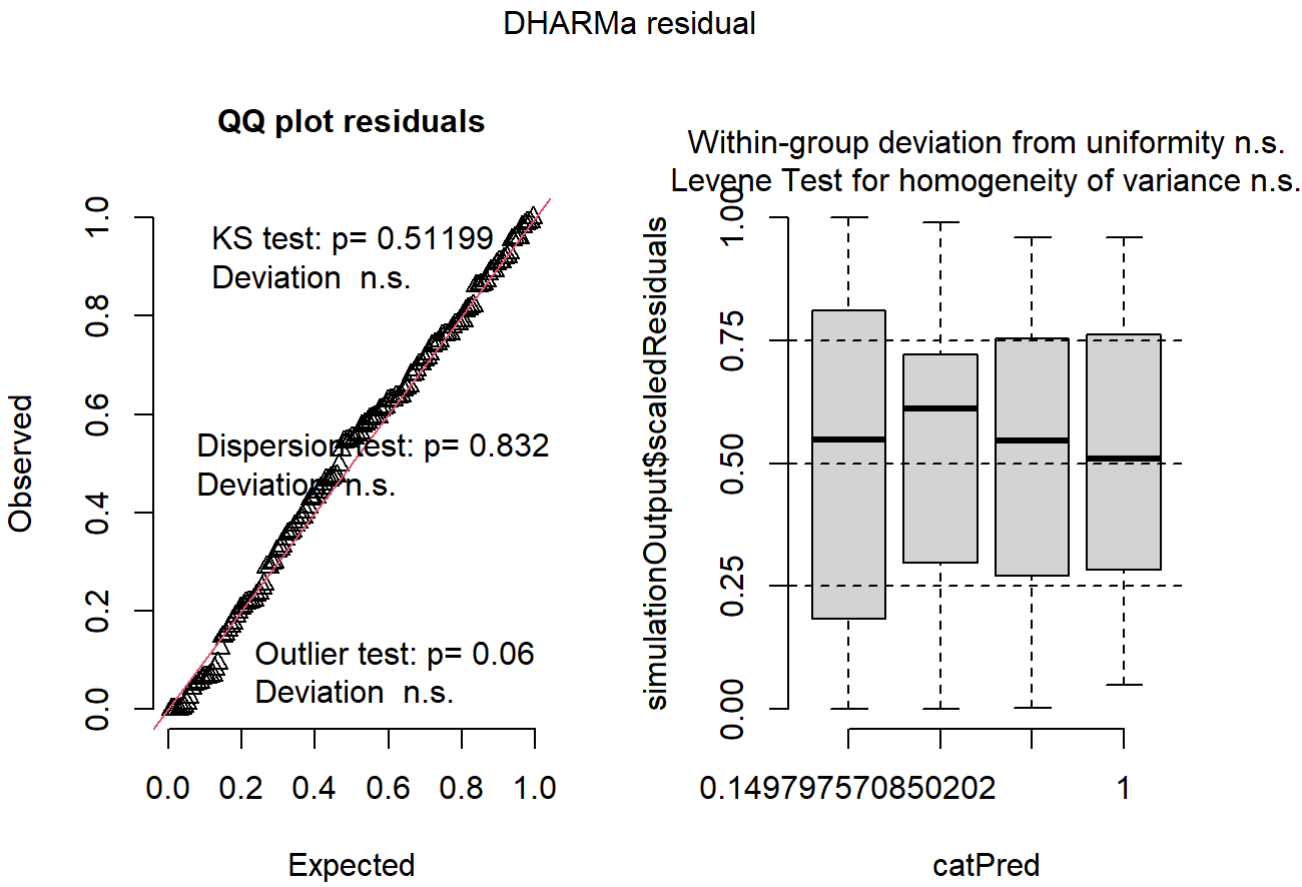
```
##  model weights
## [1] 0.994 0.006 0.000 0.000 0.000
```

The fixed effect of the interaction between the *Status* and the *Phase* seems to be the ones allowing the better explanation of the variable.

The final model take into account the fixed of the interaction between the *Status* and the *Phase*, and the random effect of the individual.

```
b1_1<-glmmTMB(Stand_nb_bud~Phase*Status+(1|ID_parent),REML = TRUE,family=poisson(log),data =
 data_sum)
```

Finally, model's residuals are checked to see if the model is equilibrated.



Overdispersion, Kolmogorov-Smirnov and outlier test end all correct.

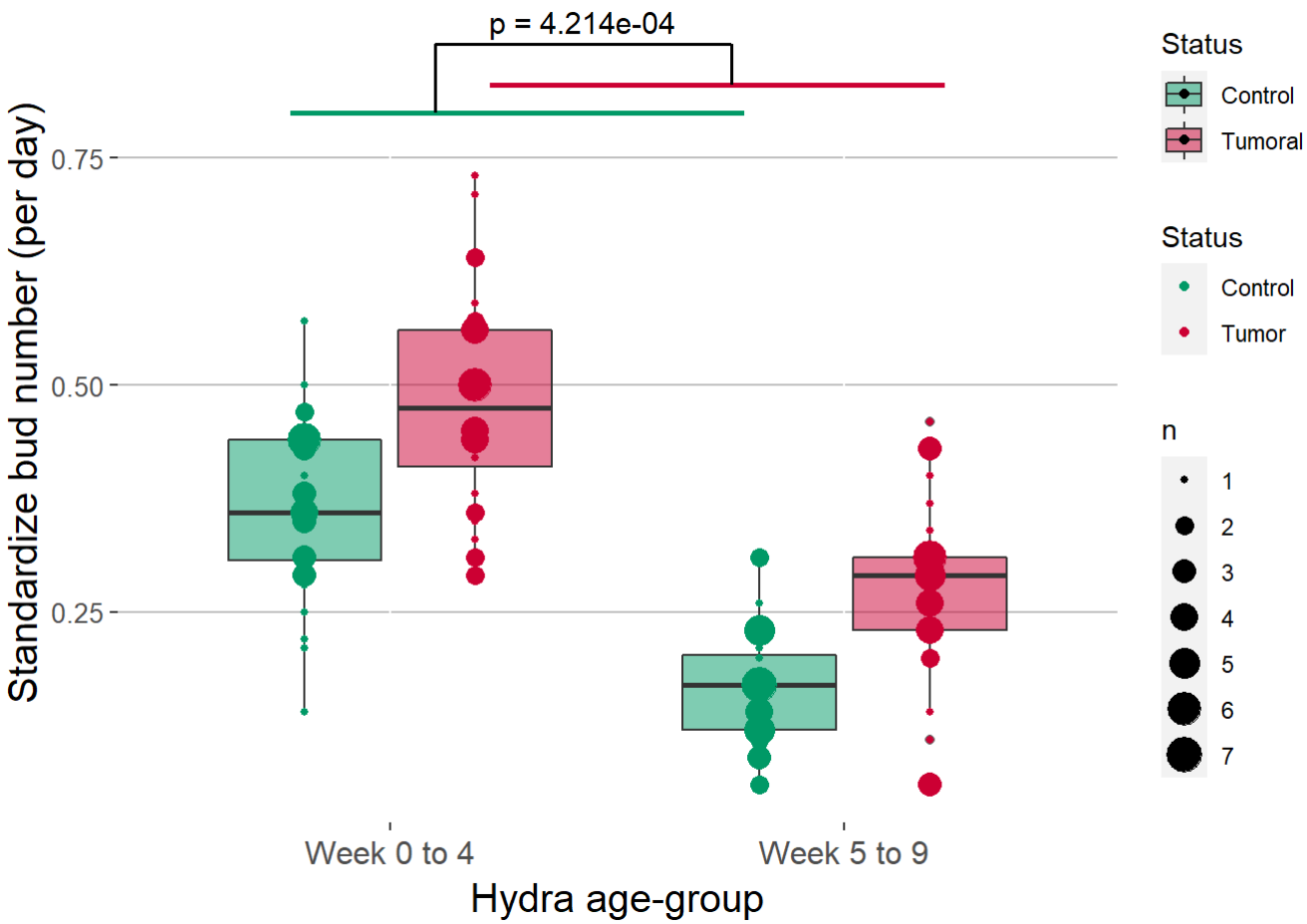
Final result

Number of buds per day ~ Status * Phase

Predictors	*Incidence Rate Ratios*	*std. Error*	*p*
Passage of the 5th week	0.45	0.02	**6.523e-59**
Tumor status	1.30	0.08	**1.618e-05**
Tumor status following the 5th week	1.26	0.08	**4.214e-04**
Observations	141		

	Stand_nb_bud		
Predictors	*Incidence Rate Ratios*	*std. Error*	*p*
Phase [Week 5 to 9]	0.45	0.02	**<0.001**
Status [Tumor]	1.30	0.08	**<0.001**
Phase [Week 5 to 9] * Status [Tumor]	1.26	0.08	**<0.001**
Observations	141		

Graphical representation



Number of bud per reproductive episode analysis

We start by creating a column describing qualitatively the type of reproductive episode observed.

```
all$bud_type<-with(all,ifelse(Nb_simultanous_buds<2,"simple","multiple"))
all$bud_type<-as.factor(all$bud_type)
```

The random effect is selected. The third and fourth model do not converge, they have been excluded.

```
sim1<-glmmTMB(bud_type~Phase*Status+(1|ID_parent),REML = TRUE, family=binomial(logit),data =
 all)
sim2<-glmmTMB(bud_type~Phase*Status+(1|Parent_birthday),REML = TRUE, family=binomial(logit),d
ata = all)
sim3<-glmmTMB(bud_type~Phase*Status+(1|ID_parent)+(1|Parent_birthday),REML = TRUE, family=bin
omial(logit),data = all)
sim4<-glmmTMB(bud_type~Phase*Status,REML = TRUE, family=binomial(logit),data = all)
```

```
##  model weights
## [1] 0.489 0.511
```

The second model seems to be the best.

