DGRP Genetic Arch

Alexander Shingleton

9/8/2020

#Outline of questions

* Section 1: SSD in fed flies
  + Question 1.1: Is there Sexual Size Dimorphism in the DGRP lines.
  + Question 1.2: Do we see a genetic variation of SSD in the DGRP lines.
  + Question 1.3: Is the SSD genetic variation due to a genetic variation in male or in female size?
* Section 2: SSD in other conditions
  + Question 2.1: Do we have the same SSD when we change environment?
  + Question 2.2: Does SSD increase or decrease when the flies are starved? Our hypothesis is that overall SSD should decrease
  + Question 2.3: Does the SSD in starved conditions vary the same way as in fed flies?

If SSD differs between environmental conditions, that means the female and male plasticity are different.

-Section 3: SSP

- Question 3.1: Is there SSP when we look at fed vs starved flies? - Question 3.2: Does this SSP vary in the DGRP flies?

- Question 3.3: Is the variation in SSP due to a variation in SSD in fed or starved flies? If SSP covaries with SSD0, that means that we have variation in fed flies. If SSP covaries with starved flies, we have variation in the reduction of flies size.

- Question 3.4: If it is due to SSD in either fed or starved flies, is it because the females vary more or the males vary more in size?

- Question 3.5: Is the covariation with SSP (identified in Question 3.4) the same as the covariation with SSD identified in Question 1.3?

#Data preparation

#dataset was manually relabeled and control lines removed  
df <- read.csv("~/Dropbox/\_Github\_reps/DGRP\_SSDSSP/Data/DGRPfinal\_clean\_first\_ctrl\_only.csv")  
  
#column names  
names(df)

## [1] "id" "line" "block" "day" "sex" "wing" "leg" "pupa"

#add character in front of line, day and block  
df$line <- sprintf('L%i', df$line)  
df$day<-sprintf('D%i', df$day)  
df$block<-sprintf('B%i', df$block)  
#summary of data  
summary(df)

## id line block day   
## Length:21410 Length:21410 Length:21410 Length:21410   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## sex wing leg pupa   
## Length:21410 Min. :11.29 Min. :12.95 Min. :13.82   
## Class :character 1st Qu.:13.79 1st Qu.:13.52 1st Qu.:14.28   
## Mode :character Median :13.93 Median :13.66 Median :14.42   
## Mean :13.93 Mean :13.65 Mean :14.41   
## 3rd Qu.:14.07 3rd Qu.:13.79 3rd Qu.:14.55   
## Max. :14.55 Max. :14.38 Max. :15.14   
## NA's :638 NA's :828 NA's :2136

## Filtering out lines with few data

#first, create a column to indicate group by line\_sex\_day. This will be easier to filter out or subset by group later  
df$group<- paste(df$line, df$sex, df$day, sep = "\_")  
head(df)

## id line block day sex wing leg pupa group  
## 1 365-A-0-F-031 L365 B5 D0 F 14.17346 13.81809 NA L365\_F\_D0  
## 2 365-A-1-M-045 L365 B5 D1 M 13.70294 13.46277 NA L365\_M\_D1  
## 3 712-A-0-F-013 L712 B5 D0 F 14.05506 13.77982 NA L712\_F\_D0  
## 4 774-B-0-F-055 L774 B5 D0 F 14.23619 13.79753 NA L774\_F\_D0  
## 5 774-B-2-F-094 L774 B5 D2 F 13.91447 13.53212 NA L774\_F\_D2  
## 6 317-B-1-F-7742 L317 B11 D1 F 13.70913 13.35505 13.817 L317\_F\_D1

#KEEP THIS HERE FOR NOW

#filtering out groups (line x sex x day) that have less than 10 flies  
#NB: check if I run that filter after na omit before each analysis might not be better?  
df\_sub<-df%>%  
group\_by(group) %>%  
filter(n() >=10)

## Block effect?

Flies of lineages were collected in different blocks, periods of time. Control lineages were selected and collected repeatedly for each block so that if there is variation, we can account for that factor.

Look at the control lineages to see whether there is variation among blocks

#removing all groups with less than 10 flies  
df\_sub<-df%>%  
group\_by(group) %>%  
filter(n() >=10)   
  
df\_sub<-na.omit(df\_sub,cols=pupa) #195 lines, and 17147 rows  
  
blocktest<-lm(data=df\_sub, pupa~block)  
summary(blocktest)

##   
## Call:  
## lm(formula = pupa ~ block, data = df\_sub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.62976 -0.12854 0.00592 0.13201 0.66806   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 14.470404 0.004891 2958.772 < 2e-16 \*\*\*  
## blockB11 -0.046127 0.006828 -6.756 1.46e-11 \*\*\*  
## blockB12 -0.001570 0.006335 -0.248 0.80421   
## blockB13 -0.016644 0.006794 -2.450 0.01430 \*   
## blockB4 -0.021145 0.006551 -3.228 0.00125 \*\*   
## blockB5 -0.033760 0.006080 -5.552 2.86e-08 \*\*\*  
## blockB6 -0.098174 0.006319 -15.537 < 2e-16 \*\*\*  
## blockB7 -0.101287 0.006202 -16.332 < 2e-16 \*\*\*  
## blockB8 -0.062617 0.006680 -9.374 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1849 on 17138 degrees of freedom  
## Multiple R-squared: 0.0385, Adjusted R-squared: 0.03806   
## F-statistic: 85.79 on 8 and 17138 DF, p-value: < 2.2e-16

#move residuals to dataframe  
df\_sub$pupa\_noblock <- blocktest$resid  
head(df\_sub)

## # A tibble: 6 x 10  
## # Groups: group [6]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 317-B-1-F-7742 L317 B11 D1 F 13.7 13.4 13.8 L317\_F\_… -0.607  
## 2 837-B-2-F-1476 L837 B13 D2 F 13.8 13.4 13.8 L837\_F\_… -0.630  
## 3 911-B-1-M-096 L911 B7 D1 M 13.3 13.0 13.8 L911\_M\_… -0.537  
## 4 038-B-2-F-6287 L38 B11 D2 F 13.9 13.5 13.8 L38\_F\_D2 -0.583  
## 5 796-A-1-M-034 L796 B7 D1 M 13.6 13.6 13.8 L796\_M\_… -0.526  
## 6 105-A-1-M-7585 L105 B11 D1 M 13.5 13.4 13.8 L105\_M\_… -0.581

#df\_sub$pupanoblock is the column I will use when I want to calculate summary and indexes for correlation analyses

There is a difference between collecting blocks so we have to account for block as a random factor, or adjust the values so there is no variation between blocks.

# SSD in fed flies

* Question 1.1: Is there Sexual Size Dimorphism in the DGRP lines. A1.1: YES
* Question 1.2: Do we see a genetic variation of SSD in the DGRP lines. A1.2: YES
* Question 1.3: Is the SSD genetic variation due to a genetic variation in male or in female size?

