Quantitative genetic analysis of germination in the residual allogamy dataset

Margaux Jullien

19 juillet 2016

What are the underlying traits influencing residual allofecondation in *Medicago truncatula*?

- Flowering date of the mother plant

- Flowering date of the "father" plant

- Environmental conditions: zone, quadrat (plant density per quadrat = covariable?), plan (intra/interpop)

- Seed numbers, pod numbers, number of viable seeds

Supplementary questions: does flowering date affect the probability of being a father?

Analysis of life history traits during seed germination: are there genetic effects, effects of the mother plant's flowering date? Does allofecondation have an effect on life history traits during seed germination?

- Seed swelling before scarification (seed dormancy)

- Speed of radicle emergence

- Speed of cotyledon emergence (deployed cotyledons)

- Seed viability

- Having developed a radicle

- having developped cotyledons

Radicle emergence

The speed of radicle emergence is computed as the difference between the first observation of a radicle and the swelling date of the seed.

Effects looked into:

- Genetic: genotype of the mother plant (within each plan, intrapop and interpop)

- Environment: Plan (intra/interpop), Zone (within intrapop), Quadrat, Pod, mother plant (= maternal environment)

- Experiment: block, plate, box

Boxplots

Vradicule ~ Block

Vradicule ~ Plate

Vradicule ~ UniqueGeno

Vradicule ~ Zone
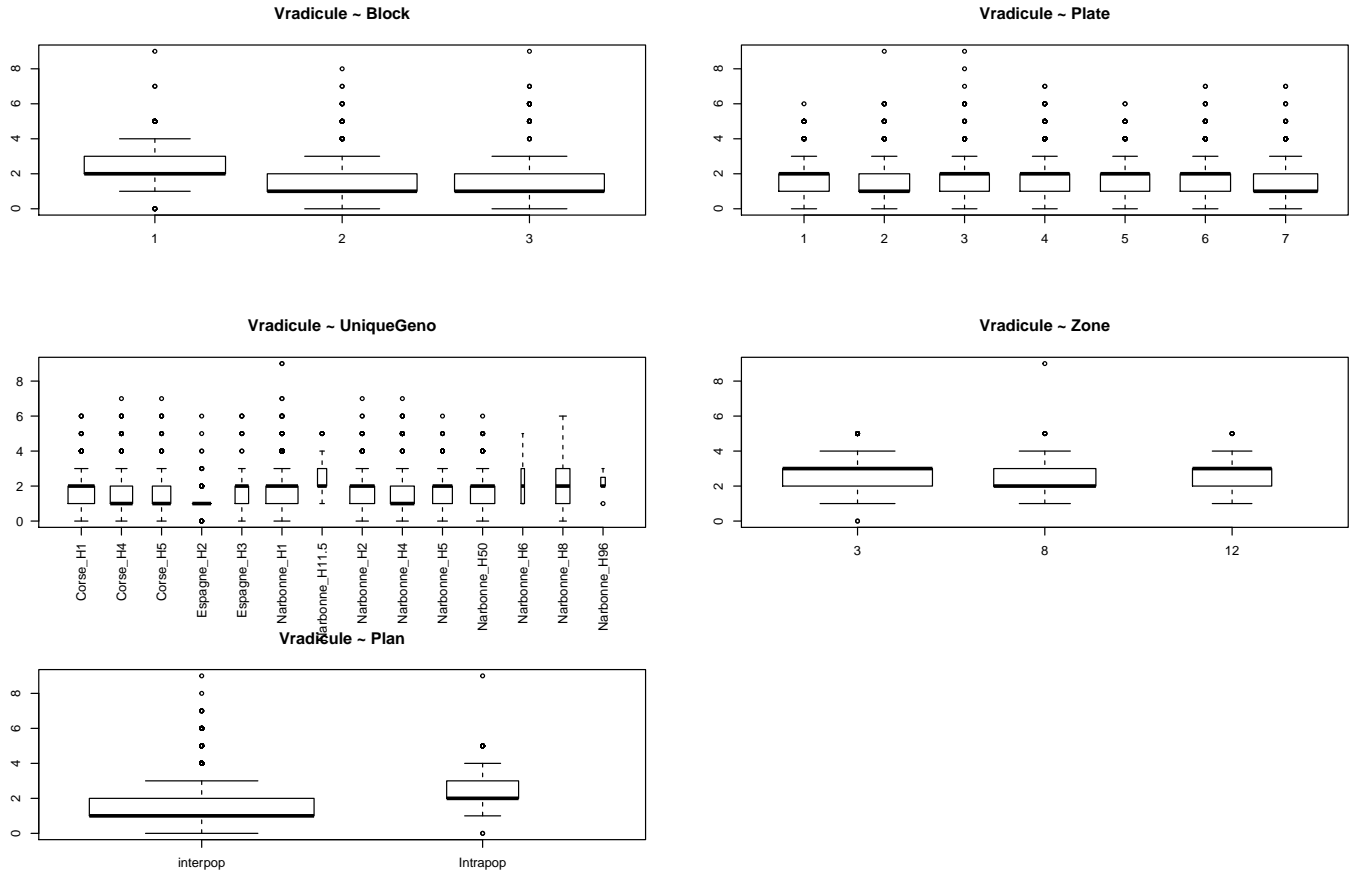
Vradicule ~ Plan

Mixed linear models

Testing zone effect within intrapop data

Vradicule ~ Zone + (1|Geno) + (1|Plate) + (1|Quadrat) + (1|Plant) + (1|IDPod) + (1|IDBox)

```
## Data: data_intrapop
## Models:
## mod2: Vradicule ~ (1 | UniqueGeno) + (1 | Plate) + (1 | UniqueQuadrat) +
## mod2:     (1 | Plant) + (1 | IDPod) + (1 | IDBox)
## mod1: Vradicule ~ Zone + (1 | UniqueGeno) + (1 | Plate) + (1 | UniqueQuadrat) +
## mod1:     (1 | Plant) + (1 | IDPod) + (1 | IDBox)
##      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod2  8 1484.4 1518.1 -734.22   1468.4
## mod1 10 1485.9 1528.0 -732.93   1465.9 2.5793      2     0.2754
```

There is no zone effect on the speed of radicle emergence in the intrapop dataset, we can continue with only a Plan effect on the whole dataset (Block effect is removed from this model as all the seed from the intrapop dataset germinated during Block 1).

Testing whole dataset

Vradicule ~ Block + Plan + (0 + Plan|Geno) + (1|Plate) + (1|Quadrat) + (1|Plant) + (1|Pod) + (1|Pod)

The model simplification is done using LRT and the results are verified with AICc.

Results

Selected model: **Vradicule ~ Block + (1|Plate) + (1|Quadrat) + (1|Plant) + (1|Pod)**

	Estimate	Std. Error	t value
(Intercept)	2.616	0.095	27.675
Block2	-1.073	0.094	-11.479
Block3	-1.175	0.094	-12.532

Percentage of explained variance:

Effect	Block	Quadrat	Plate	Plant	Pod	Residus
p-value	2.2e-16	0.007073	1.888e-07	1.647e-08	2.2e-16	
Variance (%)		2.28	2.35	4.53	19.03	71.81

On average, seeds from Block 1 have a speed of radicle emergence of 2.62 days, it decreases by 1.07 day in Block 2 and by 1.18 in Block 3.
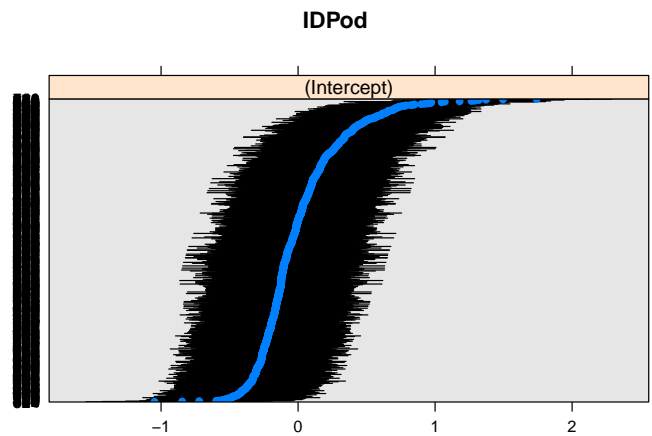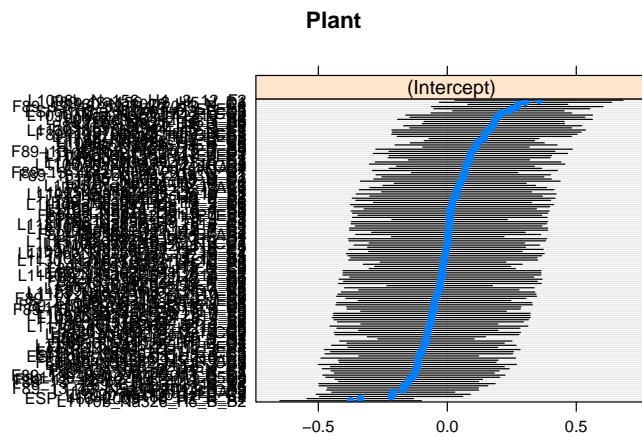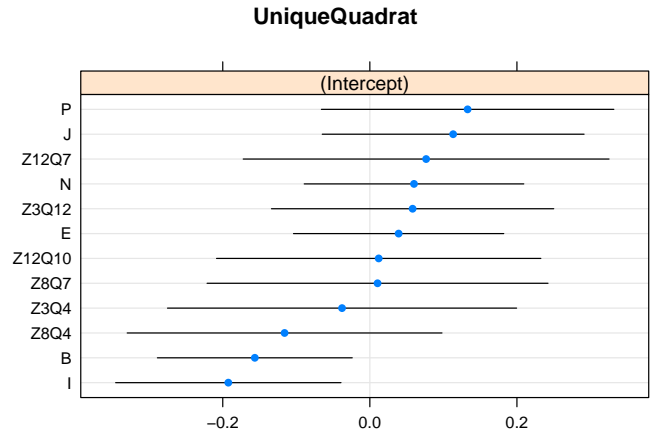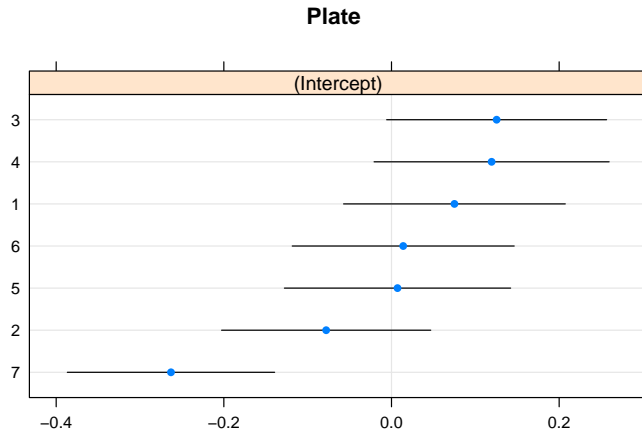
Residuals analysis
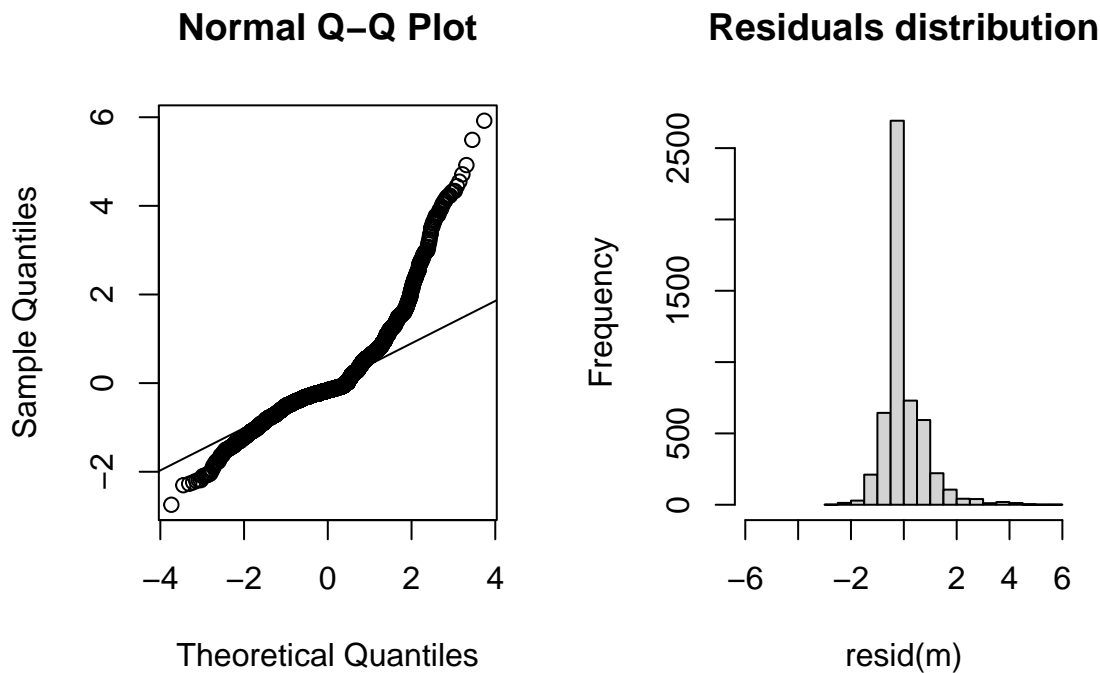

```
## 
##  Shapiro-Wilk normality test
## 
## data:  resid(m)[1:5000]
## W = 0.87881, p-value < 2.2e-16
```