The fixed effects are selected.

```
sim2_1<-glmmTMB(bud_type~Phase*Status+(1|ID_parent),REML = FALSE,family=binomial(logit),data
 =all)
sim2_2<-glmmTMB(bud_type~Phase+Status+(1|ID_parent),REML = FALSE,family=binomial(logit),data
 =all)
sim2_3<-glmmTMB(bud_type~Phase+(1|ID_parent),REML = FALSE,family=binomial(logit),data =all)
sim2_4<-glmmTMB(bud_type~Status+(1|ID_parent),REML = FALSE,family=binomial(logit),data =all)
sim2_5<-glmmTMB(bud_type~1+(1|ID_parent),REML = FALSE,family=binomial(logit),data =all)
```

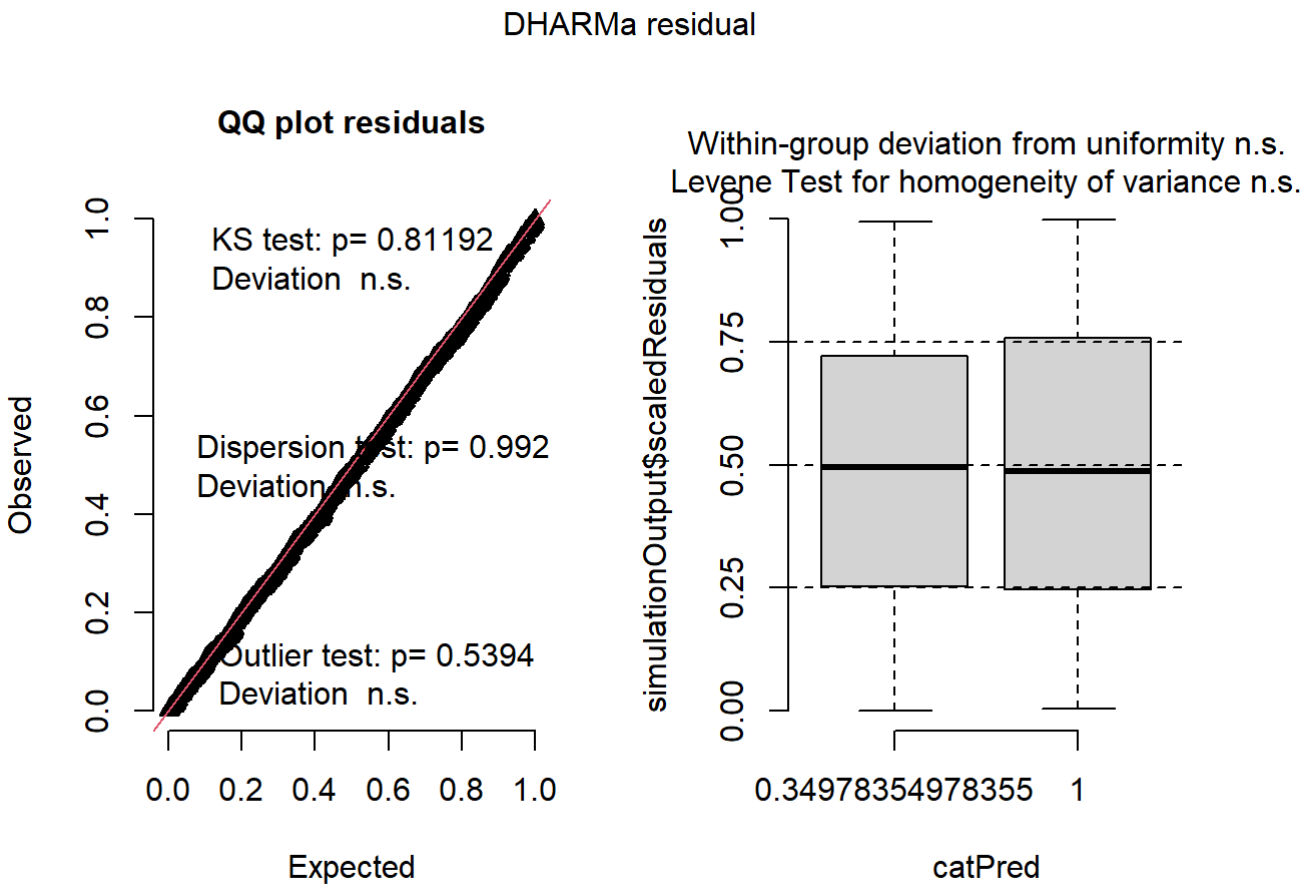
```
##  model weights
## [1] 0.158 0.332 0.000 0.510 0.000
```

The fourth model has the better AICc.

The final model take into account the fixed effect of the *Status* and the random effect of the individual.

```
sim2_4<-glmmTMB(bud_type~Status+(1|ID_parent),REML = TRUE,family=binomial(logit),data =all)
```

Finally, model's residuals are checked.



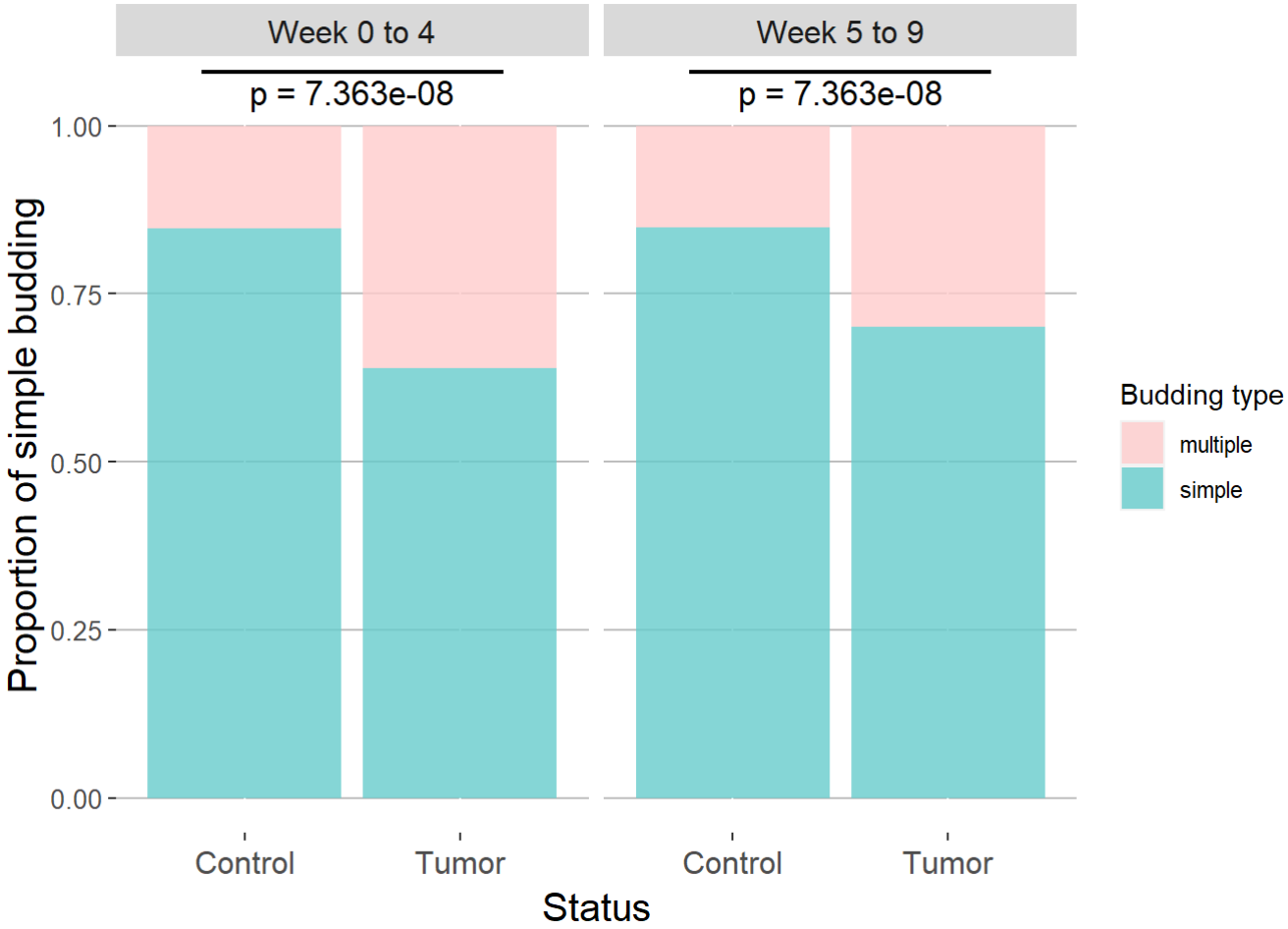
The model selected fit well the data.

Final result

Type of reproductive episode (simple/multiple) ~ Status

Predictors	*Odds Ratios*	*std. Error*	*p*
Tumorous status	0.37	0.07	**7.363e-08**
Observations	751		

Graphical representation



Analysis of the gap between two reproductive episode

The random effect is selected. The fourth model do not converge then have been excluded.

```
del1<-glmmTMB(Delta~Status*Phase+(1|ID_parent),REML = TRUE, family = poisson(log),data=all)
del2<-glmmTMB(Delta~Status*Phase+(1|Parent_birthday),REML = TRUE, family = poisson(log),data=
all)
del3<-glmmTMB(Delta~Status*Phase+(1|ID_parent)+(1|Parent_birthday),REML = TRUE, family = pois
son(log),data=all)
del4<-glmmTMB(Delta~Status*Phase,REML = TRUE, family = poisson(log),data=all)
```

```
##  model weights
## [1] 0.507 0.212 0.281
```

The first model seems to be the best.