## Question 1.1: Is there Sexual Size Dimorphism in the DGRP flies that are normally fed?

To test if we have sexual size dimorphism, we want to test the effect of size with line and block as random factor. If there is a variation due to sex, that means SSD is present.

df<-df\_sub #do not forget to reload original csv data if I want to filter out things differently  
  
#subsetting day 0, fed flies  
df0<-subset(df, day=="D0")   
  
#na.omit only if pupa has NA  
df0<-na.omit(df0, cols="pupa") #8065 rows

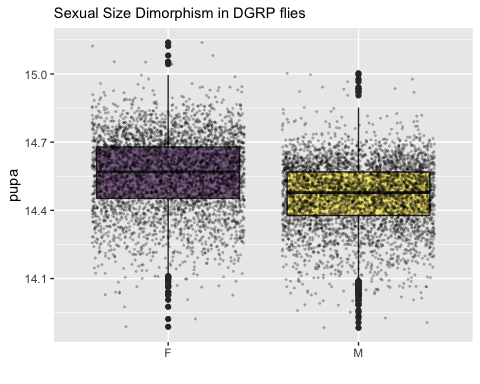
# Testing effect of sex in pupa size, with random effect for line and block.  
SSD<-lmer(pupa~sex+(1|line) +(1|block), REML=TRUE, data=df0)  
summary(SSD)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex + (1 | line) + (1 | block)  
## Data: df0  
##   
## REML criterion at convergence: -11278.3  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.4604 -0.6102 0.0874 0.6699 3.1485   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.009976 0.09988   
## block (Intercept) 0.002266 0.04760   
## Residual 0.013302 0.11534   
## Number of obs: 8065, groups: line, 187; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.456e+01 1.777e-02 9.923e+00 819.41 <2e-16 \*\*\*  
## sexM -9.436e-02 2.611e-03 7.891e+03 -36.14 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sexM -0.076

anova(SSD)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## sex 17.374 17.374 1 7890.6 1306.1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# plotting male and female size mean  
df0 %>%  
 ggplot( aes(x=sex, y=pupa, fill=sex)) +  
 geom\_boxplot() +  
 scale\_fill\_viridis(discrete = TRUE, alpha=0.6) +  
 geom\_jitter(color="black", size=0.4, alpha=0.2) +  
 #theme\_ipsum() +  
 theme(  
 legend.position="none",  
 plot.title = element\_text(size=11)  
 ) +  
 ggtitle("Sexual Size Dimorphism in DGRP flies") +  
 xlab("")



Answer 1.1: Yes, there is sexual size dimorphism in the DGRP flies, and females are larger than males on average.

## Question 1.2: Do we see a genetic variation of SSD in the DGRP lines?

### Comparing two models using ANOVA

#comparing two models  
model2<-lmer(pupa~sex+(1|line)+(1|block), data=df0) #model to test for SSD presence as we did above  
model1<-lmer(pupa~sex+(sex|line)+(1|block), data=df0)   
anova(model1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## sex 9.6134 9.6134 1 151.31 734.22 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(model1,model2)

## Data: df0  
## Models:  
## model2: pupa ~ sex + (1 | line) + (1 | block)  
## model1: pupa ~ sex + (sex | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 5 -11285 -11250 5647.3 -11295   
## model1 7 -11316 -11267 5664.9 -11330 35.264 2 2.2e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Model 1 is better, as AIC and BIC is smaller and log likelihood is higher. The difference of fit between these two models is significant.

### Do a LRT

How many parameters for each models

(AIC(model1) - REMLcrit(model1))/2 # # of parameters the model "thinks" are being estimated

## [1] 7

(AIC(model2) - REMLcrit(model2))/2 # # of parameters the model "thinks" are being estimated

## [1] 5

So lme4/lmer is treating model 1 as having two more parameters than model2.

LR.model <- -as.numeric(REMLcrit(model1) - REMLcrit(model2))  
LR.model

## [1] 35.79724

nlevels(df$line)

## [1] 0

pchisq(q = LR.model, df=2, lower=F)

## [1] 1.685498e-08

pchisq(q = LR.model, df=nlevels(df$line), lower=F)

## [1] 0

### Parametric boostrap (not ran)

Finally, we can conduct a parametric bootstrap to compare the two models.

### Finally using Bayesian Analysis

prior.2 <-list(R=list(V=0.01, nu=0.002),   
 G=list(G1=list(V=0.01\*diag(1), nu=0.002),  
 G2=list(V=0.01\*diag(2), nu=0.002)))  
  
model1M.MCMC <- MCMCglmm(pupa ~ 1 + sex,   
 random=~block + us(1 + sex):line,  
 prior = prior.2, burnin = 5000, nitt = 20000, thin = 10,  
 verbose = F, pr = T,  
 data=df0)  
summary(model1M.MCMC)

##   
## Iterations = 5001:19991  
## Thinning interval = 10  
## Sample size = 1500   
##   
## DIC: -11819.49   
##   
## G-structure: ~block  
##   
## post.mean l-95% CI u-95% CI eff.samp  
## block 0.002887 0.0005818 0.00683 1500  
##   
## ~us(1 + sex):line  
##   
## post.mean l-95% CI u-95% CI eff.samp  
## (Intercept):(Intercept).line 0.0112853 0.0088169 0.0138182 1500.0  
## sexM:(Intercept).line -0.0013976 -0.0022002 -0.0006144 1500.0  
## (Intercept):sexM.line -0.0013976 -0.0022002 -0.0006144 1500.0  
## sexM:sexM.line 0.0008737 0.0004197 0.0013529 667.8  
##   
## R-structure: ~units  
##   
## post.mean l-95% CI u-95% CI eff.samp  
## units 0.01311 0.01272 0.01355 1531  
##   
## Location effects: pupa ~ 1 + sex   
##   
## post.mean l-95% CI u-95% CI eff.samp pMCMC   
## (Intercept) 14.56386 14.52174 14.60119 1500 <7e-04 \*\*\*  
## sexM -0.09427 -0.10109 -0.08748 1500 <7e-04 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

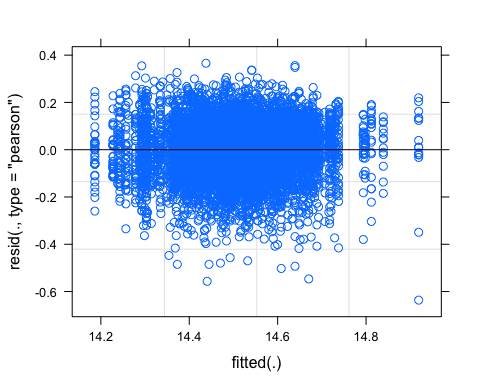
### Post model 1 fitting check

#### Residual distribution

res\_model1=residuals(model1)

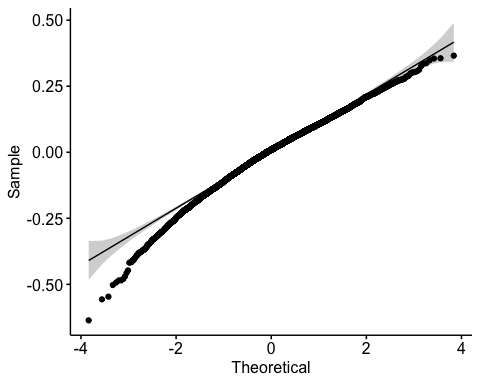
#### Model 1 residual distribution

plot(model1)



#### QQ plot

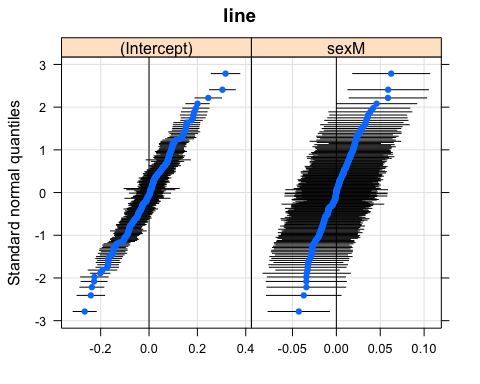
require(ggpubr)  
ggqqplot(res\_model1)



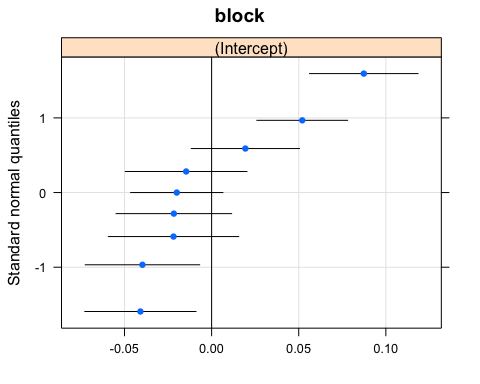
#### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block

 We have two plots, one for line and one for block

## Question 1.3: Is SSD genetic variation due to the genetic variation in male or in female size?

To test this, I need to look at the correlation between SSD and male size and SSD and female size, the correlation that is higher means that that sex contributes the most to the SSD variation.