The residuals are not normally distributed but they are centered on zero.

BLUPs

BLUPs are computed for the random effects of the model:

Covariance analysis

Flowering time

```
## Data: vradFlo
## Models:
## m: Vradicule ~ Block + (1 | Plate) + (1 | UniqueQuadrat) + (1 |
## m:      Plant) + (1 | IDPod)
## m1: Vradicule ~ Block + time_flo + (1 | Plate) + (1 | UniqueQuadrat) +
## m1:      (1 | Plant) + (1 | IDPod)
##    Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   8 13404 13457 -6694.3    13388
## m1  9 13405 13463 -6693.3    13387 1.9257      1     0.1652
```

No significative effect of flowering time on the speed of radicle emergence.

Number of pods per plant

```
## Data: vradNbPods
## Models:
## m: Vradicule ~ Block + (1 | Plate) + (1 | UniqueQuadrat) + (1 |
## m:      Plant) + (1 | IDPod)
## m1: Vradicule ~ Block + TotalNbPods + (1 | Plate) + (1 | UniqueQuadrat) +
```

```
## m1:     (1 | Plant) + (1 | IDPod)
##    Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   8 13860 13913 -6922.0    13844
## m1  9 13862 13921 -6921.8    13844 0.3797      1     0.5378
```

No significative effect of the number of pods per plant.

Number of seeds per plant

```
## Data: vradSeed
## Models:
## m: Vradicule ~ Block + (1 | Plate) + (1 | UniqueQuadrat) + (1 |
## m:     Plant) + (1 | IDPod)
## m1: Vradicule ~ Block + NumSeedPlant + (1 | Plate) + (1 | UniqueQuadrat) +
## m1:     (1 | Plant) + (1 | IDPod)
##    Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   8 14049 14102 -7016.7    14033
## m1  9 14051 14110 -7016.5    14033 0.2888      1     0.591
```

No significative effect of the number of seeds per plant.

Generalized linear mixed model (GLMM): Poisson distribution

As the residuals of the linear mixed model are not normally distributed, we analyse the speed of radicle emergence with a
GLMM fitting a Poisson distribution (count of days until radicle emergence).
Two methods have been used to identify the best model: using bootstrap to compute the confidence intervals of each
effect's variance and the LRT method.

Bootstrap (profiling method)

The confidence intervals (CI) of the model's effects were determined using 100 bootstrap simulations:

```
##            intercept/variance          CI_inf          CI_sup
## intercept       9.534645e-01   0.8273214605   1.1019041317
## block2         -5.475542e-01  -0.6708411297  -0.4446923111
## block3         -5.925968e-01  -0.7193469542  -0.4825703633
## PlanIntra       6.196819e-03  -0.1805274805   0.1663789169
## Box             0.000000e+00  -0.0060078210   0.0030926673
## Pod             2.280043e-10  -0.0072704597   0.0039264161
## Plant           8.411520e-03   0.0020205970   0.0168975476
## GenoInter       2.935613e-03  -0.0039848370   0.0088896220
## GenoIntra       1.970094e-04  -0.0082254278   0.0041250102
## Quadrat         4.459506e-03  -0.0014653535   0.0126588205
## Plate           6.395324e-03  -0.0002972868   0.0145884987
## Seed            0.000000e+00  -0.0005709717   0.0002749441
```
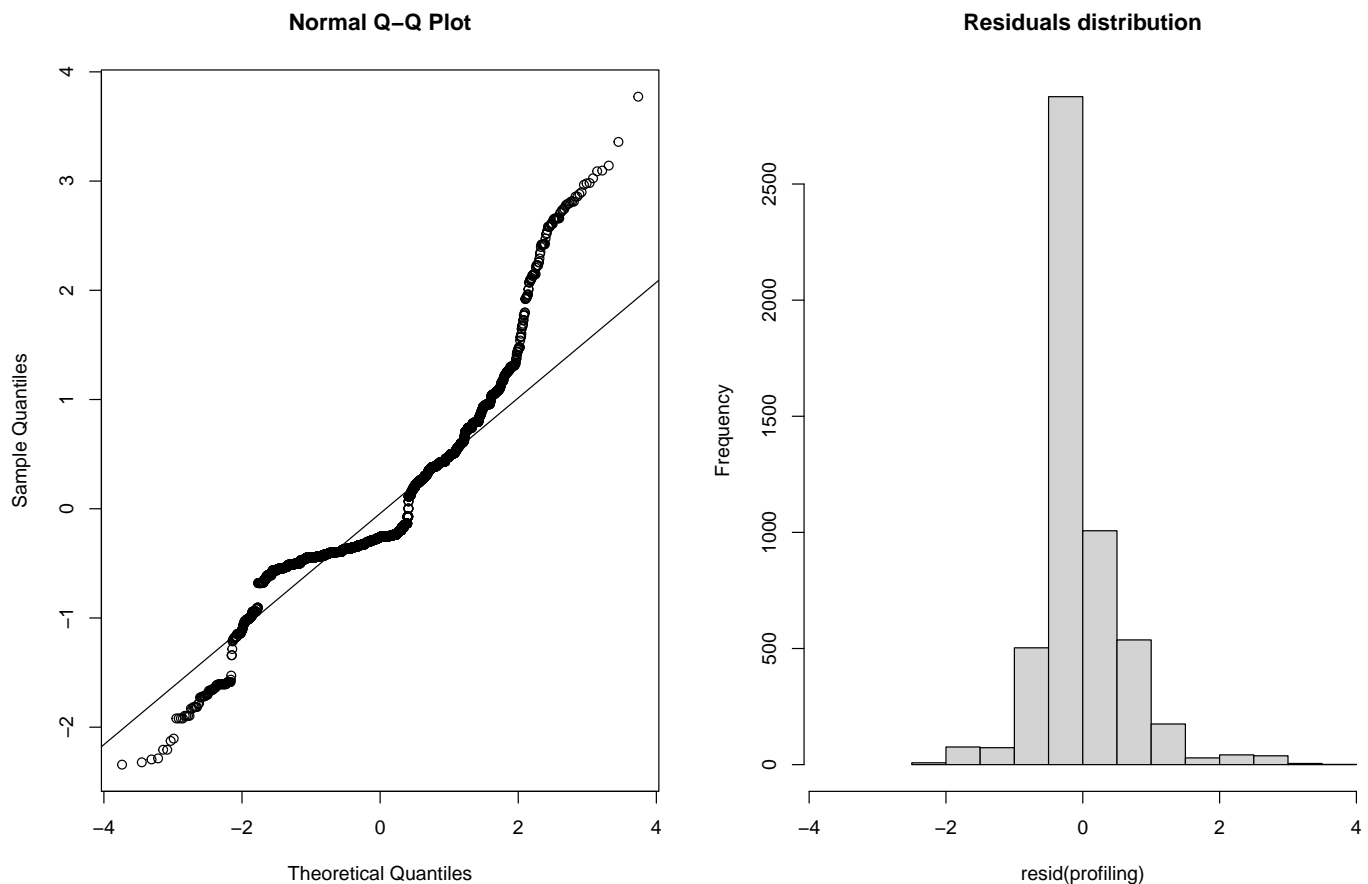
For some effects variance, the confidence intervals include 0, which means that the variance is not different from 0:

- Box

- Pod

- Genotype within interpop
- Genotype within intrapop
```

- Quadrat

- Plate

- Seed

Using the profiling method, the model for the speed of radicle emergence would simplify as: **Vrad ~ Block + (1|Plant) + (1|Plate)**

**Residuals analysis**



The residuals are not normally distributed.

**LRT method**

Using LRT method, the simplified model is **Vrad ~ Block + (1|Quadrat) + (1|Plnt) + (1|Plate)**