The fixed effect is selected.

```
del1_1<-glmmTMB(Delta~Status*Phase+(1|ID_parent),REML = FALSE, family = poisson(log),data=al
l)
del1_2<-glmmTMB(Delta~Status+Phase+(1|ID_parent),REML = FALSE,family = poisson(log),data=all)
del1_3<-glmmTMB(Delta~Status+(1|ID_parent),REML = FALSE,family = poisson(log),data=all)
del1_4<-glmmTMB(Delta~Phase+(1|ID_parent),REML = FALSE,family = poisson(log),data=all)
del1_5<-glmmTMB(Delta~1+(1|ID_parent),REML = FALSE,family = poisson(log),data=all)
```

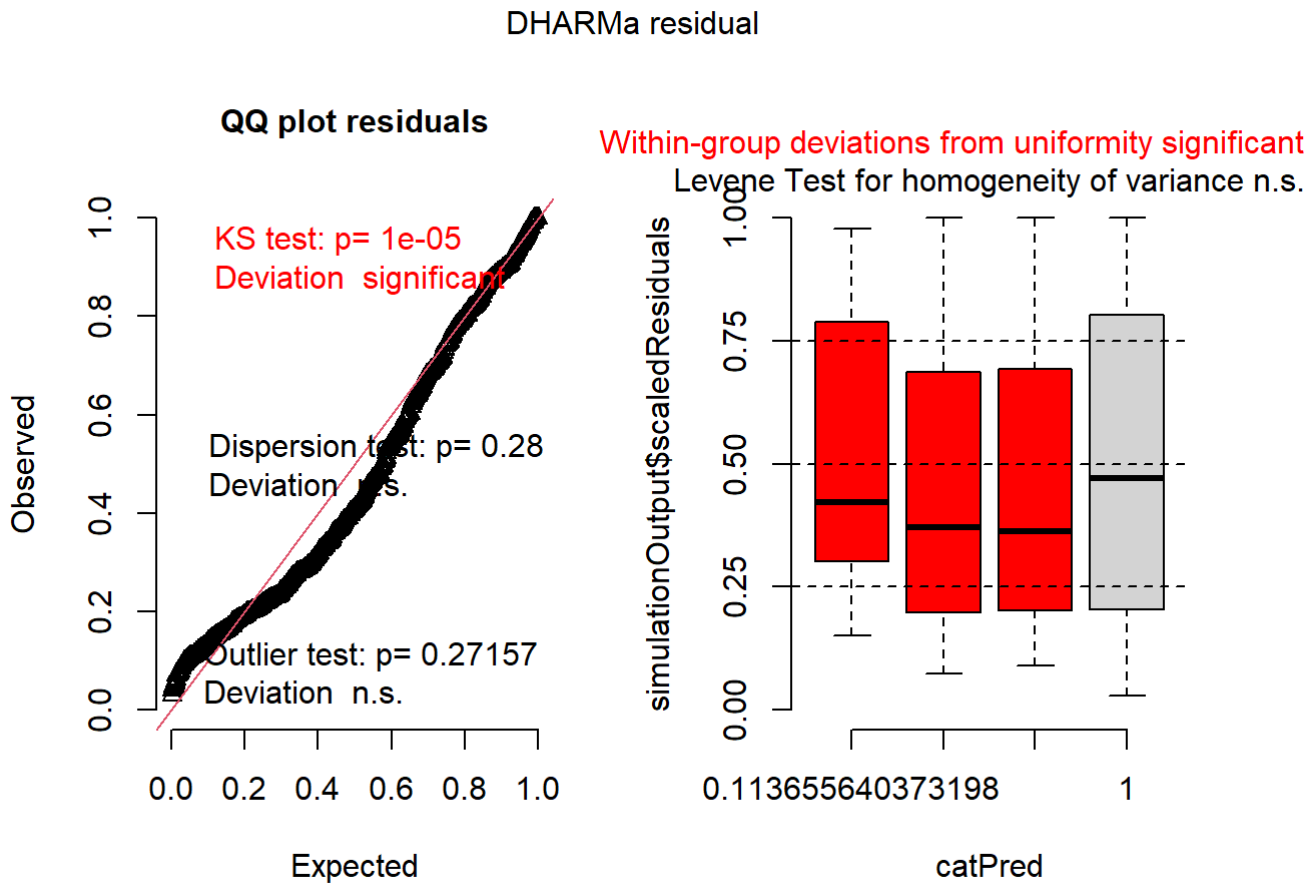
```
##  model weights
## [1] 0.397 0.603 0.000 0.000 0.000
```

The second model has the better AICc.

The final model take into account the fixed effect of the *Status* and the *Phase*, and the random effect of the individual.

```
del1_2<-glmmTMB(Delta~Status+Phase+(1|ID_parent),REML = TRUE,family = poisson(log),data=all)
```

Finally, model's residuals are checked.



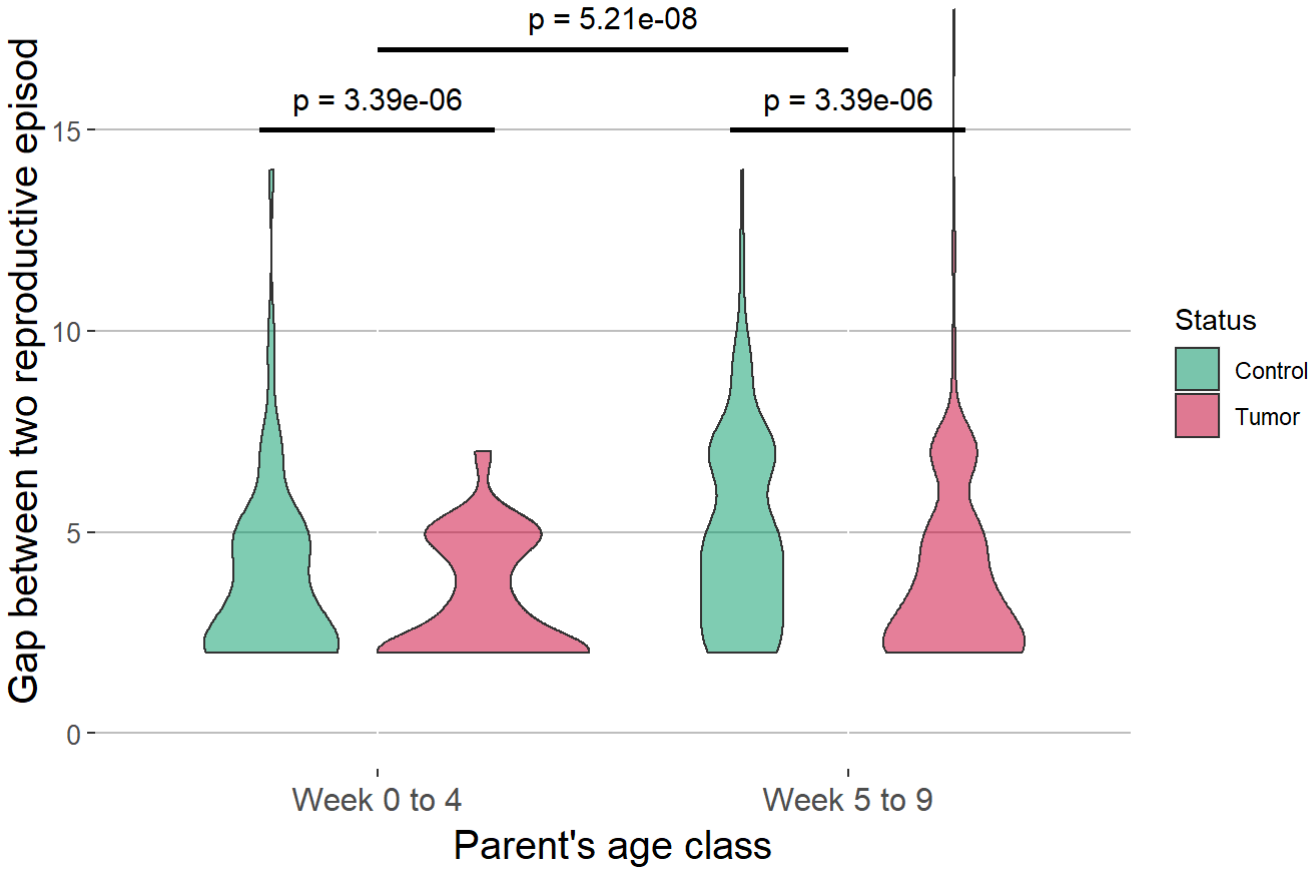
The overdispersion and the outlier tests are correct.

Final result

Delta ~ Status + Phase

Predictors	*Incidence Rate Ratios*	*std. Error*	*p*
Tumor status	0.81	0.04	**3.392e-06**
Passage of the 5th week	1.25	0.05	**5.208e-08**

Graphical representation

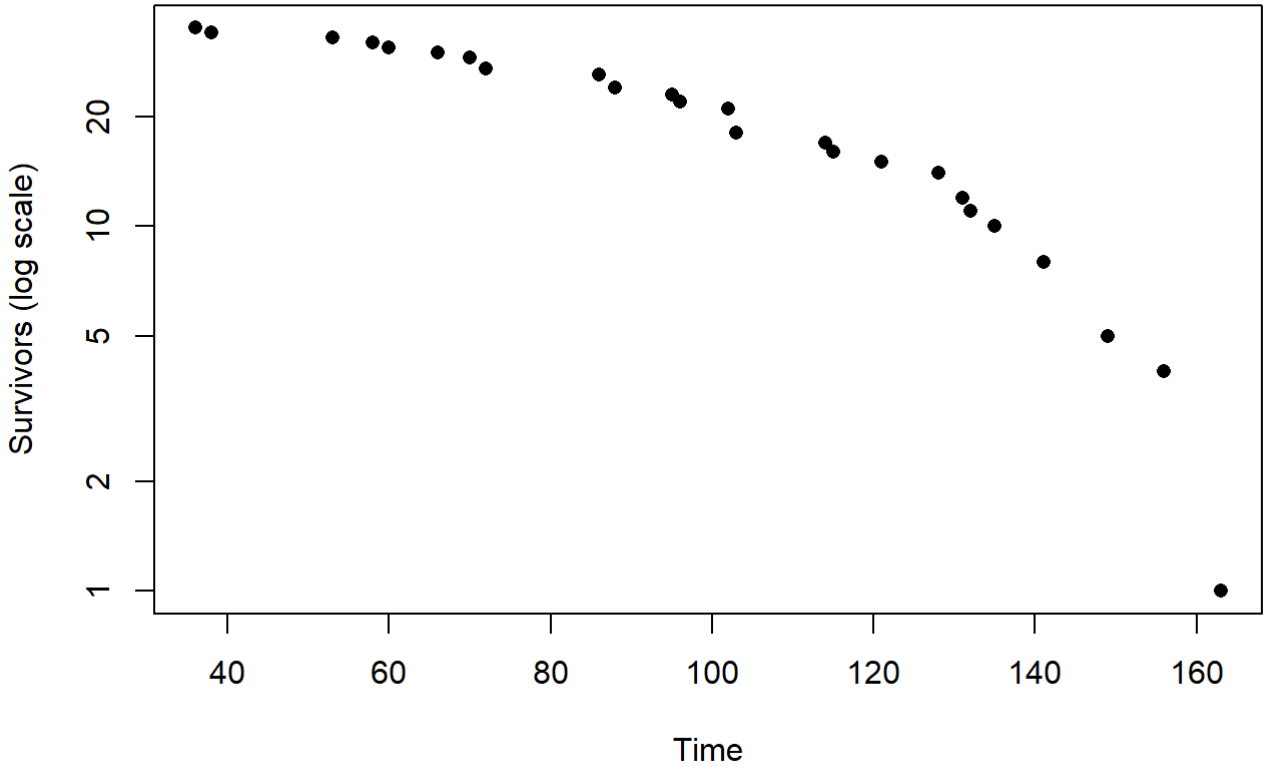


Survival time analysis

We start by calculating the survival time.

```
surv$survival_Time<-with(surv,ifelse(is.na(Death_date),as.numeric(surv$Last_check_date-surv$P
arent_birthday),as.numeric(surv$Death_date-surv$Parent_birthday)))
```

Then we traced the graph of the instantaneous risk over the time.