I need to calculate summary for pupa female size, pupa male size and SSD.

#calculate means for each group using pupa\_noblock  
head(df0)

## # A tibble: 6 x 10  
## # Groups: group [6]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 853-A-0-M-027 L853 B7 D0 M 13.9 13.9 13.9 L853\_M\_… -0.485  
## 2 317-B-0-F-6693 L317 B11 D0 F 13.8 13.6 13.9 L317\_F\_… -0.536  
## 3 787-A-0-M-011 L787 B7 D0 M 13.9 13.6 13.9 L787\_M\_… -0.462  
## 4 513-B-0-F-59 L513 B13 D0 F 13.4 13.1 13.9 L513\_F\_… -0.532  
## 5 513-B-0-M-61 L513 B13 D0 M 13.4 13.2 13.9 L513\_M\_… -0.527  
## 6 492-A-0-M-014 L492 B6 D0 M 13.4 13.2 13.9 L492\_M\_… -0.437

df0\_mean<-aggregate(df0[, 10], list(df0$group), mean)  
head(df0\_mean)

## Group.1 pupa\_noblock  
## 1 L100\_F\_D0 0.256259113  
## 2 L100\_M\_D0 0.117684442  
## 3 L101\_F\_D0 0.109003255  
## 4 L101\_M\_D0 0.089615441  
## 5 L105\_F\_D0 -0.003621951  
## 6 L105\_M\_D0 -0.053412213

#re-add line, day and sex columns  
df0\_mean<-df0\_mean %>%  
 separate(Group.1, c("line", "sex","day"), "\_")  
head(df0\_mean)

## line sex day pupa\_noblock  
## 1 L100 F D0 0.256259113  
## 2 L100 M D0 0.117684442  
## 3 L101 F D0 0.109003255  
## 4 L101 M D0 0.089615441  
## 5 L105 F D0 -0.003621951  
## 6 L105 M D0 -0.053412213

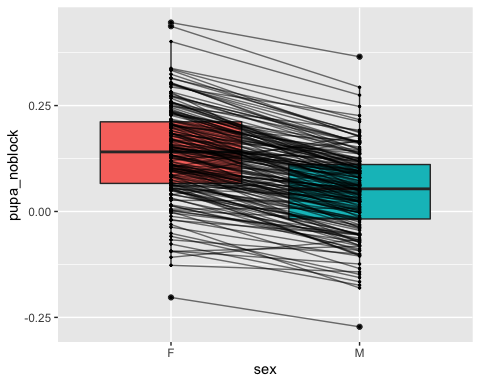
Size Mean by line plot boxplot for fed flies

# this plot is based on the variation excluding block effect  
head(df0\_mean)

## line sex day pupa\_noblock  
## 1 L100 F D0 0.256259113  
## 2 L100 M D0 0.117684442  
## 3 L101 F D0 0.109003255  
## 4 L101 M D0 0.089615441  
## 5 L105 F D0 -0.003621951  
## 6 L105 M D0 -0.053412213

df0\_mean %>%  
 ggplot(aes(sex,pupa\_noblock, fill=sex)) +  
 geom\_boxplot() +  
 geom\_point(size=0.5)+   
 geom\_line(aes(group=line, alpha=0.1, linewidth=0.5)) +  
 theme(legend.position = "none")

## Warning: Ignoring unknown aesthetics: linewidth



#calculate SSD0  
#separating males and females to put the values in columns  
df0\_mean\_F<-subset(df0\_mean, sex=="F")  
df0\_mean\_M<-subset(df0\_mean, sex=="M")  
  
df0\_mean\_2<-merge(x=df0\_mean\_F, y=df0\_mean\_M, by.x="line", by.y="line")  
head(df0\_mean\_2)

## line sex.x day.x pupa\_noblock.x sex.y day.y pupa\_noblock.y  
## 1 L100 F D0 0.256259113 M D0 0.11768444  
## 2 L101 F D0 0.109003255 M D0 0.08961544  
## 3 L105 F D0 -0.003621951 M D0 -0.05341221  
## 4 L109 F D0 0.063180232 M D0 -0.02019754  
## 5 L136 F D0 0.184305984 M D0 0.06867247  
## 6 L138 F D0 0.207235102 M D0 0.12057467

#remove extra columns  
df0\_mean\_2<-df0\_mean\_2[,c(1,4,7)] #line, pupa\_noblockF and pupa\_noblockM  
colnames(df0\_mean\_2) <- c("line", "pupaF", "pupaM") #rename col  
  
df0\_mean<-df0\_mean\_2 #move back to df0\_mean  
df0\_mean$SSD0<- df0\_mean$pupaF - df0\_mean$pupaM #since we established that females are larger than males in general, SSD is female-male sizes  
head(df0\_mean) #182 lines

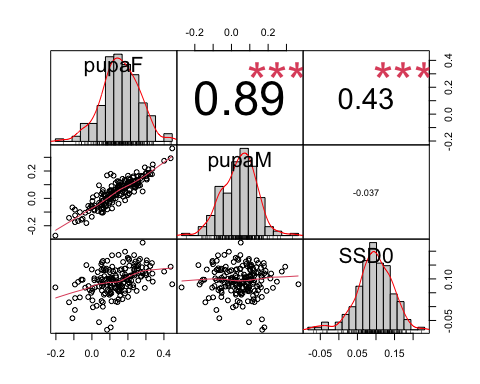
## line pupaF pupaM SSD0  
## 1 L100 0.256259113 0.11768444 0.13857467  
## 2 L101 0.109003255 0.08961544 0.01938781  
## 3 L105 -0.003621951 -0.05341221 0.04979026  
## 4 L109 0.063180232 -0.02019754 0.08337778  
## 5 L136 0.184305984 0.06867247 0.11563351  
## 6 L138 0.207235102 0.12057467 0.08666043

Correlation test

corr.test(df0\_mean[2:4],  
 use = "pairwise",  
 method = "pearson",  
 adjust = "none")

## Call:corr.test(x = df0\_mean[2:4], use = "pairwise", method = "pearson",   
## adjust = "none")  
## Correlation matrix   
## pupaF pupaM SSD0  
## pupaF 1.00 0.89 0.43  
## pupaM 0.89 1.00 -0.04  
## SSD0 0.43 -0.04 1.00  
## Sample Size   
## [1] 182  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## pupaF pupaM SSD0  
## pupaF 0 0.00 0.00  
## pupaM 0 0.00 0.62  
## SSD0 0 0.62 0.00  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

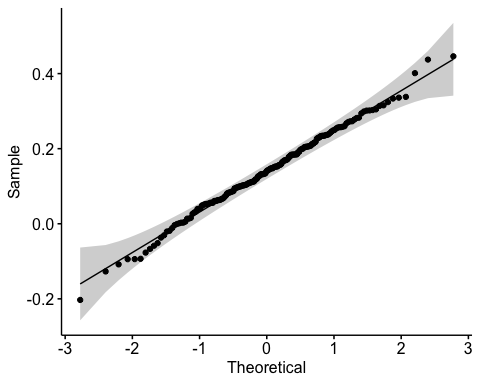
chart.Correlation(df0\_mean[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

 It seems that SSD in fed flies co-varies with the most with female size. The R2 is not very high though, so it is not all that explains SSD0 variation.

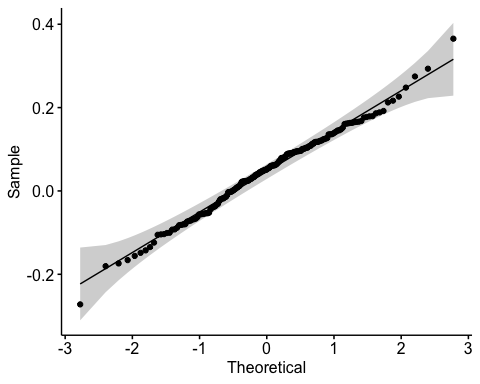
Does female size vary more genetically than male size?