```
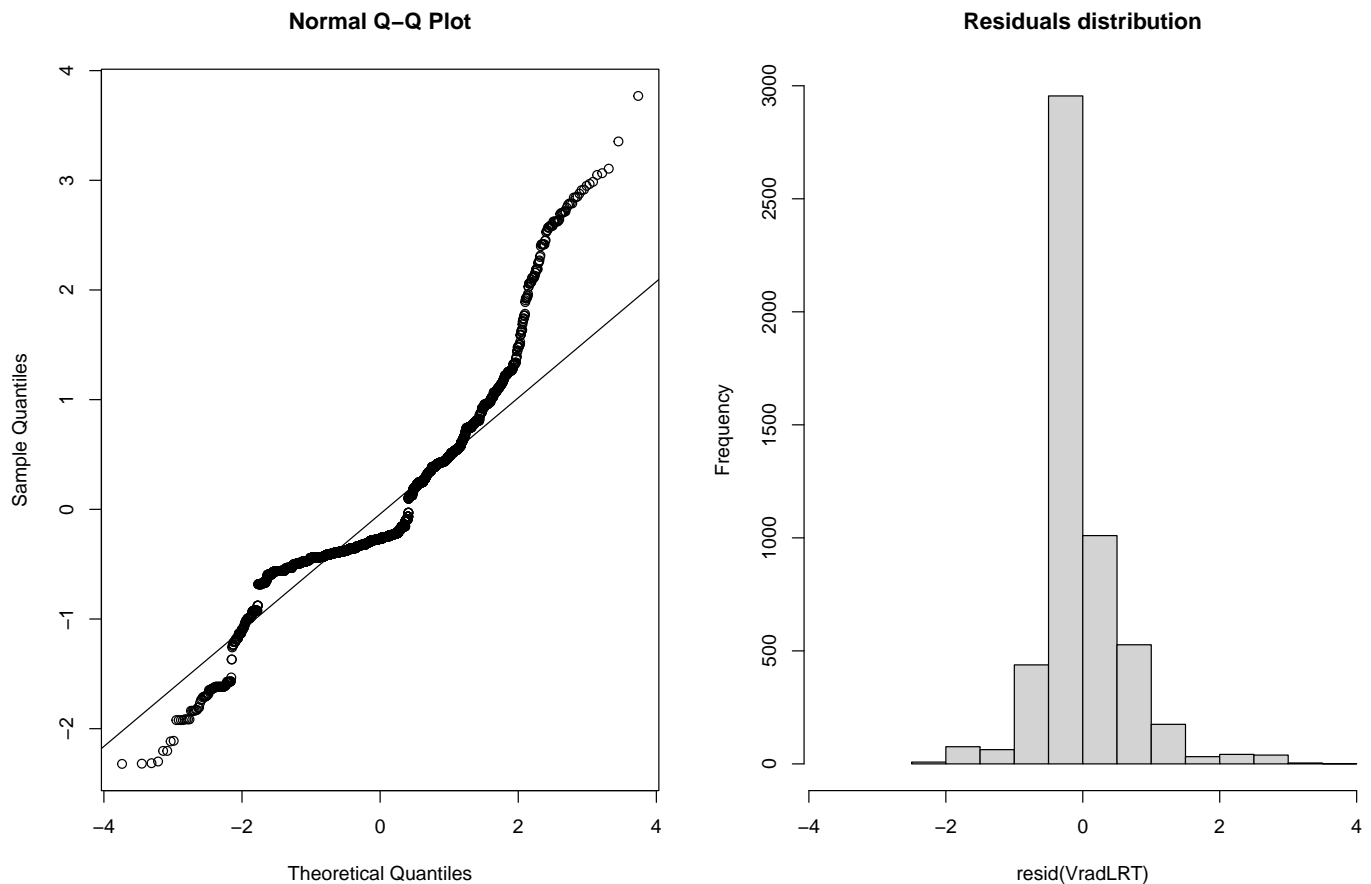## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Vradicule ~ Block + (1 | Plate) + (1 | UniqueQuadrat) + (1 |
##     Plant)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##  15039.8  15079.3  -7513.9  15027.8     5364
##
## Scaled residuals:
```

```
##      Min       1Q  Median      3Q      Max
## -1.6412 -0.3807 -0.2556  0.3275   5.2650
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  Plant         (Intercept) 0.010632 0.10311
##  UniqueQuadrat (Intercept) 0.004914 0.07010
##  Plate         (Intercept) 0.006300 0.07938
## Number of obs: 5370, groups:  Plant, 142; UniqueQuadrat, 12; Plate, 7
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.95536    0.05019   19.04   <2e-16 ***
## Block2      -0.54323    0.05059  -10.74   <2e-16 ***
## Block3      -0.60724    0.05105  -11.90   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr) Block2
## Block2 -0.540
## Block3 -0.541  0.856
```

Residuals analysis

The residuals are not normally distributed.

Cotyledon emergence

The speed of cotyledon emergence was computed as the difference between the observation of deployed cotyledons and the date at which the seed was first observed swollen.

Effects looked into:

- Genetic: genotype of the mother plant (within each plan, intrapop and interpop)

- Environment: Plan (intra/interpop), Zone (within intrapop), Quadrat, Pod, mother plant (= maternal environment)

- Experiment: block, plate, box

Boxplots

Mixed linear models

Testing zone effect within intrapop data

Vcotyledon ~ Zone + (1|Geno) + (1|Plate) + (1|Quadrat) + (1|Plant) + (1|Box) + (1|Pod)

```
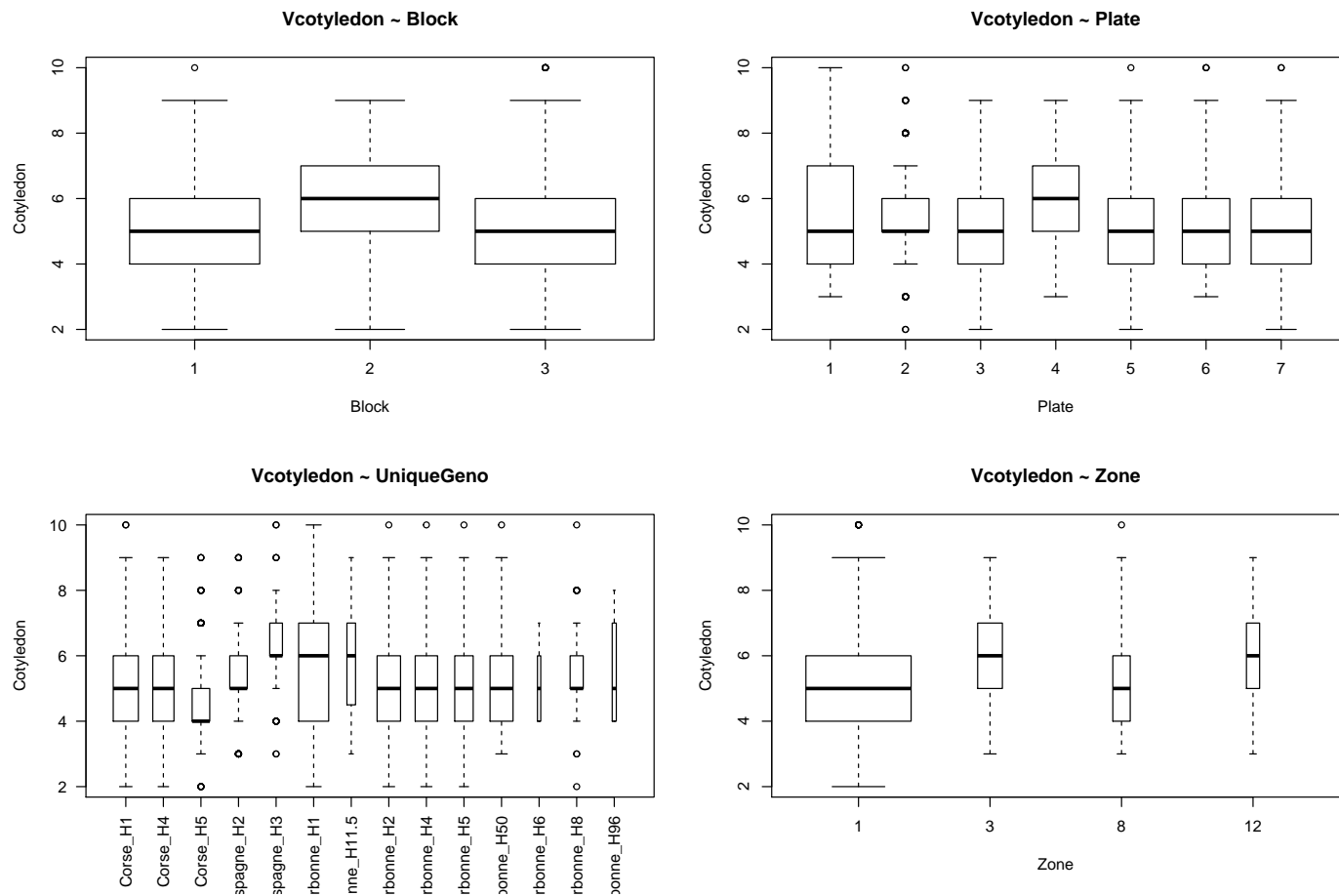## Data: data_intrapop
## Models:
## mod2: Vcotyledon ~ (1 | UniqueGeno) + (1 | Plate) + (1 | UniqueQuadrat) +
## mod2:     (1 | Plant) + (1 | IDBox) + (1 | IDPod)
## mod1: Vcotyledon ~ Zone + (1 | UniqueGeno) + (1 | Plate) + (1 | UniqueQuadrat) +
```

```
## mod1:     (1 | Plant) + (1 | IDBox) + (1 | IDPod)
##      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod2  8 1728.8 1761.9 -856.40   1712.8
## mod1 10 1724.0 1765.4 -851.98   1704.0 8.8326      2    0.01208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significative Zone effect within the Intrapop data, we cannot remove that effect.
We will use the variable Zone0, which has the value 0 when it is within Interpop, and 3, 8 or 12 when it is within Intrapop.
This variable encompasses both the Zone and the Plan effects.

Testing whole dataset

Model simplification is performed using the LRT method and confirming using AICc.

Results

Vcotyledon ~ Block + Zone0 + (1|Geno) + (1|Plant) + (1|Box) + (1|Pod)

	Estimate	Std. Error	t value
(Intercept)	4.814	0.133	36.218
Block2	0.867	0.114	7.643
Block3	0.224	0.113	1.981
Zone03	1.223	0.182	6.699
Zone08	0.337	0.196	1.718
Zone012	0.966	0.222	4.345

Percentage of explained variance:

Effect	Block	Zone0	Geno	Plant	Pod	Box	Residuals
p-value	2.2e-16	4.93e-10	2.172e-5	3.317e-08	1.009e-07	2.186e-08	
Variance (%)			5.1042496	4.3187413	11.1101446	9.3246453	70.1420561

Residuals analysis

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles

Residuals distribution

Frequency

resid(m)

The residuals seem to be normally distributed and centered around zero.

BLUPs

BLUPs are computed for the genotype random effect:

UniqueGeno

Random effects distributions

Covariance analysis - Phenotypic compromises

Is there an effect of flowering time, of the number of pods/seeds per plant on the speed of cotyledon emergence?