The instantaneous risk is not linear across the time so we will used a Weibull distribution to model this risk.

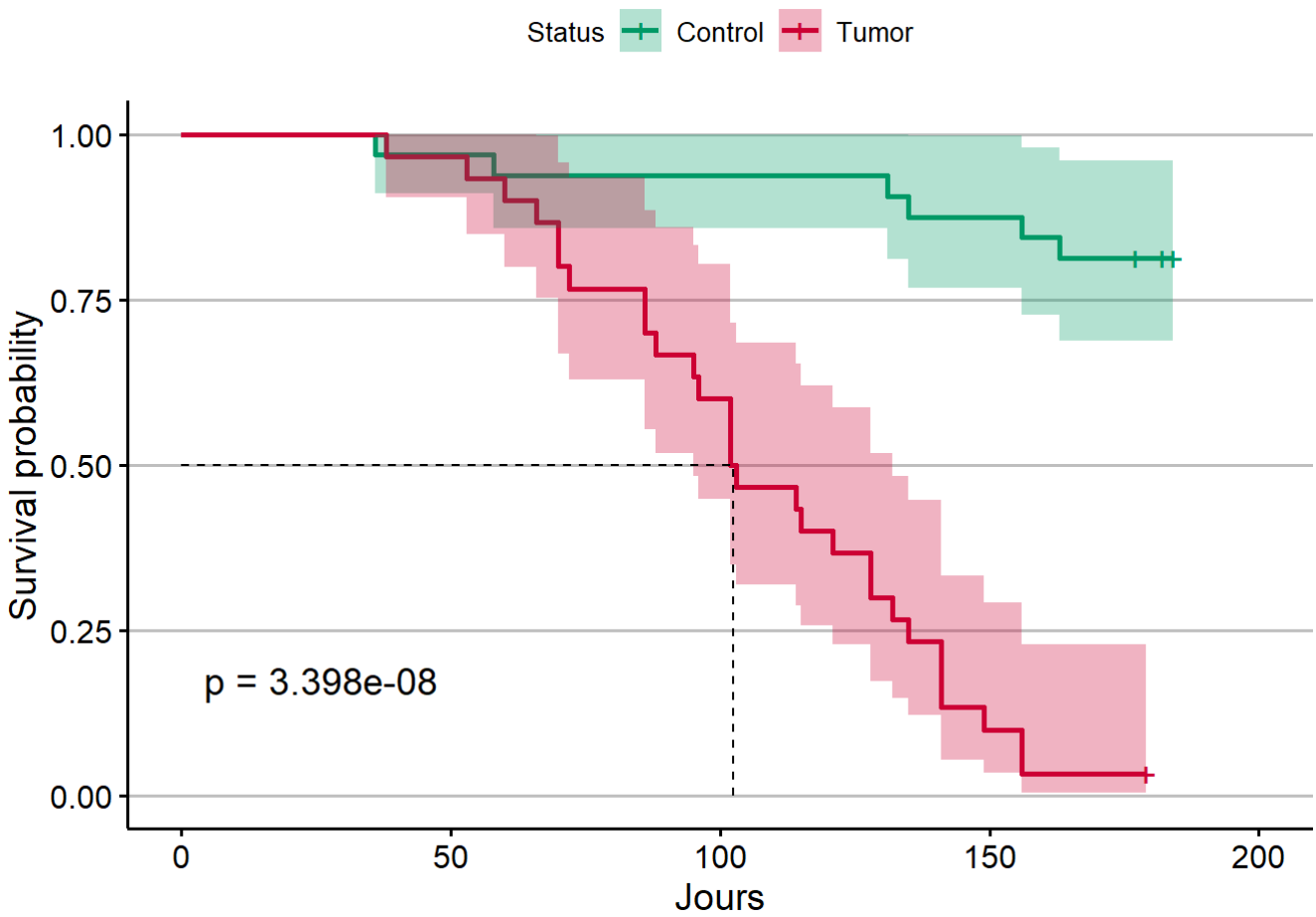
```
survie<-Surv(as.numeric(surv$survival_Time),surv$Censored)
modele<-survreg(survie~surv$Status,dist="weibull")
```

Final result

Survival time ~ Status

Predictors	*Estimates*	*std. Error*	*p*
Tumor status	0.39	0.07	**6.862e-08**
Log(scale)	0.33	0.05	**5.883e-15**
Observations	62		

Graphical representation



Analysis of bud's survival

First, the random effect is selected.

```
surv1<-glmmTMB(Survival~Status*Production_phase*Bud_type+(1|Complet_batch)+(1|Survey_date),RE
ML=TRUE, family = binomial(logit),data=transmission)
surv2<-glmmTMB(Survival~Status*Production_phase*Bud_type+(1|Complet_batch),REML=TRUE, family
 = binomial(logit),data=transmission)
surv3<-glmmTMB(Survival~Status*Production_phase*Bud_type+(1|Survey_date),REML=TRUE, family =
 binomial(logit),data=transmission)
surv4<-glmmTMB(Survival~Status*Production_phase*Bud_type+(1|Survey_date/Complet_batch),REML=T
RUE, family = binomial(logit),data=transmission)
surv5<-glmmTMB(Survival~Status*Production_phase*Bud_type,REML=TRUE, family = binomial(logit),
data=transmission)
```

```
##  model weights
## [1] 0.387 0.226 0.000 0.387 0.000
```

The first and the fourth model are equal, so we decided to keeps the nested random effect because at the same survey date different batch.

Then, we select the fixed effect.

```
surv4_1<-glmmTMB(Survival~Status*Production_phase*Bud_type+(1|Survey_date/Complet_batch),REML
=FALSE, family = binomial(logit),data=transmission)
surv4_2<-glmmTMB(Survival~Production_phase*Bud_type+(1|Survey_date/Complet_batch),REML=FALSE,
family = binomial(logit),data=transmission)
surv4_3<-glmmTMB(Survival~Status*Bud_type+(1|Survey_date/Complet_batch),REML=FALSE, family =
 binomial(logit),data=transmission)
surv4_4<-glmmTMB(Survival~Status*Production_phase+(1|Survey_date/Complet_batch),REML=FALSE, f
amily = binomial(logit),data=transmission)
surv4_5<-glmmTMB(Survival~Status+Production_phase+Bud_type+(1|Survey_date/Complet_batch),REML
=FALSE, family = binomial(logit),data=transmission)
surv4_6<-glmmTMB(Survival~Status+Production_phase*Bud_type+(1|Survey_date/Complet_batch),REML
=FALSE, family = binomial(logit),data=transmission)
surv4_7<-glmmTMB(Survival~Status*Production_phase+Bud_type+(1|Survey_date/Complet_batch),REML
=FALSE, family = binomial(logit),data=transmission)
surv4_8<-glmmTMB(Survival~Status*Bud_type+Production_phase+(1|Survey_date/Complet_batch),REML
=FALSE, family = binomial(logit),data=transmission)
surv4_9<-glmmTMB(Survival~Production_phase+Bud_type+(1|Survey_date/Complet_batch),REML=FALSE,
family = binomial(logit),data=transmission)
surv4_10<-glmmTMB(Survival~Status+Bud_type+(1|Survey_date/Complet_batch),REML=FALSE, family =
binomial(logit),data=transmission)
surv4_11<-glmmTMB(Survival~Status+Production_phase+(1|Survey_date/Complet_batch),REML=FALSE,
 family = binomial(logit),data=transmission)
surv4_12<-glmmTMB(Survival~Production_phase+(1|Survey_date/Complet_batch),REML=FALSE, family
 = binomial(logit),data=transmission)
surv4_13<-glmmTMB(Survival~Bud_type+(1|Survey_date/Complet_batch),REML=FALSE, family = binomi
al(logit),data=transmission)
surv4_14<-glmmTMB(Survival~Status+(1|Survey_date/Complet_batch),REML=FALSE, family = binomial
(logit),data=transmission)
surv4_15<-glmmTMB(Survival~1+(1|Survey_date/Complet_batch),REML=FALSE, family = binomial(logi
t),data=transmission)
```