One tail variance F test

#Before performing the F test, I need to check that both samples are normally distributed.  
require(ggpubr)  
ggqqplot(df0\_mean$pupaF)



ggqqplot(df0\_mean$pupaM)



shapiro.test(df0\_mean$pupaM) #normal distribution

##   
## Shapiro-Wilk normality test  
##   
## data: df0\_mean$pupaM  
## W = 0.99272, p-value = 0.4996

shapiro.test(df0\_mean$pupaF) #normal distribution

##   
## Shapiro-Wilk normality test  
##   
## data: df0\_mean$pupaF  
## W = 0.99563, p-value = 0.8788

#one tail F test  
  
var.test(df0\_mean$pupaF, df0\_mean$pupaM, alternative = "greater")

##   
## F test to compare two variances  
##   
## data: df0\_mean$pupaF and df0\_mean$pupaM  
## F = 1.2274, num df = 181, denom df = 181, p-value = 0.08453  
## alternative hypothesis: true ratio of variances is greater than 1  
## 95 percent confidence interval:  
## 0.9605383 Inf  
## sample estimates:  
## ratio of variances   
## 1.227407

The variance in female size is actually not larger than the variance in male size, despite the higher correlation between female size and SSD0, than male size and SS0. So does it mean that SSD0 varies partly because of a co-variation with female size but not from the fact that female size varies more genetically? So where does that SSD variation come from?

# Section 2: SSD in other conditions

We found that in fed flies, females are on average larger than males, so there is a female biased SSD. That SSD also vary genetically. We finally found out that SSD variation is in part driven by the genetic variation of female size, although the variance of female size is the same as male size.

Let’s look at different starving conditions

* Question 2.1: Do we have the same SSD when we change environment? NO
* Question 2.2: Does SSD increase or decrease when the flies are starved? Our hypothesis is that overall SSD should decrease
* Question 2.3: Does the SSD in starved conditions vary the same way as in fed flies?

## Question 2.1: Do we have the same SSD when we change environment?

### 1 day starvation data

#subset day 1 and day 2 starvation  
  
#fisrt subset Day1 data from df\_sub  
df1<-subset(df\_sub, day=="D1") #df\_sub N<10 per group filtered out already  
  
#na.omit only if pupa has NA  
df1<-na.omit(df1, cols="pupa") #6409 rows  
length(unique(df1$line)) #179 lines left

## [1] 179

# Testing effect of sex in pupa size, with random effect for line and block.  
SSD1<-lmer(pupa~sex+(1|line) +(1|block), REML=TRUE, data=df1)  
summary(SSD1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex + (1 | line) + (1 | block)  
## Data: df1  
##   
## REML criterion at convergence: -7178.2  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.7420 -0.6448 0.0154 0.6609 3.3926   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.010985 0.10481   
## block (Intercept) 0.001843 0.04293   
## Residual 0.017470 0.13217   
## Number of obs: 6409, groups: line, 179; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.440e+01 1.676e-02 8.427e+00 858.94 <2e-16 \*\*\*  
## sexM -6.611e-02 3.437e-03 6.292e+03 -19.23 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sexM -0.109

anova(SSD1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## sex 6.463 6.463 1 6291.6 369.95 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

SSD still exists at 1 day starvation

### 2 day starvation data

#subsetting day 2 starvation   
df2<-subset(df\_sub, day=="D2")   
#na.omit only if pupa has NA  
df2<-na.omit(df2, cols="pupa") #2673 rows  
length(unique(df2$line)) #88 lines

## [1] 88

# Testing effect of sex in pupa size, with random effect for line and block.  
SSD2<-lmer(pupa~sex+(1|line) +(1|block), REML=TRUE, data=df2)  
summary(SSD2)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex + (1 | line) + (1 | block)  
## Data: df2  
##   
## REML criterion at convergence: -3601.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.3563 -0.6596 0.0276 0.6644 3.5750   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.010789 0.10387   
## block (Intercept) 0.007009 0.08372   
## Residual 0.013655 0.11685   
## Number of obs: 2673, groups: line, 88; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.431e+01 3.078e-02 9.106e+00 464.85 <2e-16 \*\*\*  
## sexM -7.920e-02 4.812e-03 2.636e+03 -16.46 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sexM -0.078

Anova(SSD2)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: pupa  
## Chisq Df Pr(>Chisq)   
## sex 270.9 1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

SSD is present at 2 day starvation

How do I see if the SSDs are the same? - I could check if SSD average is the same - Then check their correlation

First I need to calculate SSD1 and SSD2

#SSD1  
# calculate size mean for male and female  
head(df1)

## # A tibble: 6 x 10  
## # Groups: group [5]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 317-B-1-F-7742 L317 B11 D1 F 13.7 13.4 13.8 L317\_F\_… -0.607  
## 2 911-B-1-M-096 L911 B7 D1 M 13.3 13.0 13.8 L911\_M\_… -0.537  
## 3 796-A-1-M-034 L796 B7 D1 M 13.6 13.6 13.8 L796\_M\_… -0.526  
## 4 105-A-1-M-7585 L105 B11 D1 M 13.5 13.4 13.8 L105\_M\_… -0.581  
## 5 105-B-1-M-7570 L105 B11 D1 M 13.5 13.2 13.8 L105\_M\_… -0.579  
## 6 861-B-1-M-101 L861 B7 D1 M 13.5 13.2 13.8 L861\_M\_… -0.524

df1\_mean<-aggregate(df1[, 10], list(df1$group), mean) #calculate mean of pupa\_noblock to remove the block effect  
head(df1\_mean)

## Group.1 pupa\_noblock  
## 1 L100\_F\_D1 0.08515325  
## 2 L100\_M\_D1 0.06014172  
## 3 L101\_F\_D1 -0.03520331  
## 4 L101\_M\_D1 -0.10079021  
## 5 L105\_F\_D1 -0.08479509  
## 6 L105\_M\_D1 -0.25153763

#re-add line, day and sex columns  
df1\_mean<-df1\_mean %>%  
 separate(Group.1, c("line", "sex","day"), "\_")  
head(df1\_mean)

## line sex day pupa\_noblock  
## 1 L100 F D1 0.08515325  
## 2 L100 M D1 0.06014172  
## 3 L101 F D1 -0.03520331  
## 4 L101 M D1 -0.10079021  
## 5 L105 F D1 -0.08479509  
## 6 L105 M D1 -0.25153763

length(unique(df1\_mean$line)) #179 lines

## [1] 179

#separating males and females to put the values in columns  
df1\_mean\_F<-subset(df1\_mean, sex=="F")  
df1\_mean\_M<-subset(df1\_mean, sex=="M")  
  
df1\_mean\_2<-merge(x=df1\_mean\_F, y=df1\_mean\_M, by.x="line", by.y="line")  
head(df1\_mean\_2)

## line sex.x day.x pupa\_noblock.x sex.y day.y pupa\_noblock.y  
## 1 L100 F D1 0.08515325 M D1 0.06014172  
## 2 L101 F D1 -0.03520331 M D1 -0.10079021  
## 3 L105 F D1 -0.08479509 M D1 -0.25153763  
## 4 L109 F D1 -0.15244216 M D1 -0.25892157  
## 5 L129 F D1 0.04989262 M D1 -0.03625669  
## 6 L136 F D1 -0.09580059 M D1 -0.14308417

#remove extra columns  
df1\_mean\_2<-df1\_mean\_2[,c(1,4,7)] #using pupa\_noblock  
colnames(df1\_mean\_2) <- c("line", "pupaF1", "pupaM1")  
  
df1\_mean<-df1\_mean\_2  
df1\_mean$SSD1<- df1\_mean$pupaF1 - df1\_mean$pupaM1 #since we established that females are larger than males in general  
head(df1\_mean)