Flowering time

```
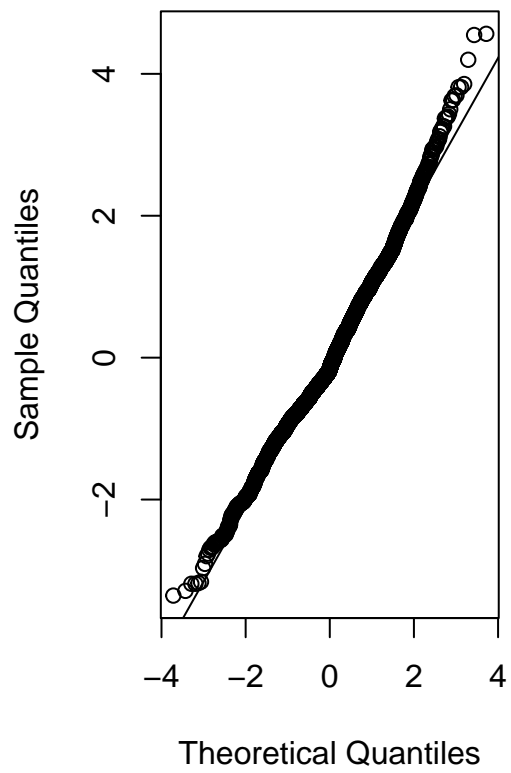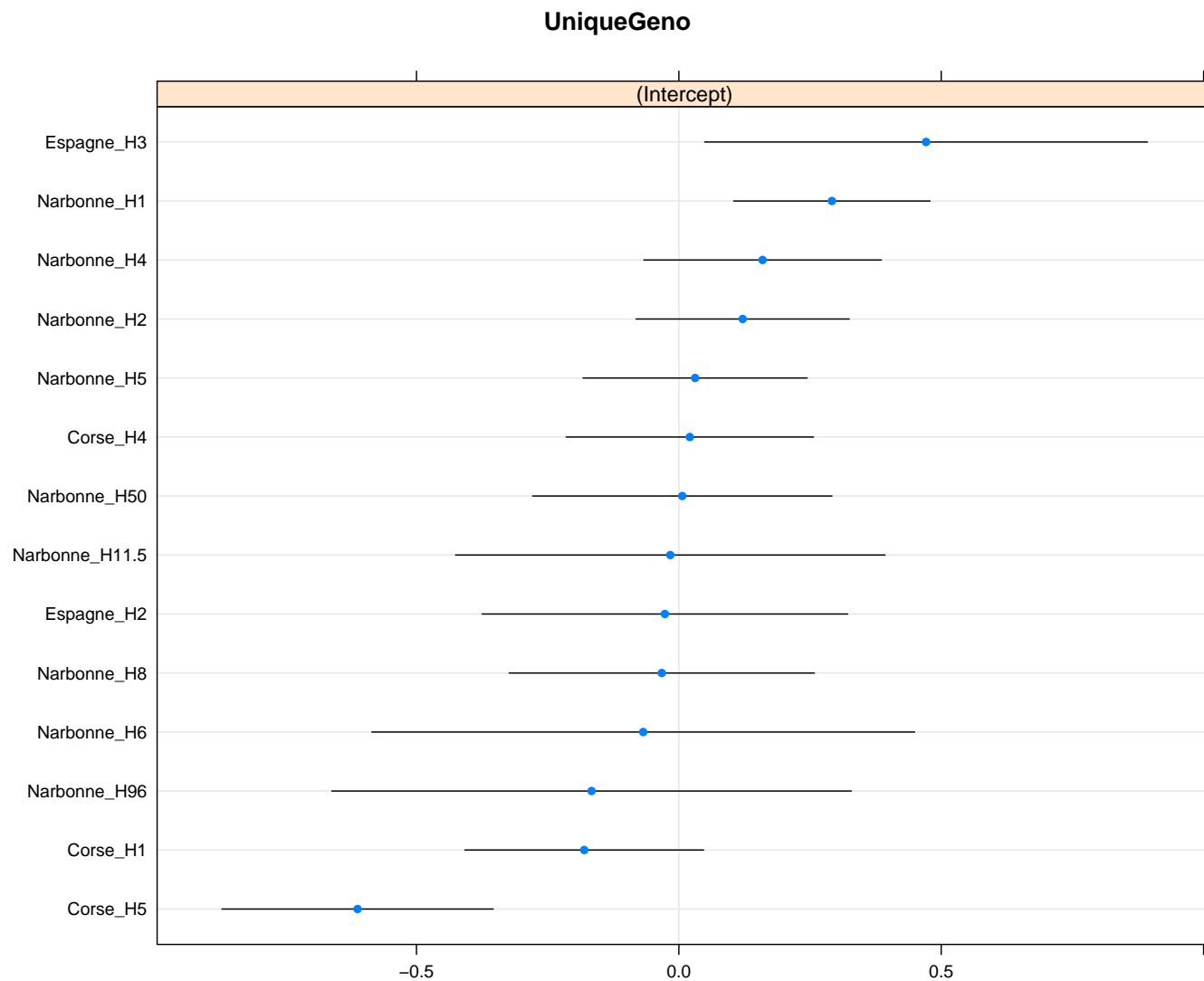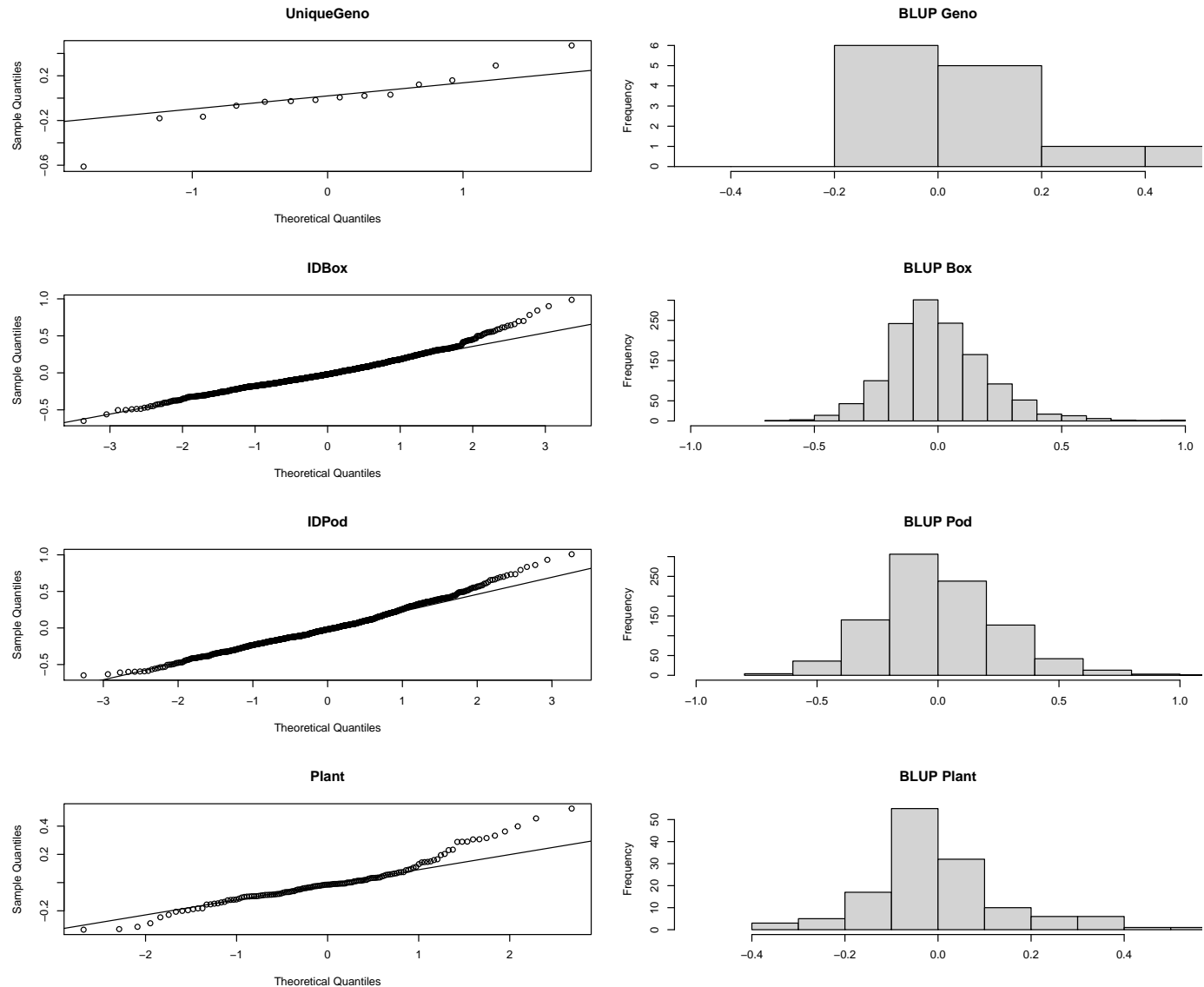## Data: dataFlowering
## Models:
## m: Vcotyledon ~ Block + Zone0 + (1 | UniqueGeno) + (1 | Plant) +
## m:     (1 | IDBox) + (1 | IDPod)
## m1: Vcotyledon ~ Block + Zone0 + time_flo + (1 | UniqueGeno) + (1 |
## m1:     Plant) + (1 | IDBox) + (1 | IDPod)
##    Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## m  11 15487 15558 -7732.4    15465
## m1 12 15485 15563 -7730.7    15461  3.31      1    0.06886 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

12

	Estimate	Std. Error	t value
(Intercept)	5.038	0.175	28.841
Block2	0.837	0.116	7.217
Block3	0.195	0.116	1.685
Zone03	1.077	0.193	5.580
Zone08	0.195	0.207	0.942
Zone012	0.915	0.224	4.094
time_flo	-0.012	0.007	-1.829

When flowering occurred late, the speed of cotyledon emergence decreases.

Number of pods per plant

```
## Data: dataPods
## Models:
## m: Vcotyledon ~ Block + Zone0 + (1 | UniqueGeno) + (1 | Plant) +
## m:      (1 | IDBox) + (1 | IDPod)
## m1: Vcotyledon ~ Block + Zone0 + TotalNbPods + (1 | UniqueGeno) +
## m1:      (1 | Plant) + (1 | IDBox) + (1 | IDPod)
##     Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   11 16240 16312 -8109.0    16218
## m1 12 16239 16317 -8107.4    16215 3.1642      1    0.07527 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	Estimate	Std. Error	t value
(Intercept)	4.695	0.145	32.335
Block2	0.903	0.114	7.895
Block3	0.260	0.114	2.278
Zone03	1.289	0.185	6.977
Zone08	0.426	0.202	2.113
Zone012	1.061	0.228	4.656
TotalNbPods	0.004	0.003	1.738

When the number of pods per plant increases, the cotyledons go out more slowly.