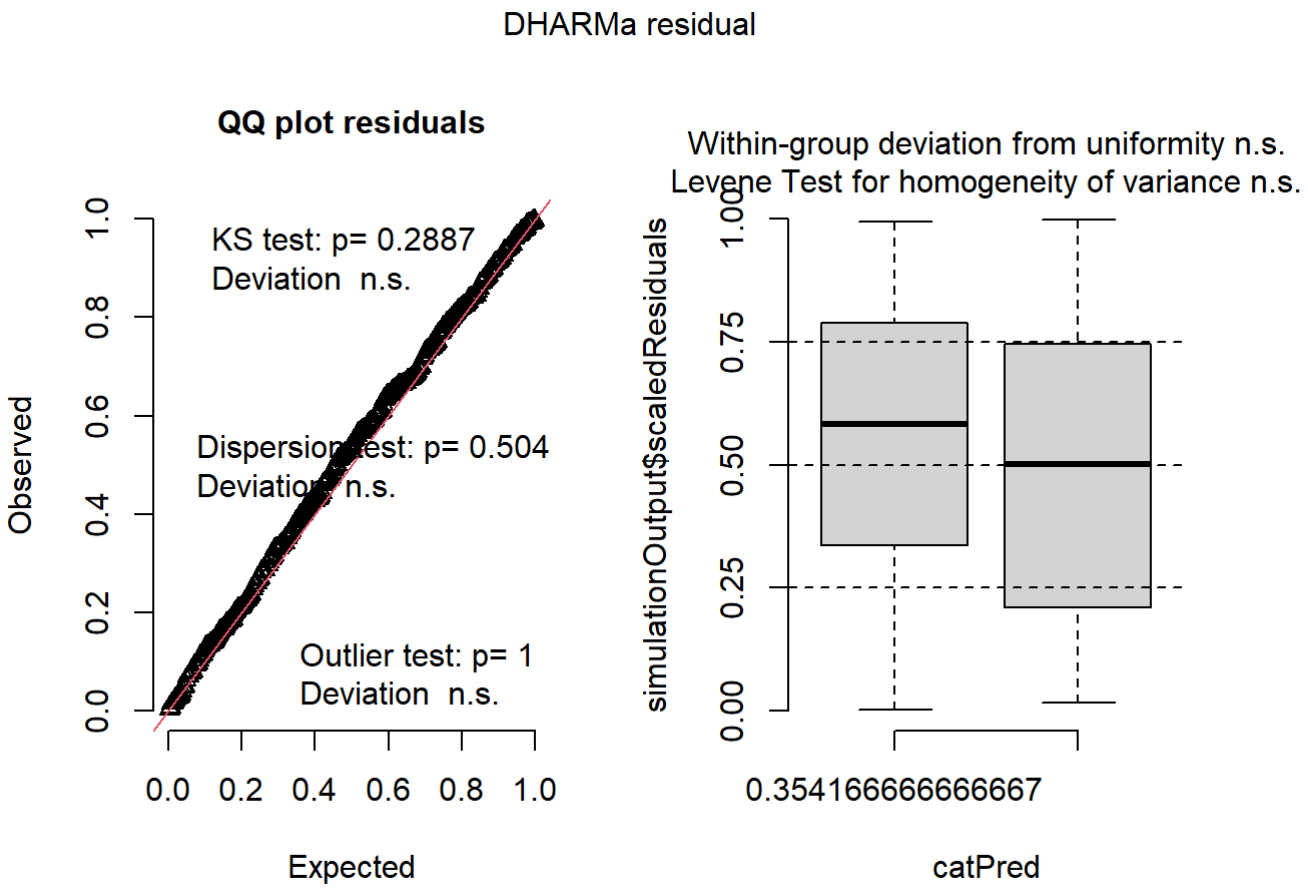
```
##  model weights
##  [1] 0.003 0.074 0.001 0.055 0.160 0.062 0.063 0.057 0.171 0.002 0.133 0.212
## [13] 0.003 0.001 0.003
```

The twelfth model has the better AICc.

The final model take into account the fixed of the *Production_phase*, and the nested random effect of the *Survey_date* and the *Complet_batch*.

```
surv4_12<-glmmTMB(Survival~Production_phase+(1|Survey_date/Complet_batch),REML=TRUE, family =
binomial(logit),data=transmission)
```

The modele's residuals are checked.



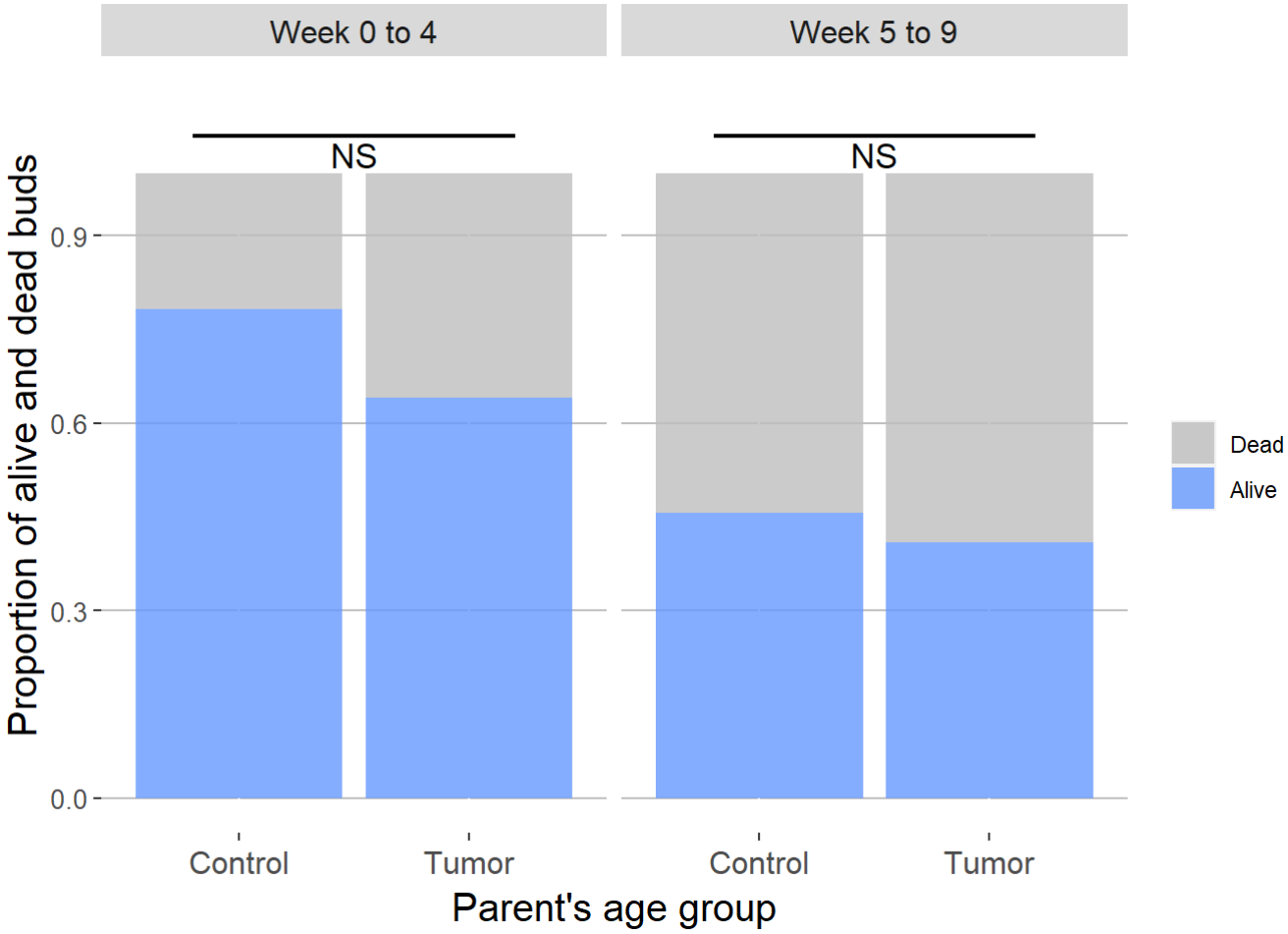

```
## Object of Class DHARMa with simulated residuals based on 250 simulations with refit = FALS
E . See ?DHARMa::simulateResiduals for help.
##
## Scaled residual values: 0.4767229 0.5598669 0.1379513 0.388332 0.02551519 0.02524903 0.852
6626 0.67773 0.8209898 0.6766341 0.6859425 0.3158154 0.5510649 0.4938672 0.1657359 0.6996305
0.317379 0.6103911 0.2221154 0.5752245 ...
```

Final result

Survival~Production_phase

Predictors	*Odds Ratios*	*std. Error*	*p*
Passage of the fifth week	5.22	2.89	**2.907e-03**
N $_{\text{Complet_batch}}$	32		
N $_{\text{Survey_date}}$	9		
Observations	372		

Graphical representation



Analysis of tumor transmission

We begin with isolating, inside the surviving buds, those which are from a tumor strain.

```
transmission_alive<-subset(transmission,Survival=="Alive")
transmission_alive$Bud_status<-as.factor(transmission_alive$Bud_status)
tum<-subset(transmission_alive, Status=="Tumor")
```

Afterward, we select the random effect needed. The last model does not converged, so it has been excluded.

```
devtum10<-glmmTMB(Bud_status~Production_phase+Bud_type+(1|Complet_batch)+(1|Survey_date),REML
=TRUE, family = binomial(logit),data=tum)
devtum20<-glmmTMB(Bud_status~Production_phase+Bud_type+(1|Complet_batch),REML=TRUE, family =
 binomial(logit),data=tum)
devtum30<-glmmTMB(Bud_status~Production_phase+Bud_type+(1|Survey_date),REML=TRUE, family = bi
nomial(logit),data=tum)
devtum40<-glmmTMB(Bud_status~Production_phase+Bud_type+(1|Survey_date/Complet_batch),REML=TRU
E, family = binomial(logit),data=tum)
devtum50<-glmmTMB(Bud_status~Production_phase+Bud_type,REML=TRUE, family = binomial(logit),da
ta=tum)
```

```
##  model weights
## [1] 0.125 0.375 0.375 0.125
```

The second and the third model have identical AICc so we decided to selected the random effect of the *Complet_batch* because it has more level than the *Survey_date*.