## line pupaF1 pupaM1 SSD1  
## 1 L100 0.08515325 0.06014172 0.02501154  
## 2 L101 -0.03520331 -0.10079021 0.06558689  
## 3 L105 -0.08479509 -0.25153763 0.16674254  
## 4 L109 -0.15244216 -0.25892157 0.10647941  
## 5 L129 0.04989262 -0.03625669 0.08614931  
## 6 L136 -0.09580059 -0.14308417 0.04728358

length(unique(df1\_mean$line)) #157 lines left after calculating SSD1

## [1] 157

#SSD 2  
# calculate size mean for male and female  
#head(df2)  
df2\_mean<-aggregate(df2[, 10], list(df2$group), mean)  
head(df2\_mean)

## Group.1 pupa\_noblock  
## 1 L100\_M\_D2 -0.23292373  
## 2 L153\_F\_D2 0.20834910  
## 3 L153\_M\_D2 0.08699819  
## 4 L158\_F\_D2 0.01987062  
## 5 L158\_M\_D2 -0.07201030  
## 6 L177\_M\_D2 -0.35313746

#re-add line, day and sex columns  
df2\_mean<-df2\_mean %>%  
 separate(Group.1, c("line", "sex","day"), "\_")  
head(df2\_mean)

## line sex day pupa\_noblock  
## 1 L100 M D2 -0.23292373  
## 2 L153 F D2 0.20834910  
## 3 L153 M D2 0.08699819  
## 4 L158 F D2 0.01987062  
## 5 L158 M D2 -0.07201030  
## 6 L177 M D2 -0.35313746

#separating males and females to put the values in columns  
df2\_mean\_F<-subset(df2\_mean, sex=="F")  
df2\_mean\_M<-subset(df2\_mean, sex=="M")  
  
df2\_mean\_2<-merge(x=df2\_mean\_F, y=df2\_mean\_M, by.x="line", by.y="line")  
head(df2\_mean\_2)

## line sex.x day.x pupa\_noblock.x sex.y day.y pupa\_noblock.y  
## 1 L153 F D2 0.208349100 M D2 0.086998187  
## 2 L158 F D2 0.019870621 M D2 -0.072010303  
## 3 L229 F D2 0.025244072 M D2 -0.171367258  
## 4 L256 F D2 -0.051082781 M D2 -0.174299175  
## 5 L26 F D2 -0.002092585 M D2 -0.104164760  
## 6 L28 F D2 0.092029050 M D2 -0.003598201

#remove extra columns  
df2\_mean\_2<-df2\_mean\_2[,c(1,4,7)]   
colnames(df2\_mean\_2) <- c("line", "pupaF2", "pupaM2")  
  
df2\_mean<-df2\_mean\_2  
df2\_mean$SSD2<- df2\_mean$pupaF2 - df2\_mean$pupaM2 #since we established that females are larger than males in general  
head(df2\_mean)

## line pupaF2 pupaM2 SSD2  
## 1 L153 0.208349100 0.086998187 0.12135091  
## 2 L158 0.019870621 -0.072010303 0.09188092  
## 3 L229 0.025244072 -0.171367258 0.19661133  
## 4 L256 -0.051082781 -0.174299175 0.12321639  
## 5 L26 -0.002092585 -0.104164760 0.10207217  
## 6 L28 0.092029050 -0.003598201 0.09562725

length(unique(df2\_mean$line)) #59 lines left after I calculate SSD2

## [1] 59

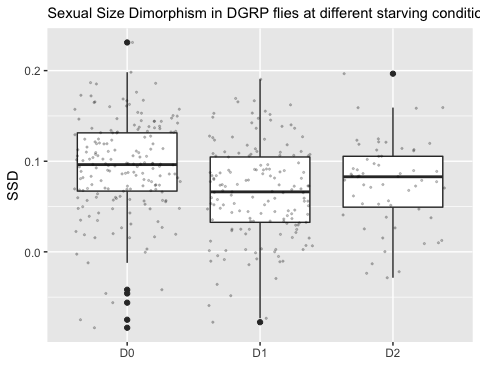
##Question 2.2: Does overall SSD increase or decrease when the flies are starved? Our hypothesis is that overall SSD should decrease

To make the boxplot, I initially merged SSDs dataframes but it discards all lines that do not have SSD2, so this time I will rbind the dataframes to keep a maximum number of rows

#combining SSDs values without discarding rows, which would happen if I merged the dataframes (we would end up with 59 lines for all SSD2)  
SSD0<-df0\_mean[,c(1,4)]  
SSD0$day<-"D0"  
names(SSD0)[names(SSD0) == "SSD0"] <- "SSD"  
  
SSD1<-df1\_mean[,c(1,4)]  
SSD1$day<-"D1"  
names(SSD1)[names(SSD1) == "SSD1"] <- "SSD"  
  
SSD2<-df2\_mean[,c(1,4)]  
SSD2$day<-"D2"  
names(SSD2)[names(SSD2) == "SSD2"] <- "SSD"  
  
SSD\_all<-rbind(SSD0,SSD1,SSD2)

SSD boxplot, the overall SSD decreases when flies are starved

# plotting male and female size mean  
SSD\_all %>%  
 ggplot( aes(x=day, y=SSD)) +  
 geom\_boxplot() +  
 scale\_fill\_viridis(discrete = TRUE, alpha=0.6) +  
 geom\_jitter(color="black", size=0.4, alpha=0.2) +  
 #theme\_ipsum() +  
 theme(  
 legend.position="none",  
 plot.title = element\_text(size=11)  
 ) +  
 ggtitle("Sexual Size Dimorphism in DGRP flies at different starving conditions") +  
 xlab("")

 SSD decreases at Day1 starvation but seems to reincrease at day2, let’s check if this is significant

I just want to compare the means of SSD.

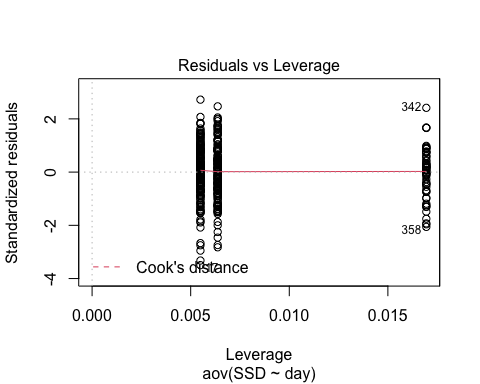
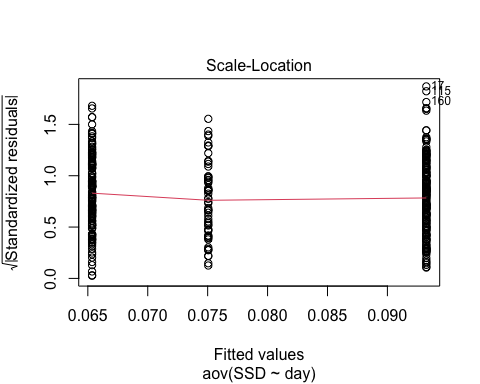
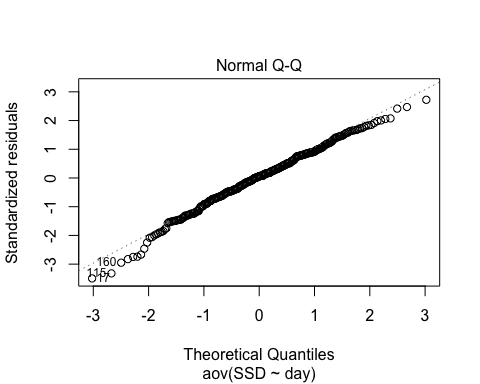
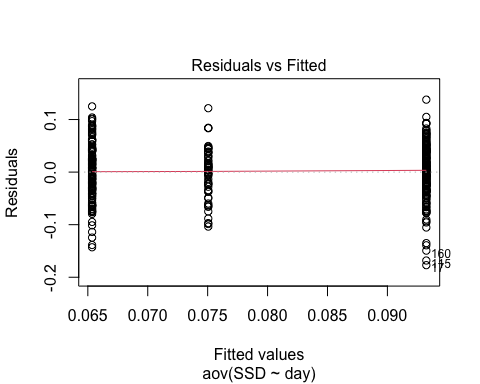
#ANOVA on SSD per day  
head(SSD\_all)