Number of seeds per plant

```
## Data: dataSeed
## Models:
## m: Vcotyledon ~ Block + Zone0 + (1 | UniqueGeno) + (1 | Plant) +
## m:      (1 | IDBox) + (1 | IDPod)
## m1: Vcotyledon ~ Block + Zone0 + NumSeedPlant + (1 | UniqueGeno) +
## m1:      (1 | Plant) + (1 | IDBox) + (1 | IDPod)
##     Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   11 16240 16312 -8109.0    16218
## m1 12 16236 16314 -8106.2    16212 5.5931      1    0.01803 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	Estimate	Std. Error	t value
(Intercept)	4.654	0.144	32.227

	Estimate	Std. Error	t value
Block2	0.901	0.109	8.258
Block3	0.258	0.109	2.370
Zone03	1.300	0.179	7.259
Zone08	0.451	0.198	2.280
Zone012	1.100	0.226	4.871
NumSeedPlant	0.001	0.001	2.364

When the number of seeds per plant increases, it takes a longer time for the cotyledon to emerge.

Complete model with all the covariables: **Vcot ~ Block + Zone + time_flo + NbPods + NbSeeds + (1|Geno) + (1|Plant) + (1|Box) + (1|Pod)**

	Estimate	Std. Error	t value
(Intercept)	4.562	0.238	19.130
Block2	0.831	0.111	7.500
Block3	0.189	0.110	1.714
Zone03	1.314	0.205	6.399
Zone08	0.498	0.231	2.159
Zone012	1.129	0.234	4.835
time_flo	0.004	0.008	0.428
TotalNbPods	0.006	0.007	0.857
NumSeedPlant	0.001	0.002	0.838

Normal Q–Q Plot

Sample Quantiles vs Theoretical Quantiles

Residuals distribution

Frequency vs resid(m)

Bootstrap analysis

		CI_inf	CI_sup
(Intercept)	4.562	4.061	5.039
Block2	0.831	0.633	1.022
Block3	0.189	-0.008	0.377
Zone03	1.314	0.937	1.712
Zone08	0.498	-0.005	0.968
Zone012	1.129	0.667	1.631
time_flo	0.004	-0.013	0.021
TotalNbPods	0.006	-0.006	0.021
NumSeedPlant	0.001	-0.002	0.004

grp	vcov	CI_inf	CI_sup
IDBox	0.1731731	0.1100221	0.2513653
IDPod	0.2004776	0.1085844	0.2783680
Plant	0.0525182	0.0100148	0.0974417
UniqueGeno	0.0648348	-0.0220013	0.1448212
Residual	1.2632337	1.2063105	1.3203407

Number of seeds per pod

Are there genotypes that produce more/less seeds per pod? Is there a type of environment in which plants produce more seeds per pod? Is there a maternal effect on the production of seeds?

Boxplots

Mixed linear models

Testing Zone effect within Intrapop data

```
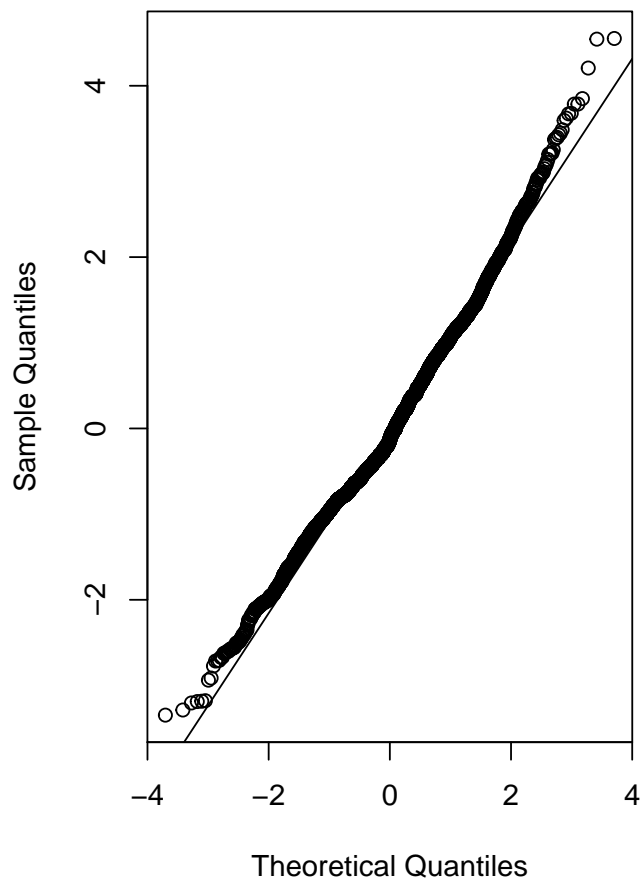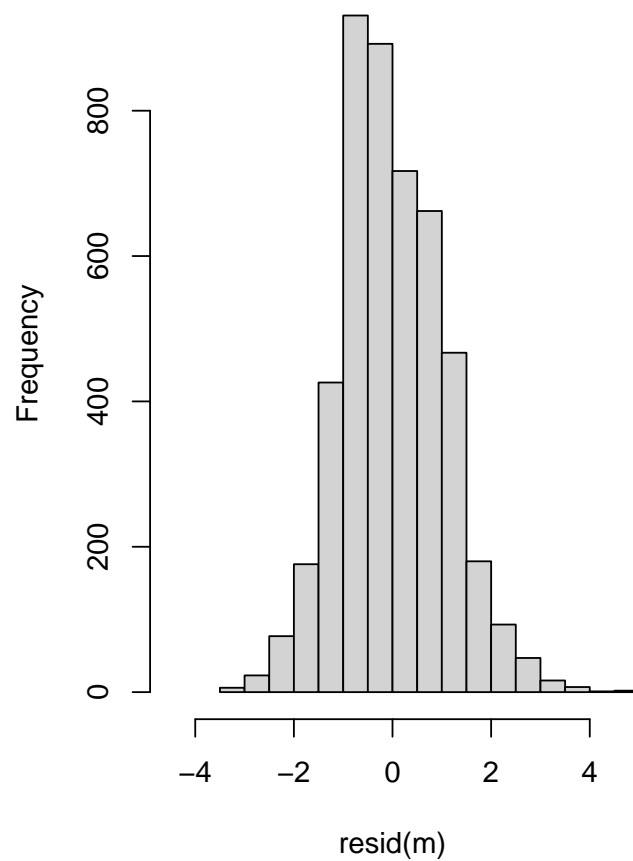## Data: seed_intrapop
## Models:
## mod2: NbSeed ~ time_flo + (1 | UniqueGeno) + (1 | UniqueQuadrat) +
## mod2:     (1 | Plant)
## mod1: NbSeed ~ Zone + time_flo + (1 | UniqueGeno) + (1 | UniqueQuadrat) +
## mod1:     (1 | Plant)
##      Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## mod2  6 418.02 434.38 -203.01   406.02
## mod1  8 418.72 440.54 -201.36   402.72 3.297      2     0.1923
```

There is no Zone effect within the intrapop dataset.

Testing the whole dataset

Results

Model selected: **NbSeed ~ Plan + time_flo + (1|Quadrat) + (1|Plant)**

	Estimate	Std. Error	t value
(Intercept)	6.863	0.289	23.750
PlanIntrapop	-1.980	0.349	-5.666
time_flo	-0.052	0.011	-4.627

Percentage of explained variance :

Effect	Plan	time_flo	Plant	Quadrat	Residuals
p-value	2.773e-5	1.063e-5	0.004367	9.498e-5	
Variance (%)			7.97	8.62	83.4

Plants located in Intrapop produce on average 1.98 seeds less than plants located in interpop (6.86 seeds per pod on average).

residuals analysis

Normal Q–Q Plot

Residuals distribution

BLUPs

BLUPs were computed for the random effects of the model:

Plant

BLUP Plant

Quadrat

BLUP Quadrat

Bootstrap

Confidence intervals (CI) of the estimated parameters (slopes and variances) are computed using the bootstrap method.

		CI_inf	CI_sup
(Intercept)	6.863	6.310	7.433
PlanIntrapop	-1.980	-2.657	-1.321
time_flo	-0.052	-0.074	-0.029

grp	vcov	CI_inf	CI_sup
Plant	0.2177181	0.0509046	0.3791522
UniqueQuadrat	0.2354769	-0.0619754	0.5308335
Residual	2.2771126	2.0511885	2.5056994

The value 0 is comprised in the confidence interval of UniqueQuadrat estimation of variance

Number of seeds per plant

Is there a genetic effect on the amount of seeds produced per plant? Is there an environmental effect? Does flowering date impact the amount of seeds produced per plant?

Boxplots

Mixed linear models

Testing Zone effect within Intrapop data

```
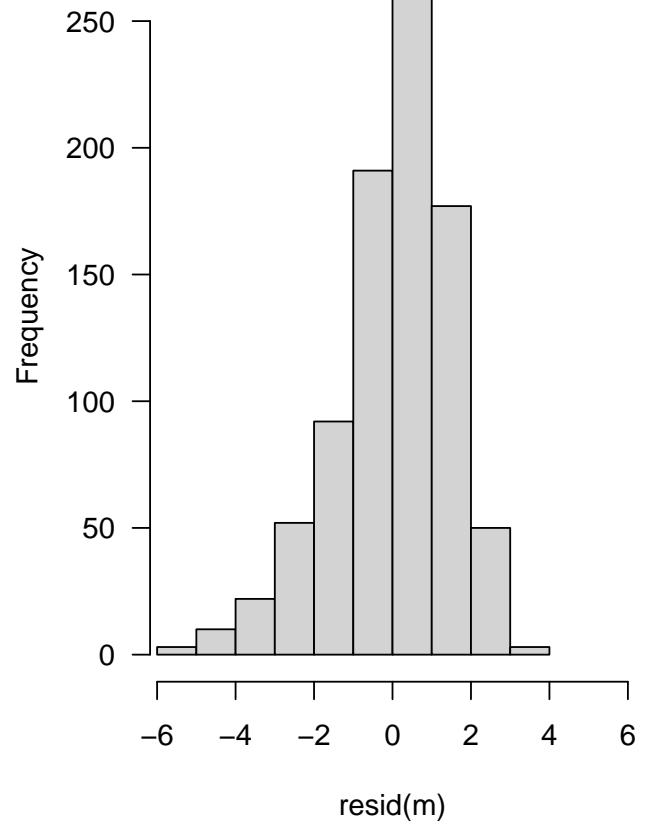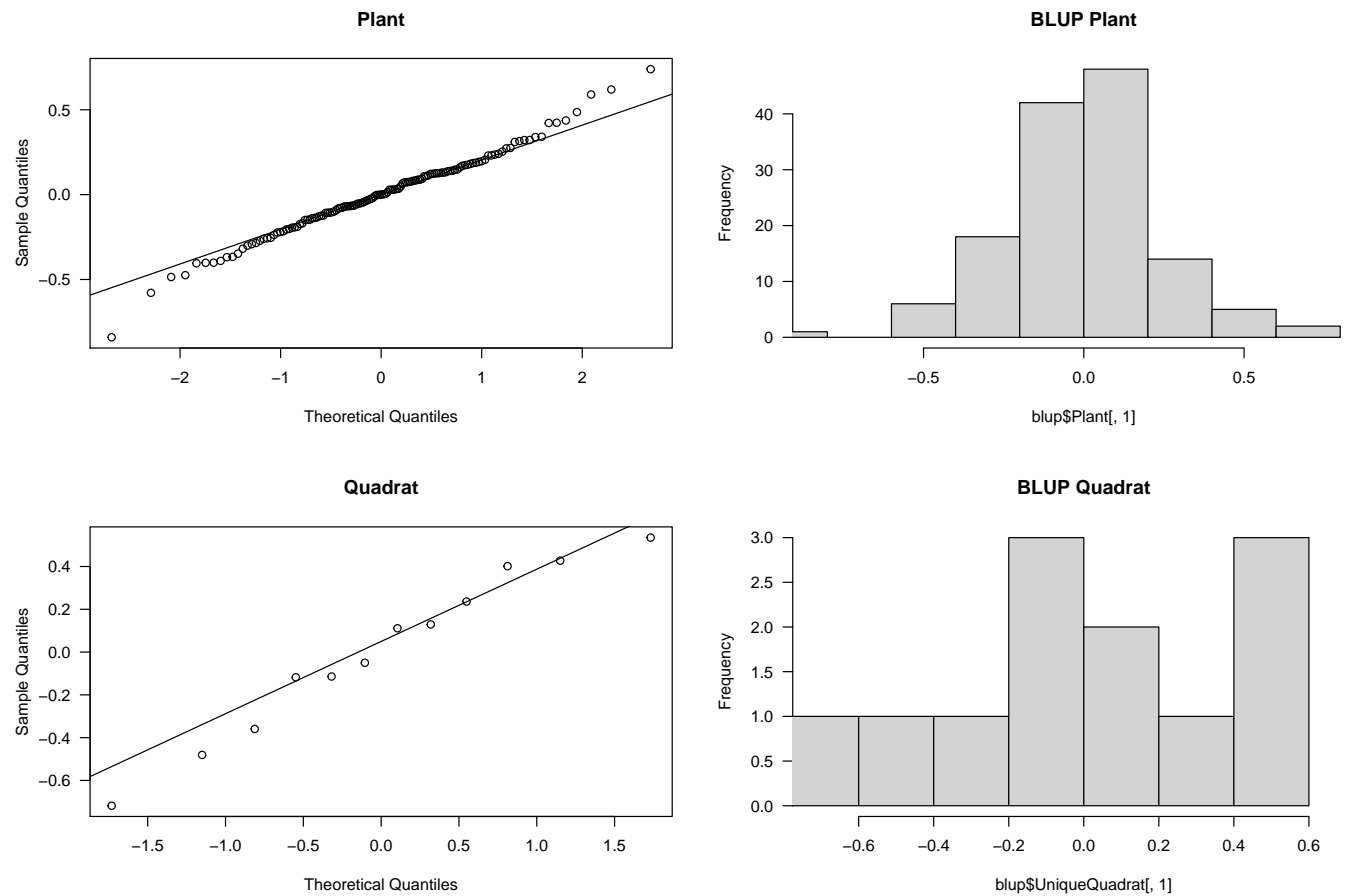## Data: seedPlant_intra
## Models:
## mod2: NbSeedPlant ~ (1 | UniqueGeno) + (1 | UniqueQuadrat)
## mod1: NbSeedPlant ~ Zone + (1 | UniqueGeno) + (1 | UniqueQuadrat)
##      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod2  4 407.70 415.58 -199.85   399.70
## mod1  6 407.67 419.50 -197.84   395.67 4.0261      2     0.1336
```

There is no significative Zone effect within the intrapop dataset (which comprises 53 plants instead of 143 for the whole dataset).

Testing the whole dataset

Results

Model selected: **NbSeed ~ Plan + (0 + Plan|UniqueGeno)**
The genetic variance is different in intrapop and in interpop.

	Estimate	Std. Error	t value
(Intercept)	70.448	14.538	4.846
PlanIntrapop	-60.734	15.662	-3.878

Percentage of explained variance :

Effect	Plan	(1 + Plan:Geno)	Residuals
p-value	0.0012	0.002774	
Variance (%)		Interpop: 53.26 Intrapop: 0	46.74

Correlations

Flowering time