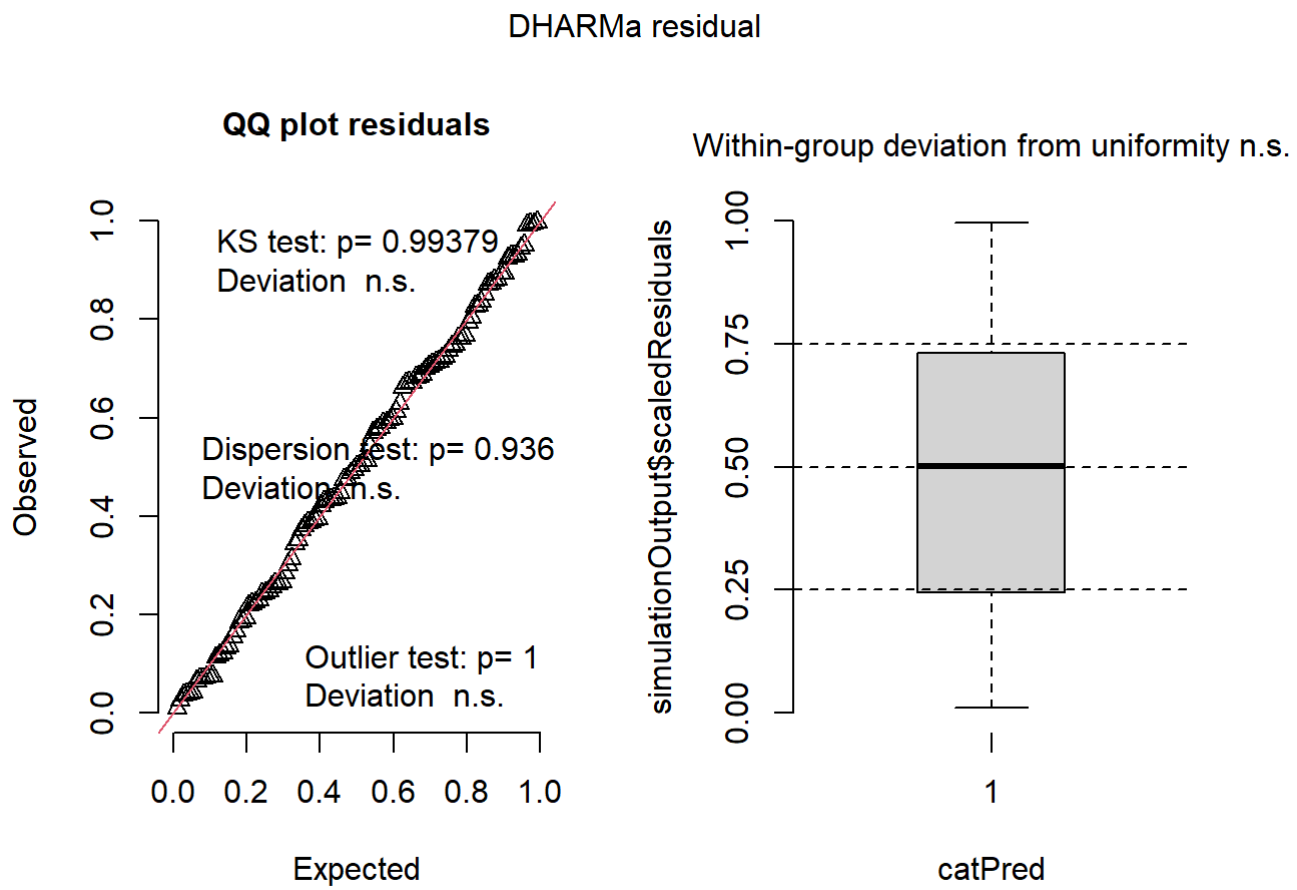
```
devtum41<-glmmTMB(Bud_status~Production_phase*Bud_type+(1|Complet_batch),REML=FALSE, family =
binomial(logit),data=tum)
devtum42<-glmmTMB(Bud_status~Production_phase+Bud_type+(1|Complet_batch),REML=FALSE, family =
binomial(logit),data=tum)
devtum43<-glmmTMB(Bud_status~Production_phase+(1|Complet_batch),REML=FALSE, family = binomial
(logit),data=tum)
devtum44<-glmmTMB(Bud_status~Bud_type+(1|Complet_batch),REML=FALSE, family = binomial(logit),
data=tum)
devtum45<-glmmTMB(Bud_status~1+(1|Complet_batch),REML=FALSE, family = binomial(logit),data=tu
m)
```

The last model has the better AICc.

The final model take into account no fixed effect, and the random effect of the *Complet_batch*.

```
devtum45<-glmmTMB(Bud_status~1+(1|Complet_batch),REML=TRUE, family = binomial(logit),data=tu
m)
```

The model's residual are checked.

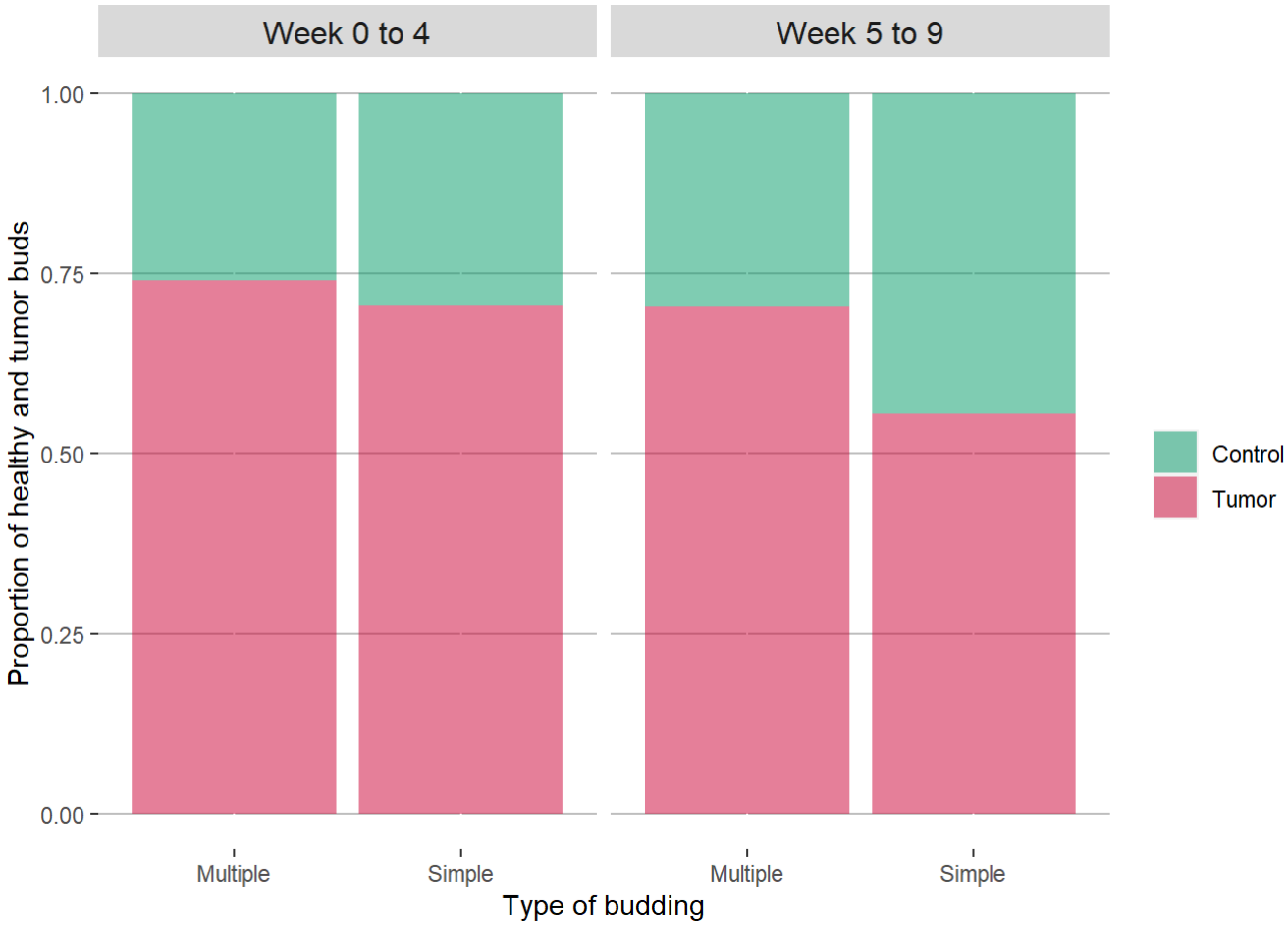


final results

Tumoral transmission ~1 + (1|Complet_batch)

Predictors	*Odds Ratios*	*std. Error*	*p*
Intercept	2.47	0.52	**1.612e-05**
N $_{\text{Complet_batch}}$	15		
Observations	111		

```
##  Family: binomial  ( logit )
## Formula:          Bud_status ~ 1 + (1 | Complet_batch)
## Data: tum
##
##      AIC      BIC   logLik deviance df.resid
##    138.6    144.0    -67.3    134.6      110
##
## Random effects:
##
## Conditional model:
##  Groups        Name        Variance  Std.Dev.
##  Complet_batch (Intercept) 3.037e-09 5.511e-05
## Number of obs: 111, groups:  Complet_batch, 15
##
## Conditional model:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9037     0.2095   4.313 1.61e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Code S2 : Code of the analysis of the traits related to sexual reproduction related to STAR Methods.

Tumors alter life-history traits in the freshwater cnidarian, Hydra oligactis

Justine BOUTRY and Sophie TISSOT

R Packages loading

```
library(readr)
library(lme4)
library(MASS)
library(DHARMa)
library(MuMIn)
library(glmmTMB)
library(ggplot2)
library(sjPlot)
library(ade4)
library(RVAideMemoire)
library(survival)
library(car)
library(survminer)
library(factoextra)
library(ggridges)
```

Data formatting

The dataframe *sexual_reproduction* (corresponding to *sexual_reproduction.csv*) contains the detail of sexual reproduction's following for each individual. The variables *Cold_survival_time*, *Eggs_number* and *First_reproduction_age* are kept in numerical form. The other variables are formatted in the factor form.

```
##                     Complet_ID          Phase         Status    Batch
##  Tumor_Week_5_to_9_7_24  :  2   Week_0_to_4:128   Control:141   2:23
##  Control_Week_0_to_4_2_1 :  1   Week_5_to_9:148   Tumor  :135   3:41
##  Control_Week_0_to_4_2_10:  1                                    4:55
##  Control_Week_0_to_4_2_11:  1                                    7:81
##  Control_Week_0_to_4_2_12:  1                                    8:76
##  Control_Week_0_to_4_2_2 :  1
##  (Other)                 :269
##        ID      Cold_placement_date      Death_date  Cold_survival_time
##  3      : 15   2021-03-11:23       2021-07-21: 38   Min.   :  6.00
##  2      : 14   2021-03-18:20       2021-08-04: 37   1st Qu.: 41.00
##  12     : 14   2021-04-22:35       2021-07-20: 32   Median : 61.00
##  1      : 13   2021-05-06:34       2021-07-07: 20   Mean   : 62.41
##  4      : 13   2021-05-20:42       2021-08-25: 16   3rd Qu.: 81.50
##  5      : 13   2021-06-08:84       2021-06-23: 14   Max.   :131.00
##  (Other):194   2021-07-08:38       (Other)   :119
##   Eggs_number     First_reproduction_date First_reproduction_age
##  Min.   :0.0000   2021-04-12:  7          Min.   :22.00
##  1st Qu.:0.0000   2021-05-25:  5          1st Qu.:29.00
##  Median :0.0000   2021-08-04:  4          Median :32.00
##  Mean   :0.3007   2021-05-31:  3          Mean   :32.65
##  3rd Qu.:0.0000   2021-07-09:  3          3rd Qu.:36.00
##  Max.   :5.0000   (Other)   : 27          Max.   :46.00
##                   NA's      :227          NA's   :227
```

Lexical

- *Complet_ID* : the name of each individual composed by the status, the phase, the bacth and the ID
- *Phase* : the age of the individual when it has been placed in the sexual reproduction conditions
- *Status* : the strain from which the individual comes from (tumorous or healthy)
- *Batch* : the number of the batch, each batch contains 12 individuals in each status
- *ID* : the number attributed to each individual inside each batch during the tests
- *Cold_placement_date* : the date when the individual have been place in the sexual reproduction condition
- *Death_date* : the date of the death of each individual
- *Cold_survival_time* : the number of days that each individual survived under the sexual reproduction conditions
- *Eggs_number* : the total number of eggs produced by each individuals
- *First_reproduction_date* : the date where the first eggs produced were observed
- *First_reproduction_age* : the number of days separating the starting of the experiment and the first egg production

Analysis of the age at the first reproduction

Three on the four groups respect the normality hypothesis.