## line SSD day  
## 1 L100 0.13857467 D0  
## 2 L101 0.01938781 D0  
## 3 L105 0.04979026 D0  
## 4 L109 0.08337778 D0  
## 5 L136 0.11563351 D0  
## 6 L138 0.08666043 D0

SSDtest<-aov(data=SSD\_all, SSD ~day)  
summary(SSDtest)

## Df Sum Sq Mean Sq F value Pr(>F)   
## day 2 0.067 0.03349 13 3.42e-06 \*\*\*  
## Residuals 395 1.018 0.00258   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(SSDtest)



model.tables(SSDtest, type="means", se = TRUE)

## Design is unbalanced - use se.contrast() for se's

## Tables of means  
## Grand mean  
##   
## 0.07954706   
##   
## day   
## D0 D1 D2  
## 0.09325 0.06536 0.07505  
## rep 182.00000 157.00000 59.00000

#Pairwise comparison  
TukeyHSD(SSDtest, which = "day")

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = SSD ~ day, data = SSD\_all)  
##   
## $day  
## diff lwr upr p adj  
## D1-D0 -0.027891722 -0.040899261 -0.0148841833 0.0000021  
## D2-D0 -0.018196675 -0.036087387 -0.0003059631 0.0451745  
## D2-D1 0.009695047 -0.008541048 0.0279311421 0.4240973

SSD0 and SSD1 are significantly different, with SSD0 > SSD1 SSD0 and SSD2 are significantly different (p=0.0451745), SSD0 > SSD2 SSD1 and SSD2 are not significantly different

Starvation decreases SSD but it does not change at different levels of starvation. high p value of SSD0 vs SSD2 due to low number of lines? Should I fit lmer with day0 and day2 raw data subset instead?

## Question 2.3: Does SSD in starved conditions vary, and does it covary with SSD0?

First, we want to see if there is genetic variation in SSD1 and SSD2 ### SSD1

#Comparing two model fit for SSD1 to see if there is genetic variation  
  
model2<-lmer(pupa~sex+(1|line)+(1|block), data=df1) #model to test for SSD presence as we did above  
model1<-lmer(pupa~sex+(sex|line)+(1|block), data=df1)   
anova(model1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## sex 4.1858 4.1858 1 166.44 242.44 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(model1,model2)

## refitting model(s) with ML (instead of REML)

## Data: df1  
## Models:  
## model2: pupa ~ sex + (1 | line) + (1 | block)  
## model1: pupa ~ sex + (sex | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 5 -7184.1 -7150.2 3597.0 -7194.1   
## model1 7 -7198.5 -7151.2 3606.3 -7212.5 18.442 2 9.893e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Model 1 is better, as AIC and BIC is smaller and log likelihood is higher. The difference of fit between these two models is significant.

### Do a LRT

How many parameters for each models

(AIC(model1) - REMLcrit(model1))/2 # # of parameters the model "thinks" are being estimated

## [1] 7

(AIC(model2) - REMLcrit(model2))/2 # # of parameters the model "thinks" are being estimated

## [1] 5

So lme4/lmer is treating model 1 as having two more parameters than model2.

LR.model <- -as.numeric(REMLcrit(model1) - REMLcrit(model2))  
LR.model

## [1] 18.79382

nlevels(df1$line)

## [1] 0

pchisq(q = LR.model, df=2, lower=F)

## [1] 8.297996e-05

pchisq(q = LR.model, df=nlevels(df1$line), lower=F)

## [1] 0

### Parametric boostrap

Finally, we can conduct a parametric bootstrap to compare the two models.

### Finally using Bayesian Analysis

#prior.2 <-list(R=list(V=0.01, nu=0.002),   
# G=list(G1=list(V=0.01\*diag(1), nu=0.002),  
 # G2=list(V=0.01\*diag(2), nu=0.002)))  
  
#model1M.MCMC <- MCMCglmm(pupa ~ 1 + sex,   
# random=~block + us(1 + sex):line,  
 # prior = prior.2, burnin = 5000, nitt = 20000, thin = 10,  
# verbose = F, pr = T,  
 # data=df1)  
#summary(model1M.MCMC)

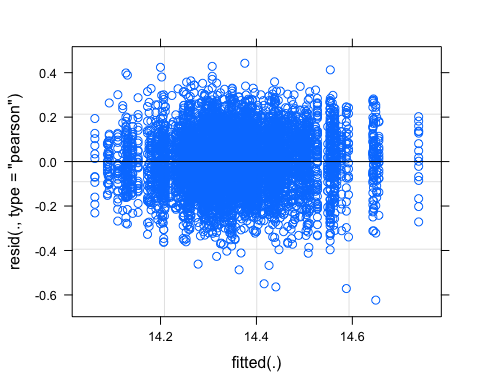
### Post model 1 fitting check

#### Residual distribution

res\_model1=residuals(model1)

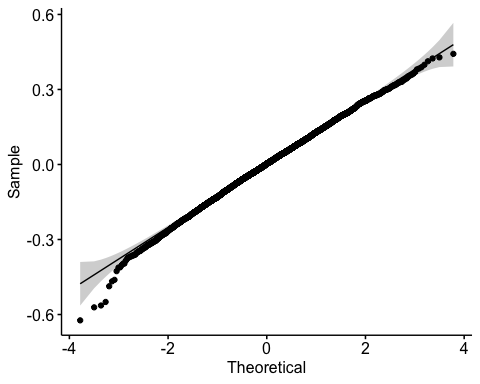
#### Model 1 residual distribution

plot(model1)



#### QQ plot

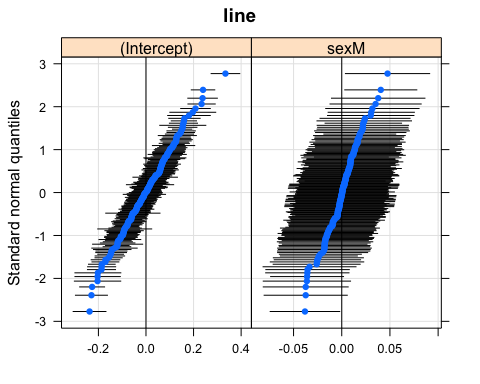
require(ggpubr)  
ggqqplot(res\_model1)



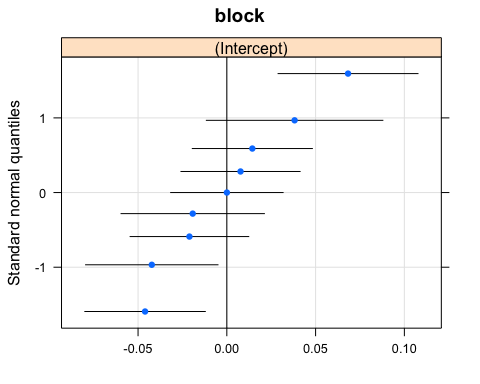
#### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block



### SSD2

#Comparing two model fit for SSD2 to see if there is genetic variation  
  
model2<-lmer(pupa~sex+(1|line)+(1|block), data=df2) #model to test for SSD presence as we did above  
model1<-lmer(pupa~sex+(sex|line)+(1|block), data=df2)

## boundary (singular) fit: see ?isSingular

anova(model1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## sex 2.998 2.998 1 227.51 220.25 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(model1,model2)

## refitting model(s) with ML (instead of REML)

## Data: df2  
## Models:  
## model2: pupa ~ sex + (1 | line) + (1 | block)  
## model1: pupa ~ sex + (sex | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 5 -3605.6 -3576.1 1807.8 -3615.6   
## model1 7 -3608.9 -3567.7 1811.5 -3622.9 7.3459 2 0.0254 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Model 1 is better, as AIC and BIC is smaller and log likelihood is higher. The difference of fit between these two models is significant.