```
## Data: seedPlant
## Models:
## m: NbSeedPlant ~ Plan + (0 + Plan | UniqueGeno)
## m1: NbSeedPlant ~ Plan + time_flo + (0 + Plan | UniqueGeno)
##    Df    AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## m   6 1422.6 1440 -705.30   1410.6
## m1  7 1380.7 1401 -683.35   1366.7 43.89      1  3.473e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	Estimate	Std. Error	t value
(Intercept)	132.267	15.164	8.722
PlanIntrapop	-81.505	13.603	-5.992
time_flo	-3.409	0.478	-7.123

There is a significative effect of flowering time on the number of seeds produced per plant.

Number of pods per plant

```
## Data: seedPlant
## Models:
## m: NbSeedPlant ~ Plan + time_flo + (0 + Plan | UniqueGeno)
## m1: NbSeedPlant ~ Plan + time_flo + TotalNbPods + (0 + Plan | UniqueGeno)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   7 1380.7 1401.0 -683.35   1366.7
## m1  8 1179.0 1202.3 -581.52   1163.0 203.66      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	Estimate	Std. Error	t value
(Intercept)	3.326	13.956	0.238
PlanIntrapop	-1.101	12.958	-0.085
time_flo	-0.308	0.254	-1.216
TotalNbPods	4.370	0.196	22.308

There is a significative effect of the number of pods per plant on the number of seeds produced.

Percentage of explained variance :

|Effect | Plan | (1 + Plan:Geno) | Residuals| |:——|——:|——:|——:|
|p-value | 0.0012 | 0.002774 | | |Variance (%) | |Interpop: 86.14 Intrapop: 0.01|13.85|

There is far less variability in Intrapop than in Interpop.

Residuals analysis

Normal Q–Q Plot

Residuals distribution


```
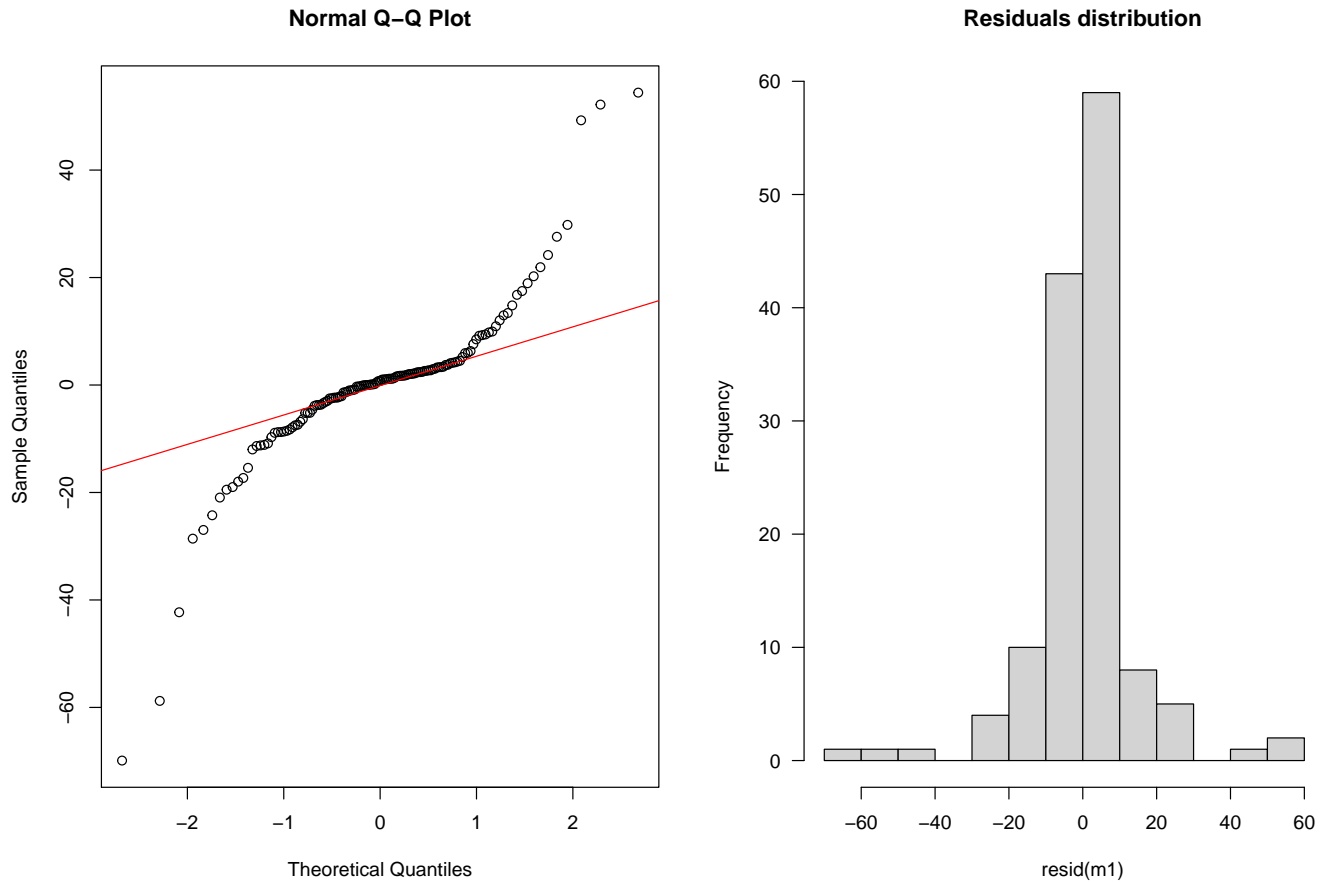##
##   Shapiro-Wilk normality test
##
## data:  resid(m1)
## W = 0.82869, p-value = 3.006e-11
```

Bootstrap

Confidence intervals are computed for the estimations of the model (slopes and variances) using bootstrap method.

		CI_inf	CI_sup
(Intercept)	3.326	-23.254	30.091
PlanIntrapop	-1.101	-25.828	23.238
time_flo	-0.308	-0.809	0.208
TotalNbPods	4.370	3.986	4.771

grp	vcov	CI_inf	CI_sup
UniqueGeno	1536.1029560	58.90518	2969.64038
UniqueGeno	0.1502108	-55.48206	27.46251
UniqueGeno	-15.1901056	-240.47516	218.59765
Residual	247.0583934	189.49655	315.00703

BLUPs

What are the effects of each genotype in the different environments (intrapop and interpop)?

All genotypes are not represented in each environment: there are only the genotypes from Narbonne in Intrapop and there are all the genotypes except Narb_H11.5, Narb_H6 and Narb_H96 in Interpop.

It is difficult to compare the effects of the genotypes in the two environments, as they are not necessarily repeated.

- Narbonne_H8: 2.99 in interpop vs. -0.03 in intrapop

- Narbonne_H50: 28.96 in interpop vs. -0.29 in intrapop

- Narbonne_H5: 0.76 in interpop vs. -0.01 in intrapop

- Narbonne_H4: 7.10 in interpop vs. -0.07 in intrapop

- Narbonne_H2: -1.32 in interpop vs. 0.01 in intrapop

- Narbonne_H1: 22.62 in interpop vs. -0.22 in intrapop

It seems that the genotype Espagne_H3 produces far less seeds than the other genotypes in Interpop, but this genotype is represented by only one plant.

Genotype qqplots:

Fixed effects correlation:

Seed viability analysis

(Analysis with data pooled by box -> proportion of viable seed in each box)

We perform model simplification looking at the effects of different parameters on seed viability's variance, then covariables are added to the simplified model :

- NbTotalPods: is there a compromise between the number of pods per mother plant and the viability of the seeds within?

- time_flo: is there a linear relationship between seed viability and the flowering time of the mother plant?

The viability data is coded as binomial data, we use GLMM for the analysis. Moreover, several link functions can be used, we will test them (using AICc) and keep the one that best fits the data to perform the model simplification.

Link function choice

The link functions are tested using AICc because all models will have the same number of degrees of freedom and so LRT is impossible.