```
## $`Week_0_to_4:Control`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.96092, p-value = 0.7963
##
##
## $`Week_0_to_4:Tumor`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.79135, p-value = 0.007483
##
##
## $`Week_5_to_9:Control`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.95393, p-value = 0.5215
##
##
## $`Week_5_to_9:Tumor`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.97495, p-value = 0.9326
```

Because three groups on four respect the normality hypothesis, we choose to used linear mixed effects models. We start by testing if the random effect of the date is useful for explaining the variable.

```
f1<-lm(First_reproduction_age~Status*Phase,data=sexual_reproduction)
f2<-lmer(First_reproduction_age~Status*Phase+(1|Cold_placement_date),REML = TRUE,data=sexual_
reproduction)
```

```
##  model weights
## [1] 0.004 0.996
```

The weights of AICc suggest that the model with the random effect fit more to the data than the other one.

Then we select the fixed effects needed.

```
f2_1<-lmer(First_reproduction_age~Status*Phase+(1|Cold_placement_date),REML = FALSE,data=sexu
al_reproduction)
f2_2<-lmer(First_reproduction_age~Status+Phase+(1|Cold_placement_date),REML = FALSE,data=sexu
al_reproduction)
f2_3<-lmer(First_reproduction_age~Status+(1|Cold_placement_date),REML = FALSE,data=sexual_rep
roduction)
f2_4<-lmer(First_reproduction_age~Phase+(1|Cold_placement_date),REML = FALSE,data=sexual_repr
oduction)
f2_5<-lmer(First_reproduction_age~1+(1|Cold_placement_date),REML = FALSE,data=sexual_reproduc
tion)
```

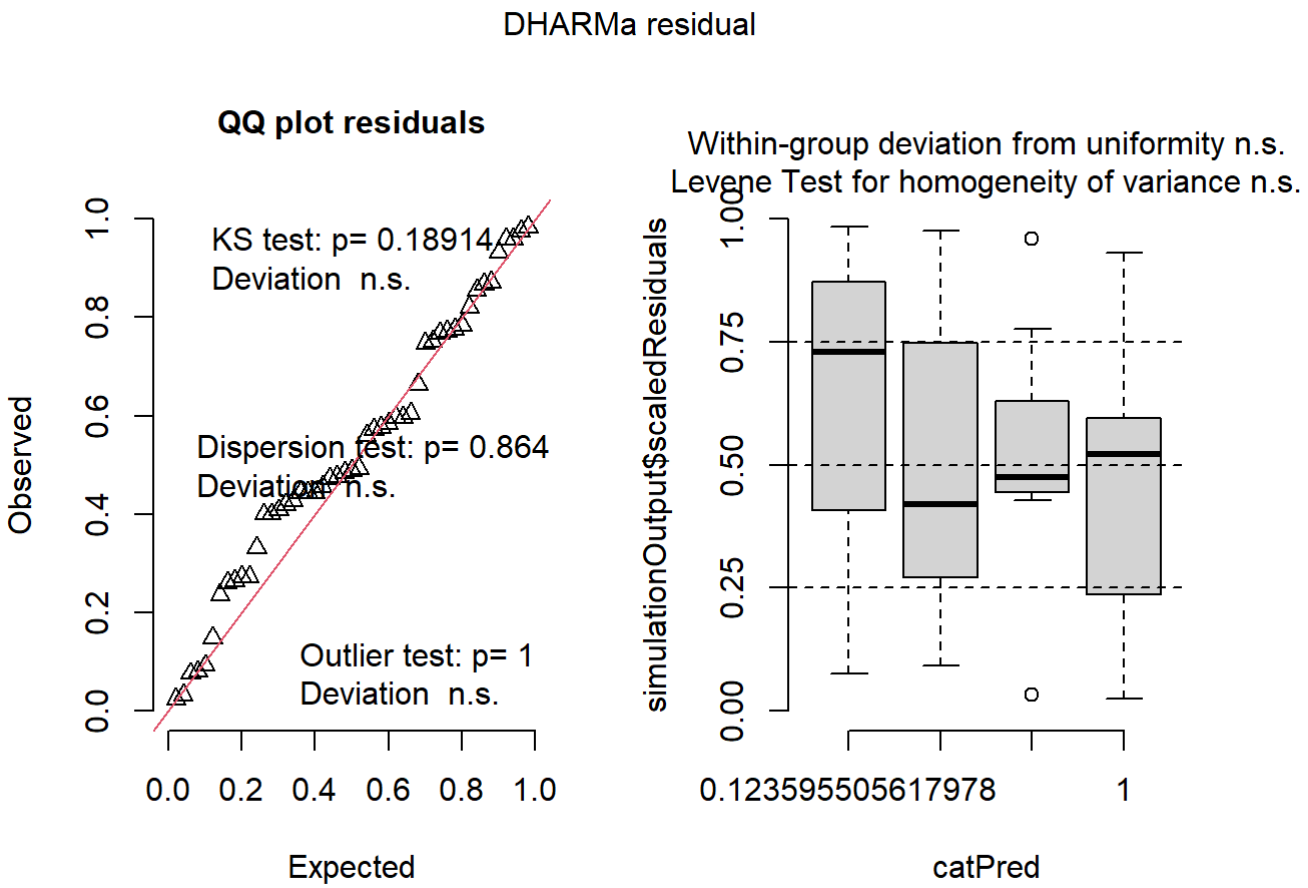
```
##  model weights
## [1] 0.734 0.258 0.000 0.008 0.000
```

The fixed effect of the interaction between the *Status* and the *Phase* seems to be the one allowing the better explanation of the variable.

The final model take into account the fixed of the interaction between the *Status* and the *Phase*, and the random effect of the date.

```
f2_1<-lmer(First_reproduction_age~Status*Phase+(1|Cold_placement_date),REML = TRUE,data=sexua
l_reproduction)
```

Finally, model's residuals are checked to see if the model fit and explain really the data.



Overdispersion, Kolmogorov-Smirnov and outlier test end all correct.

Final result

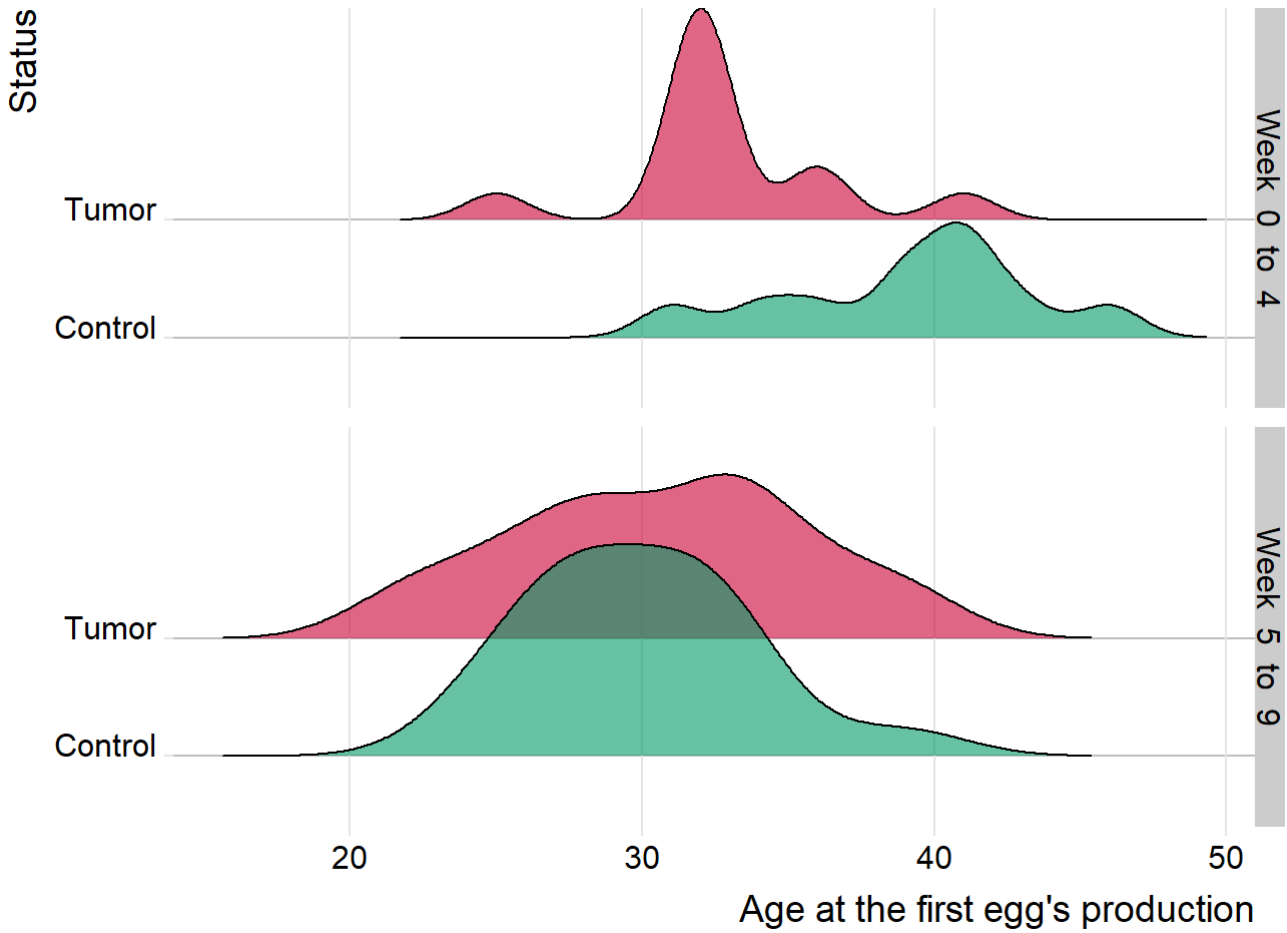
First reproduction age~Status*Phase

Predictors	*Estimates*	*std. Error*	*p*
Tumor status	-7.01	1.77	**2.792e-04**
Passage of the 5th week	-9.88	1.91	**5.743e-06**
Tumor status following the 5th week	5.11	2.44	**4.246e-02**

Random Effects

σ^2	13.63
$\tau_{00\ Cold_placement_date}$	6.50
Observations	49

Graphical representation



Analysis of the number of eggs produced

First we select the random effect in the model.