### Do a LRT

How many parameters for each models

(AIC(model1) - REMLcrit(model1))/2 # # of parameters the model "thinks" are being estimated

## [1] 7

(AIC(model2) - REMLcrit(model2))/2 # # of parameters the model "thinks" are being estimated

## [1] 5

So lme4/lmer is treating model 1 as having two more parameters than model2.

LR.model <- -as.numeric(REMLcrit(model1) - REMLcrit(model2))  
LR.model

## [1] 7.454988

nlevels(df2$line)

## [1] 0

pchisq(q = LR.model, df=2, lower=F)

## [1] 0.02405304

pchisq(q = LR.model, df=nlevels(df2$line), lower=F)

## [1] 0

### Parametric boostrap

Finally, we can conduct a parametric bootstrap to compare the two models.

### Finally using Bayesian Analysis

#prior.2 <-list(R=list(V=0.01, nu=0.002),   
 # G=list(G1=list(V=0.01\*diag(1), nu=0.002),  
 # G2=list(V=0.01\*diag(2), nu=0.002)))  
  
#model1M.MCMC <- MCMCglmm(pupa ~ 1 + sex,   
 # random=~block + us(1 + sex):line,  
 #prior = prior.2, burnin = 5000, nitt = 20000, thin = 10,  
# verbose = F, pr = T,  
 # data=df2)  
#summary(model1M.MCMC)

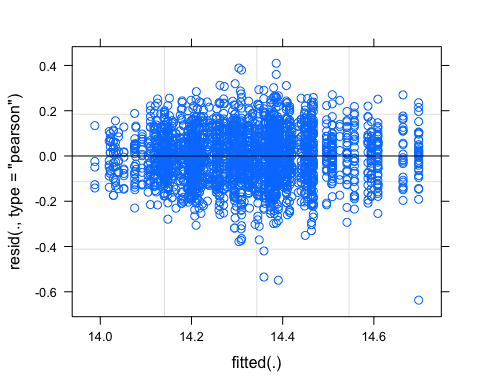
### Post model 1 fitting check

#### Residual distribution

res\_model1=residuals(model1)

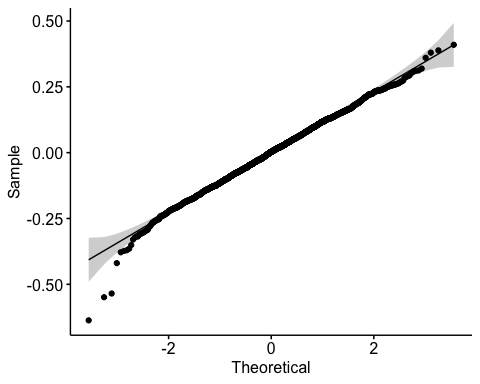
#### Model 1 residual distribution

plot(model1)



#### QQ plot

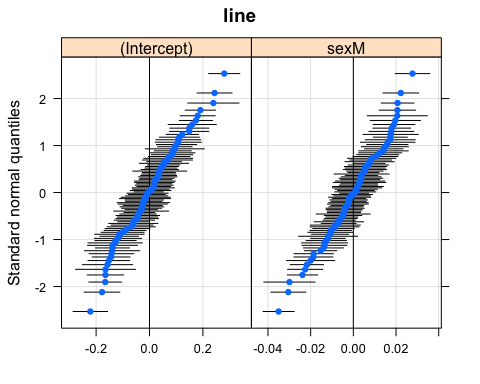
require(ggpubr)  
ggqqplot(res\_model1)



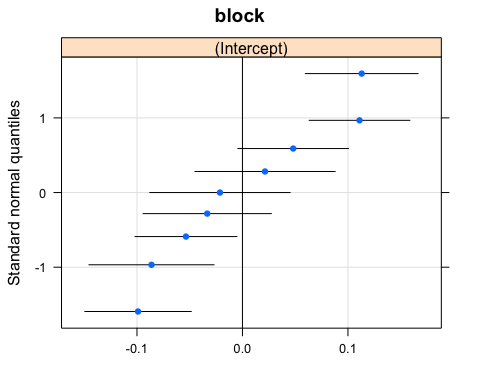
#### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block

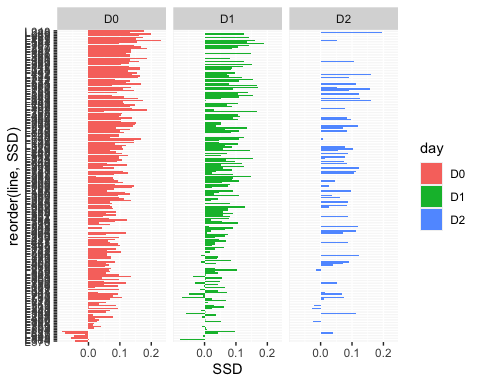
 We have two plots, one for line and one for block

SSD1 and SSD2 vary also significantly.

Does that variation correlate with SSD0?

If SSD0 covaries with SSD1/SSD2, that means we should not see a difference in plasticity. We first can visualize how SSD changes across lineages.

ggplot(SSD\_all, aes(x=reorder(line,SSD), y=SSD)) +  
 geom\_col(aes(fill = day)) +  
 facet\_wrap(~ day) +  
 coord\_flip()

 Correlation between SSDs. Hypothesis, if SSD changes with environment, which is what we expect, we will not see a correlation between SSD0 and SSD1 and/or SSD2

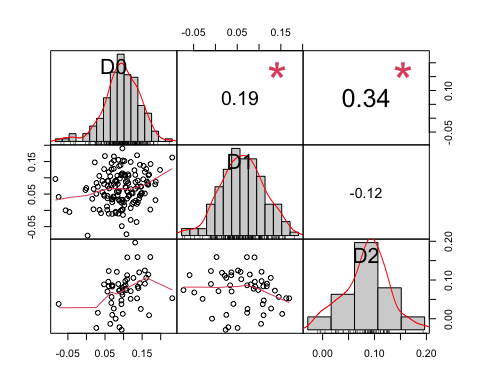
#correlation SSD0, SSD1, SSD2  
SSD\_all\_col<-spread(SSD\_all,day ,SSD)  
  
head(SSD\_all\_col) #D0 is SSD0, D1 is SSD1 and D2 is SSD2

## line D0 D1 D2  
## 1 L100 0.13857467 0.02501154 NA  
## 2 L101 0.01938781 0.06558689 NA  
## 3 L105 0.04979026 0.16674254 NA  
## 4 L109 0.08337778 0.10647941 NA  
## 5 L129 NA 0.08614931 NA  
## 6 L136 0.11563351 0.04728358 NA

corr.test(SSD\_all\_col[2:4],  
 use = "pairwise",  
 method = "pearson",  
 adjust = "none")

## Call:corr.test(x = SSD\_all\_col[2:4], use = "pairwise", method = "pearson",   
## adjust = "none")  
## Correlation matrix   
## D0 D1 D2  
## D0 1.00 0.19 0.34  
## D1 0.19 1.00 -0.12  
## D2 0.34 -0.12 1.00  
## Sample Size   
## D0 D1 D2  
## D0 182 146 55  
## D1 146 157 51  
## D2 55 51 59  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## D0 D1 D2  
## D0 0.00 0.02 0.01  
## D1 0.02 0.00 0.40  
## D2 0.01 0.40 0.00  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

chart.Correlation(SSD\_all\_col[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

 There is a weak correlation between SSD0 and SSD2, still higher than SSD1 and SSD0

Since there is just a weak correlation between SSD0 and SSD2, I want to know if SSD1 and SSD2 varies because of female size variation as it was found in SSD0.