```
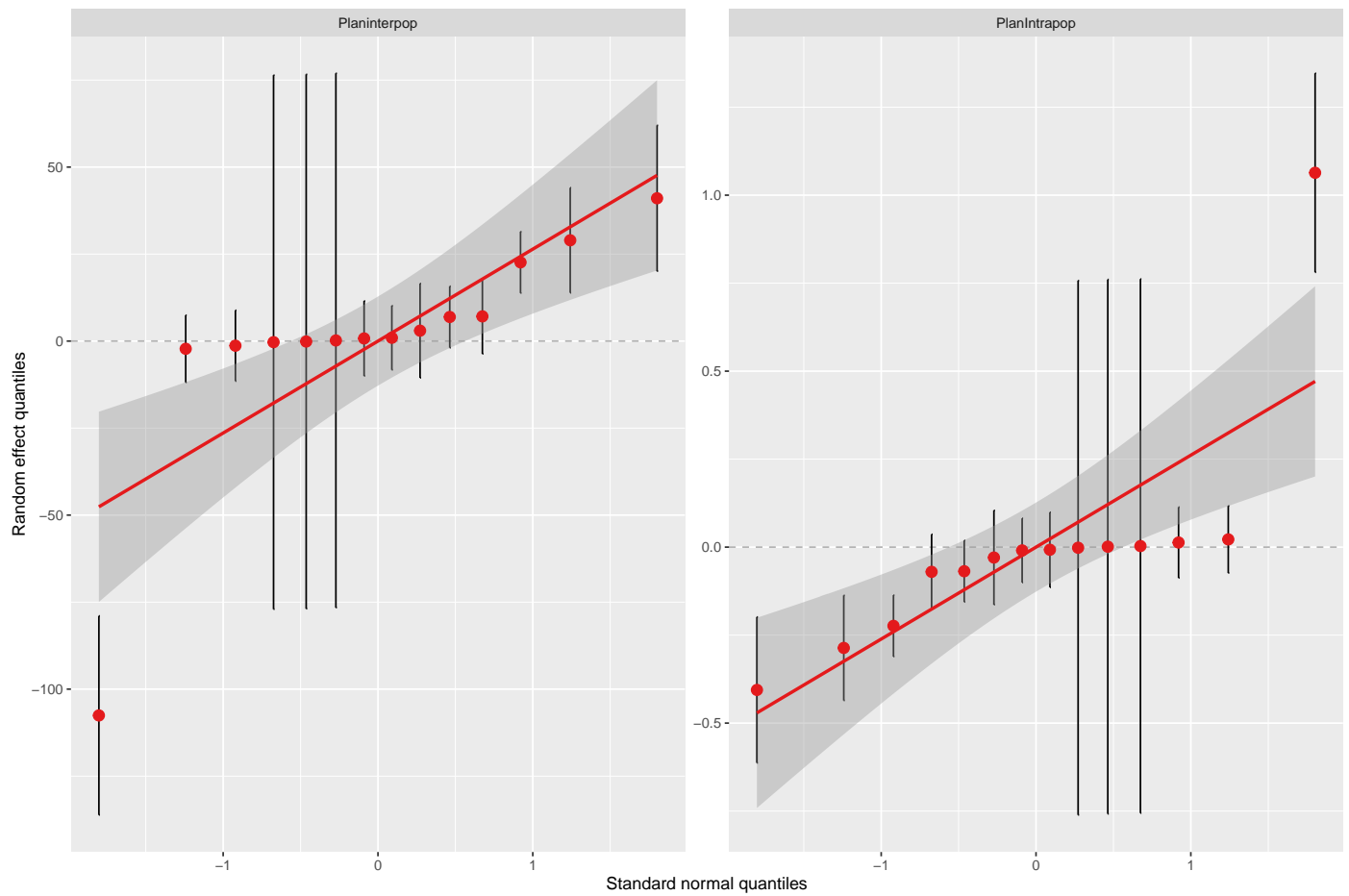## Model selection table
##         (Intrc) Block Plan family df    logLik   AICc delta weight
## cauchit  1.2800     +    + bn(cc) 13 -1807.404 3641.1  0.00      1
## logit    1.2100     +    +     bn 13 -1816.902 3660.1 19.00      0
## probit   0.7130     +    + bn(pr) 13 -1825.748 3677.8 36.69      0
## cloglog  0.3248     +    + bn(cl) 13 -1850.995 3728.3 87.18      0
```

```
## Abbreviations:
## family: bn = 'binomial', bn(cc) = 'binomial(cauchit)',
##           bn(cl) = 'binomial(cloglog)', bn(pr) = 'binomial(probit)'
## Models ranked by AICc(x)
## Random terms (all models):
## '1 | IDBox', '1 | IDPod', '1 | Plant', '1 | UniqueQuadrat', '0 + Plan | UniqueGeno', '1 | Plate', '1 | Blo
```

The Cauchit link function is better adapted than the Logit link. We will use the Cauchit link to simplify the model but the Logit link to compute heritability (not defined for Cauchit).

Model simplication

Zone effect within intrapop

Model selection is performed using the LRT method and verifying with AICc.

```
## Data: box_intrapop
## Models:
## m1: propViabilityNA ~ (1 | IDBox) + (1 | IDPod) + (1 | Plant) + (1 |
## m1:     UniqueQuadrat) + (1 | UniqueGeno) + (1 | Plate) + (1 | Plate)
## m0: propViabilityNA ~ Zone + (1 | IDBox) + (1 | IDPod) + (1 | Plant) +
## m0:     (1 | UniqueQuadrat) + (1 | UniqueGeno) + (1 | Plate) + (1 |
## m0:     Plate)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m1  8 359.21 381.90 -171.61   343.21
## m0 10 358.38 386.75 -169.19   338.38 4.8287      2    0.08943 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


## Model selection table
##    (Intrc) Zone df   logLik  AICc delta weight
## m0  0.5648    + 10 -169.192 360.3  0.00  0.518
## m1  0.9902      8 -171.607 360.4  0.15  0.482
## Models ranked by AICc(x)
## Random terms (all models):
## '1 | IDBox', '1 | IDPod', '1 | Plant', '1 | UniqueQuadrat', '1 | UniqueGeno', '1 | Plate'
```

There is a marginally significative Zone effect within the intrapop dataset (p-val = 0.08943). We use the variable Zone0 for the simplification with the whole dataset, as it encompasses the information within Zone and Plan.

Whole dataset analysis

Results

Model selection: **SeedViability ~ (1|Box) + (1|Pod) + (1|Plant) + (1|Block:Plate)**

Model simplification using the logit link results in the same best model (with AICc = 3646.737 instead of 3627.677), we can then use this link function to compute R^2 and repeatability for the model (not defined with cauchit link). Observations are clustered within Petri boxes, and boxes are nested within pods (themselves nested within mother plants)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( cauchit )
## Formula: propViabilityNA ~ (1 | IDBox) + (1 | IDPod) + (1 | Plant) + (1 |
##      Block:Plate)
##    Data: box
## Weights: NbSeedminusNA
```

```
##
##      AIC      BIC   logLik deviance df.resid
##   3627.6   3653.7  -1808.8   3617.6     1354
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -1.8761 -0.3572  0.1809  0.8110  1.3078
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  IDBox        (Intercept) 0.4285   0.6546
##  IDPod        (Intercept) 1.0399   1.0197
##  Plant        (Intercept) 0.9168   0.9575
##  Block:Plate  (Intercept) 0.3304   0.5748
## Number of obs: 1359, groups:
## IDBox, 1359; IDPod, 962; Plant, 143; Block:Plate, 21
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.2909     0.1804   7.157 8.25e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Mean viability = 0.7752713

Percentage of explained variance with logit link:

Effect	Box	Pod	Plant	Block:Plate
p-value	0.002293	2.261e-09	2.2e-16	8.305e-10
Variance (%)	12.63	42.29	32.24	12.85

Bootstrap

grp	vcov	CI_inf	CI_sup
IDBox	0.2911616	0.2255135	0.6780031
IDPod	0.9749219	0.7612417	1.3119057
Plant	0.7431849	0.5119047	1.0928218
Block:Plate	0.2961805	0.0800223	0.5673272

Intercept: 1.2383194 [0.9350477 - 1.5735991]

Percentage of explained variance with cauchit link:

Effect	Box	Pod	Plant	Block:Plate
p-value	0.002293	2.261e-09	2.2e-16	8.305e-10
Variance (%)	15.78	38.29	33.76	12.17

Covariable analysis

Flowering time of the mother plant

Model comparison

We test the fixed effect flowering time in the simplified model we obtained before (with logit link).

```
## Data: boxNA
## Models:
## m0: propViabilityNA ~ (1 | IDBox) + (1 | IDPod) + (1 | Plant) + (1 |
## m0:     Block:Plate)
## m1: propViabilityNA ~ time_flo + (1 | IDBox) + (1 | IDPod) + (1 |
## m1:     Plant) + (1 | Block:Plate)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m0  5 3456.4 3482.2 -1723.2   3446.4
## m1  6 3452.8 3483.8 -1720.4   3440.8 5.6289      1    0.01767 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significative effect of flowering time on the probability that a seed is viable (p-val = 0.01767).

Results

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.816	0.292	6.221	0.000
time_flo	-0.035	0.015	-2.418	0.016

The new mean in the proportion of viable seeds per pod is 0.8600289, and the slope according to flowering time is -0.0043. This means that when flowering is late, there are less viable seeds produced.

R² computation

Marginal $(R^2_{GLMM}(m))$ and conditional$(R^2_{GLMM}(c))$ R² are computed using the method from Nakagawa and Schielzeth (2013).

Marginal R² (variance explained by fixed factors):

$$R^2_{GLMM}(m) = \frac{\sigma^2_f}{\sigma^2_f + \sum_{l=1}^u \sigma^2_l + \sigma^2_e + \frac{\pi^2}{3}}$$

where

- σ^2_f is the variance calculated from the fixed effect components of the GLMM

- σ^2_l is the variance of the lth random factor (with u the number of random factors in the GLMM)

- σ^2_e is the additive dispersion component (residual variance, so in our model it corresponds to the variance of IDBox)

- $\frac{\pi^2}{3}$ is the distribution specific variance (for binary and proportion data under logit link)

Conditional R² (variance explained by both fixed and random factors):

$$R^2_{GLMM}(c) = \frac{\sigma^2_f + \sum_{l=1}^u \sigma^2_l}{\sigma^2_f + \sum_{l=1}^u \sigma^2_l + \sigma^2_e + \frac{\pi^2}{3}}$$

$R^2_{GLMM}(m) = 0.0115083$
-> there is only 1.2% of variance that is explained by flowering time

$R^2_{GLMM}(c) = 0.3685001$
The whole model explains 36.9% of the variance in seed viability.