```
m1<-glmmTMB(Eggs_number~Status*Phase+(1|Cold_placement_date),REML = TRUE,family = nbinom1,dat
a=sexual_reproduction)
m2<-glmmTMB(Eggs_number~Status*Phase,REML = TRUE,family = nbinom1,data=sexual_reproduction)
```

```
##  model weights
## [1] 0.998 0.002
```

The model with the random effect of the date has the lower AICc.

Then, it's the fixed effect which is tested.

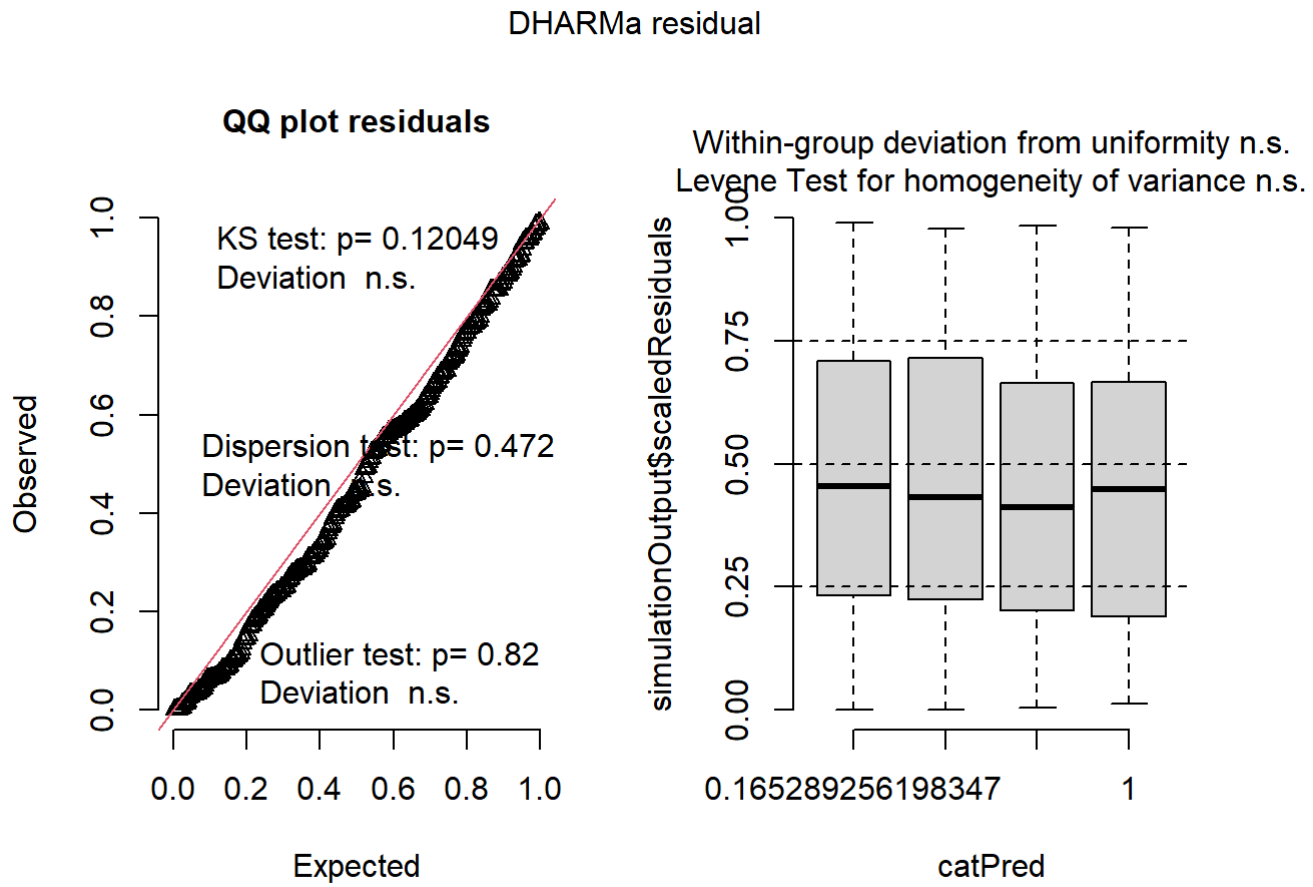
```
m1_1<-glmmTMB(Eggs_number~Status*Phase+(1|Cold_placement_date),REML = FALSE,family = nbinom1,
data=sexual_reproduction)
m1_2<-glmmTMB(Eggs_number~Status+Phase+(1|Cold_placement_date),REML = FALSE,family = nbinom1,
data=sexual_reproduction)
m1_3<-glmmTMB(Eggs_number~Status+(1|Cold_placement_date),REML = FALSE,family = nbinom1,data=s
exual_reproduction)
m1_4<-glmmTMB(Eggs_number~Phase+(1|Cold_placement_date),REML = FALSE,family = nbinom1,data=se
xual_reproduction)
m1_5<-glmmTMB(Eggs_number~1+(1|Cold_placement_date),REML = FALSE,family = nbinom1,data=sexual
_reproduction)
```

```
##  model weights
## [1] 0.418 0.043 0.114 0.116 0.308
```

The effect of the interaction is kept.

The final model take into account the fixed of the interaction between the *Status* and the *Phase*, and the random effect of the date.

Finally, model's residuals are checked.



The model selected fit well the data.

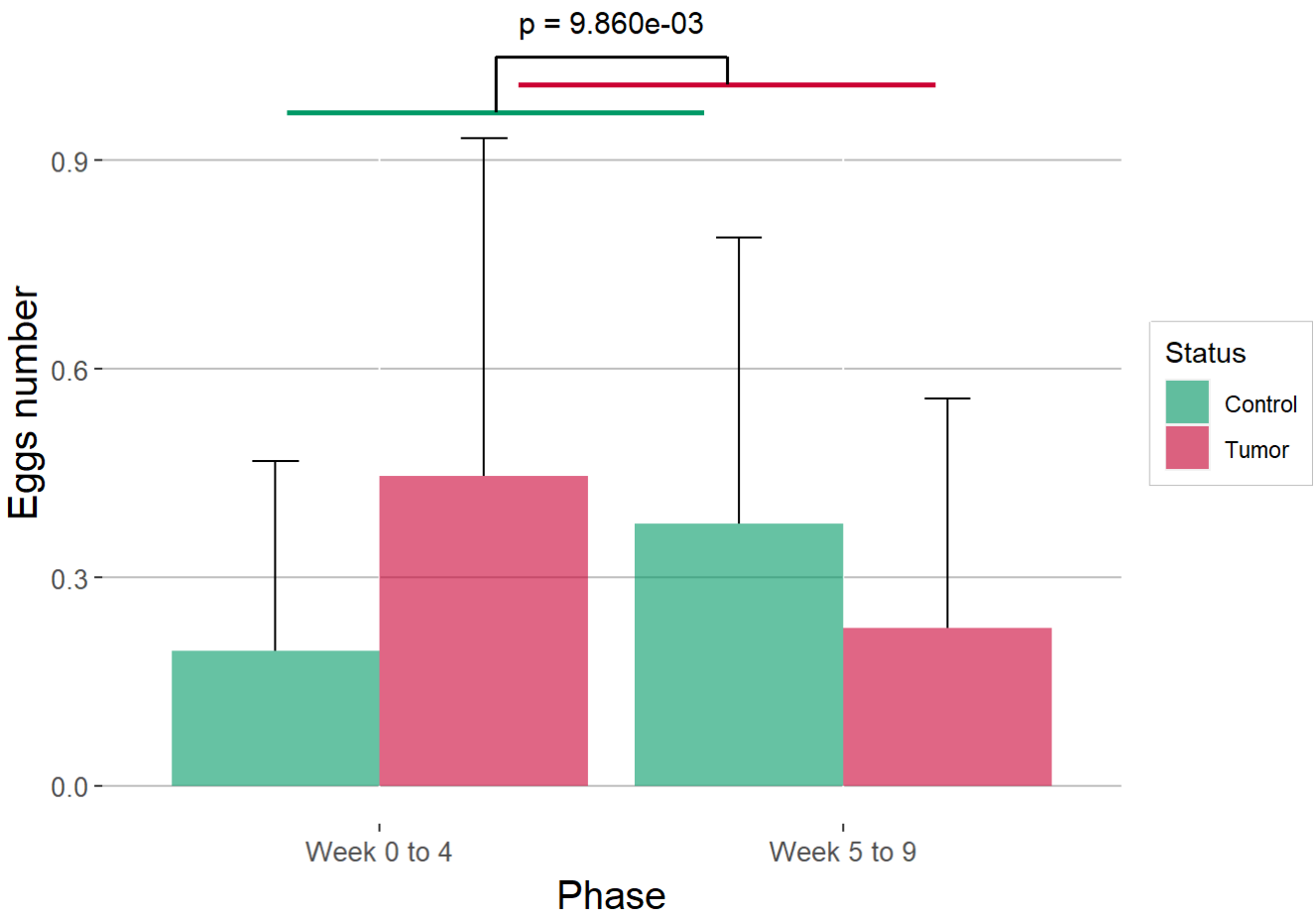
Final result

Eggs number~Status*Phase

Predictors	*Incidence Rate Ratios*	*std. Error*	*CI*	*p*
Tumor status	2.01	0.82	0.00 – Inf	8.701e-02
Passage of the 5th week	2.44	1.38	0.00 – Inf	1.144e-01
Tumor status following the 5th week	0.23	0.13	0.00 – Inf	**9.860e-03**

Random Effects

σ^2	2.03
$\tau_{00\ Cold_placement_date}$	0.69
Observations	276

Graphical representation

```
## Warning: le package 'plyr' a été compilé avec la version R 4.1.3
```