Residual analysis

Number of pods per plant

Is there a correlation between the number of pods produced by a plant and the viability of the seeds?

```
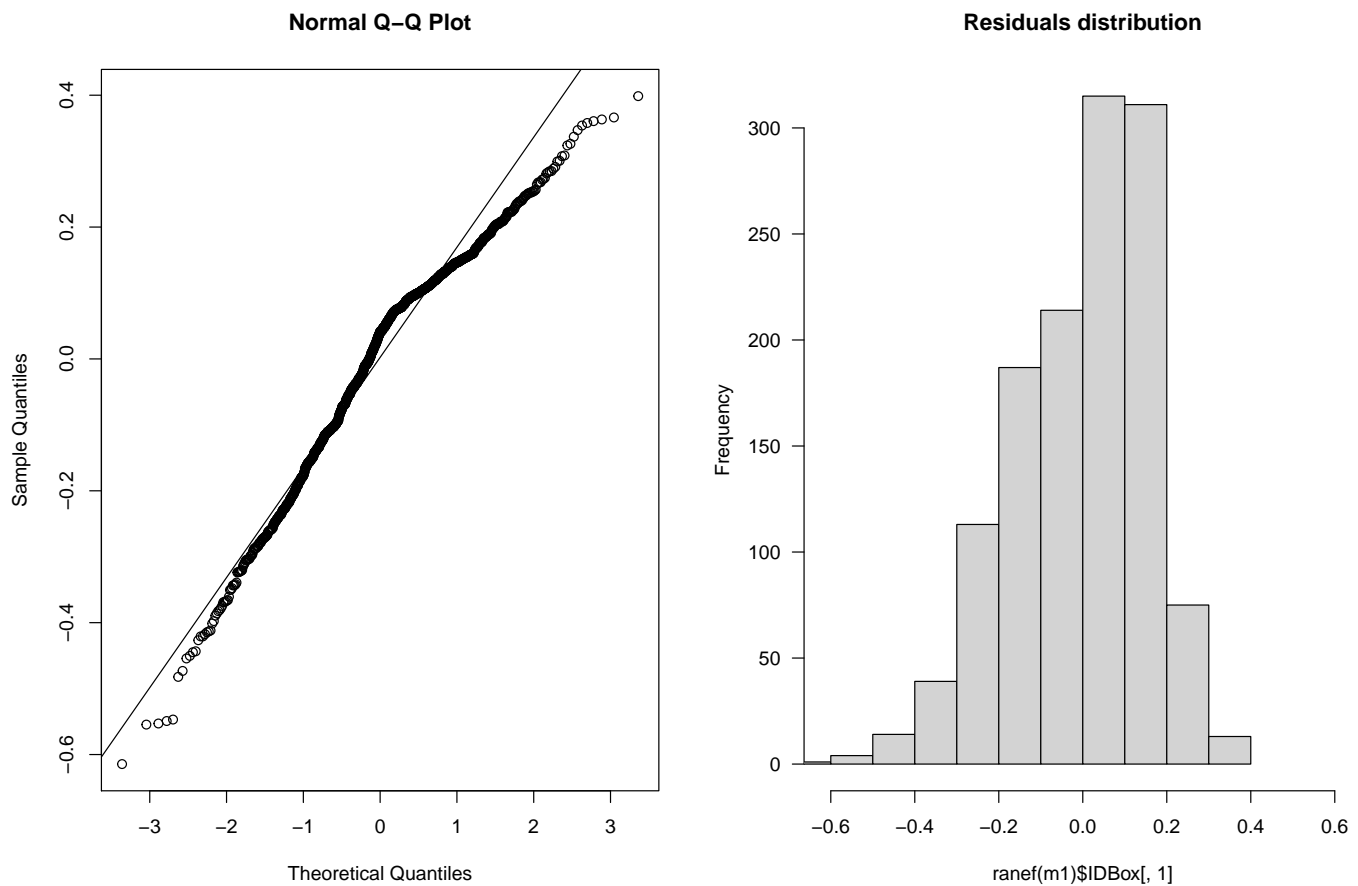## Data: NbPods
## Models:
## m0: propViabilityNA ~ (1 | IDBox) + (1 | IDPod) + (1 | Plant) + (1 |
## m0:     Block:Plate)
## m1: propViabilityNA ~ TotalNbPods + (1 | IDBox) + (1 | IDPod) + (1 |
## m1:     Plant) + (1 | Block:Plate)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m0  5 3597.5 3623.6 -1793.8   3587.5
## m1  6 3599.5 3630.7 -1793.8   3587.5 0.0107      1     0.9177
```

There is no significative effect of the total number of pods per plant on seed viability.

Number of seeds per plant

```
## Data: boxNumSeed
## Models:
## m0: propViabilityNA ~ (1 | IDBox) + (1 | IDPod) + (1 | Plant) + (1 |
## m0:     Block:Plate)
## m1: propViabilityNA ~ NumSeedPlant + (1 | IDBox) + (1 | IDPod) +
## m1:     (1 | Plant) + (1 | Block:Plate)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m0  5 3646.7 3672.8 -1818.3   3636.7
## m1  6 3648.3 3679.6 -1818.2   3636.3 0.3763      1     0.5396
```

There is no significative effect of the number of seeds per plant on seed viability.

Number of seeds per pod

```
## Data: box
## Models:
## m0: propViabilityNA ~ (1 | IDBox) + (1 | IDPod) + (1 | Plant) + (1 |
## m0:     Block:Plate)
## m1: propViabilityNA ~ NbSeed + (1 | IDBox) + (1 | IDPod) + (1 | Plant) +
## m1:     (1 | Block:Plate)
##    Df   AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m0  5 3646.7 3672.8 -1818.3   3636.7
## m1  6 3645.8 3677.1 -1816.9   3633.8 2.8951      1    0.08885 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a marginally significative effect of the number of seed per pods on seed viability.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.880	0.260	3.381	0.001
NbSeed	0.082	0.047	1.752	0.080

The mean in the proportion of viable seeds per pod is 0.706748, and the slope of the number of seed per pod is 0.0167262. This means that when the number of seeds per pod increases, the viability of those seeds increases too.

Dormancy ending analysis

Seeds are considered non-dormant if they swelled without needing scarification.

Link function choice

The logit link function best fits the data.

Results

Selected model: **Dormancy ~ Plan + (1|Pod) + (1|Plant)**

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.428	0.093	-26.063	0
PlanIntrapop	1.811	0.176	10.263	0

The mean proportion of non-dormant seeds in interpop is 0.081, while it is 0.35 in intrapop.

Percentage of variance explained:

Effect	Plan	Pod	Plant
p-value	0.002293	2.261e-09	2.2e-16
Variance		0.630512	0.1893725

BLUP analysis

Covariance analysis

Flowering time

```
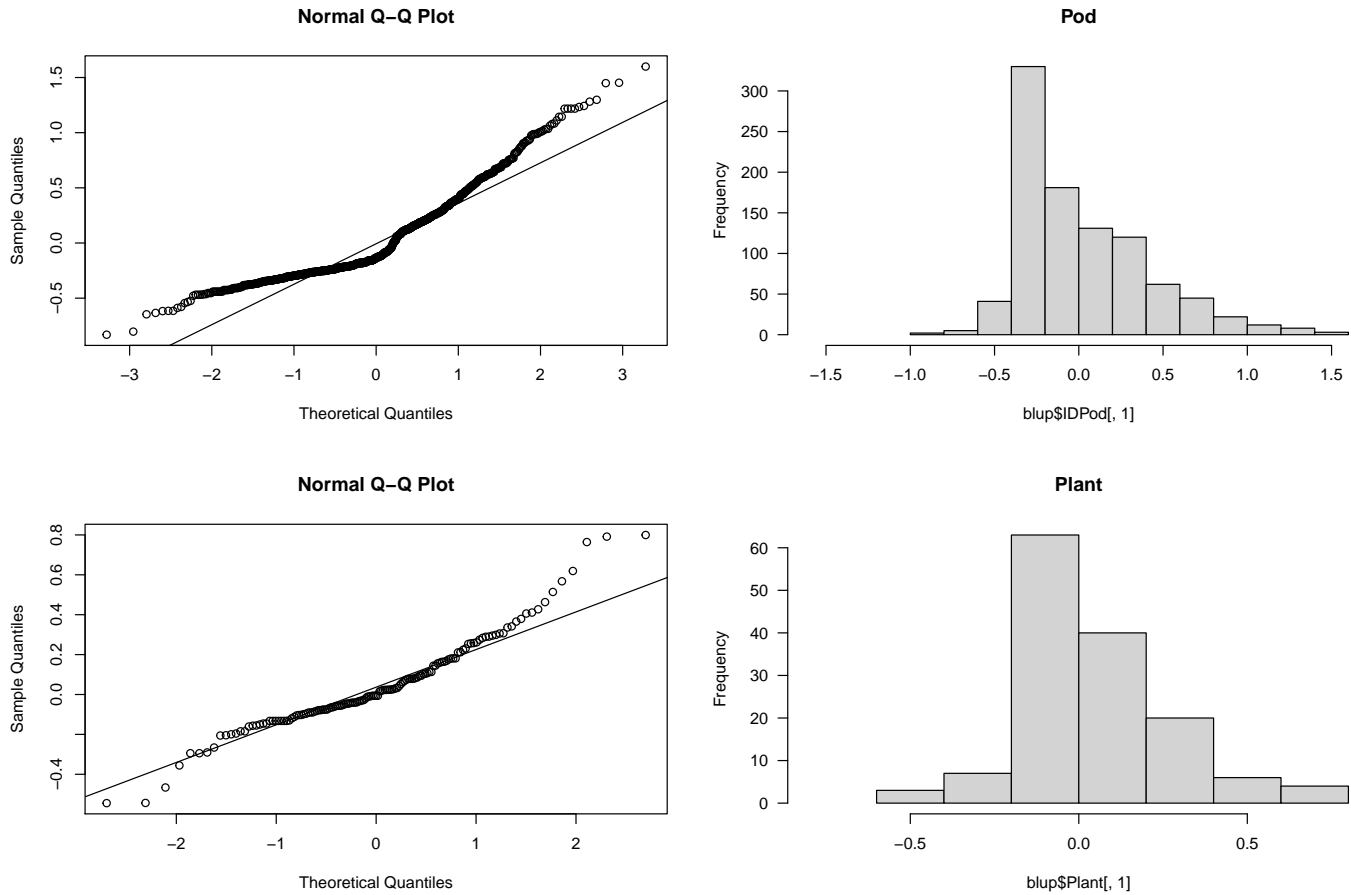## Data: boxFlo
## Models:
## m: Dormancy ~ Plan + (1 | IDPod) + (1 | Plant) + (1 | IDBox)
## m1: Dormancy ~ Plan + time_flo + (1 | IDPod) + (1 | Plant) + (1 |
## m1:     IDBox)
##   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   5 2361.3 2387.2 -1175.7    2351.3
## m1  6 2363.3 2394.3 -1175.7    2351.3 0.0114      1      0.9149
```

There is no significative effect of the flowering date on the end of dormancy of the seeds.

Number of pods per plant

```
## Data: boxPods
## Models:
## m: Dormancy ~ Plan + (1 | IDPod) + (1 | Plant) + (1 | IDBox)
## m1: Dormancy ~ Plan + TotalNbPods + (1 | IDPod) + (1 | Plant) + (1 |
## m1:     IDBox)
##   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   5 2471.6 2497.6 -1230.8    2461.6
## m1  6 2473.1 2504.3 -1230.5    2461.1 0.5069      1      0.4765
```

There is no significative effect of the number of pods per plant on the end of dormancy.

Number of seeds per plant

```
## Data: boxSeed
## Models:
## m: Dormancy ~ Plan + (1 | IDPod) + (1 | Plant) + (1 | IDBox)
## m1: Dormancy ~ Plan + NumSeedPlant + (1 | IDPod) + (1 | Plant) +
## m1:     (1 | IDBox)
##    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## m   5 2492.8 2518.9 -1241.4   2482.8
## m1  6 2494.6 2525.9 -1241.3   2482.6 0.1662      1     0.6835
```

There is no significative effect of the number of seeds per plant on the end of dormancy.

R² computation

Marginal ($R^2_{GLMM}(m)$) and conditional($R^2_{GLMM}(c)$) R² are computed using the method from Nakagawa and Schielzeth (2013).

$R^2_{GLMM}(m) = 0.0626904$
-> there is only 6.3% of variance that is explained by Plan

$R^2_{GLMM}(c) = 0.2496811$
The whole model explains 25% of the variance in seed dormancy ending.