Correlation SSD2 and female and male size

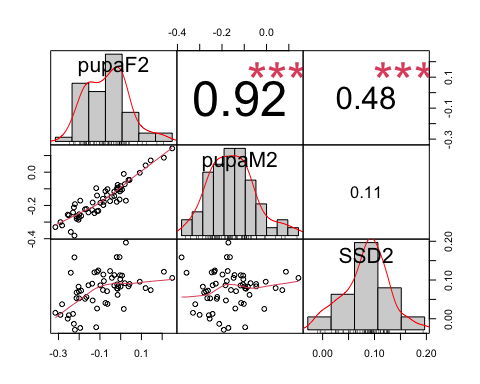
#SSD2 and female and male size dataset  
head(df2\_mean)

## line pupaF2 pupaM2 SSD2  
## 1 L153 0.2083490996 0.086998187 0.1213509129  
## 2 L158 0.0198706211 -0.072010303 0.0918809238  
## 3 L229 0.0252440717 -0.171367258 0.1966113298  
## 4 L256 -0.0510827809 -0.174299175 0.1232163941  
## 5 L26 -0.0020925854 -0.104164760 0.1020721748

corr.test(df2\_mean[2:4],  
 use = "pairwise",  
 method = "pearson",  
 adjust = "none")

## Call:corr.test(x = df2\_mean[2:4], use = "pairwise", method = "pearson",   
## adjust = "none")  
## Correlation matrix   
## pupaF2 pupaM2 SSD2  
## pupaF2 1.00 0.92 0.48  
## pupaM2 0.92 1.00 0.11  
## SSD2 0.48 0.11 1.00  
## Sample Size   
## [1] 59  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## pupaF2 pupaM2 SSD2  
## pupaF2 0 0.0 0.0  
## pupaM2 0 0.0 0.4  
## SSD2 0 0.4 0.0  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

#correlation plot  
chart.Correlation(df2\_mean[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)



SSD2 covaries with female size in 2 days starvation.

Will still check SSD1, even though it does not correlate with SSD0, we can see if SSD1 variation covaries with either female or male variation

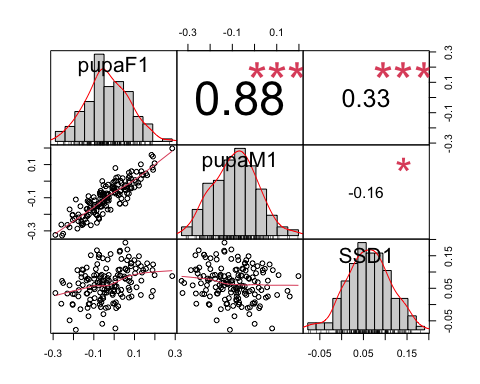
#SSD2 and female and male size dataset  
head(df1\_mean)

## line pupaF1 pupaM1 SSD1  
## 1 L100 0.0851532538 0.0601417184 2.501154e-02  
## 2 L101 -0.0352033147 -0.1007902097 6.558689e-02  
## 3 L105 -0.0847950896 -0.2515376332 1.667425e-01  
## 4 L109 -0.1524421631 -0.2589215723 1.064794e-01  
## 5 L129 0.0498926174 -0.0362566925 8.614931e-02

corr.test(df1\_mean[2:4],  
 use = "pairwise",  
 method = "pearson",  
 adjust = "none")

## Call:corr.test(x = df1\_mean[2:4], use = "pairwise", method = "pearson",   
## adjust = "none")  
## Correlation matrix   
## pupaF1 pupaM1 SSD1  
## pupaF1 1.00 0.88 0.33  
## pupaM1 0.88 1.00 -0.16  
## SSD1 0.33 -0.16 1.00  
## Sample Size   
## [1] 157  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## pupaF1 pupaM1 SSD1  
## pupaF1 0 0.00 0.00  
## pupaM1 0 0.00 0.05  
## SSD1 0 0.05 0.00  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

#correlation plot  
chart.Correlation(df1\_mean[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

 SSD1 has the highest correlation with female size variation

DOES IT MAKE SENSE TO COMPARE CORRELATIONS (is there a statistical test), I want to know if female Day0 variation controls more SSD0 than Day2 female size variation does for SSD2

We have shown that SSD in different condition changes, that in starved conditions, we have a decrease in overall SSD, and that the variation of SSD partly covariates with female size for SSD0 and SSD2. But the degree of correlation varies? (how to test that?)

But if SSD changes under different conditions, that means that females and males must have a different response to environmental changes, so we should see a sex specific plasticity or SSP.

#Section 3: SSP01 and SSP02 - Question 3.1: Is there SSP when we look at fed vs starved flies? - Question 3.2: Does this SSP vary in the DGRP flies? - Question 3.3: Is the variation in SSP due to a variation in SSD in fed or starved flies? If SSP covaries with SSD0, that means that we have variation in fed flies. If SSP covaries with starved flies, we have variation in the reduction of flies size. - Question 3.4: If it is due to SSD in either fed or starved flies, is it because the females vary more or the males vary more in size? - Question 3.5: Does the variation causing SSP (Question 3.4) is the same as the variation caused by SSD in Question 1.3?

## Question 3.1: Is there SSP when we look at fed vs starved flies?

### SSP01: between fed and 1 day starved flies

# use subset Day 1 and Day 0  
df01<-subset(df\_sub, day=="D0"|day=="D1")  
df01<-na.omit(df01, cols="pupa") #1447 rows  
  
SSP01\_test<-lmer(pupa~sex\*day+(1|line)+(1|block), REML=TRUE, data=df01) #random effect, there is variation in sex by line and there is variation in plasticity by line.  
summary(SSP01\_test)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex \* day + (1 | line) + (1 | block)  
## Data: df01  
##   
## REML criterion at convergence: -17613.2  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.9221 -0.6382 0.0577 0.6666 4.1394   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.009614 0.09805   
## block (Intercept) 0.001823 0.04269   
## Residual 0.016434 0.12819   
## Number of obs: 14474, groups: line, 195; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.457e+01 1.617e-02 9.561e+00 900.844 < 2e-16 \*\*\*  
## sexM -9.310e-02 2.883e-03 1.428e+04 -32.296 < 2e-16 \*\*\*  
## dayD1 -1.751e-01 3.238e-03 1.436e+04 -54.085 < 2e-16 \*\*\*  
## sexM:dayD1 2.610e-02 4.342e-03 1.429e+04 6.012 1.88e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexM dayD1   
## sexM -0.092   
## dayD1 -0.088 0.455   
## sexM:dayD1 0.061 -0.660 -0.698

Anova(SSP01\_test)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: pupa  
## Chisq Df Pr(>Chisq)   
## sex 1423.22 1 < 2.2e-16 \*\*\*  
## day 4858.28 1 < 2.2e-16 \*\*\*  
## sex:day 36.14 1 1.837e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### SSP02: between fed and 1 day starved flies

# use subset Day 1 and Day 0  
df02<-subset(df, day=="D0"|day=="D2")  
df02<-na.omit(df02, cols="pupa") #10738 rows  
SSP02\_test<-lmer(pupa~sex\*day+(1|line)+(1|block), REML=TRUE, data=df02) #random effect, there is variation in sex by line and there is variation in plasticity by line.  
summary(SSP02\_test)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex \* day + (1 | line) + (1 | block)  
## Data: df02  
##   
## REML criterion at convergence: -13989.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.1457 -0.6418 0.0686 0.6856 3.4675   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.009720 0.09859   
## block (Intercept) 0.003069 0.05540   
## Residual 0.014884 0.12200   
## Number of obs: 10738, groups: line, 189; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.457e+01 2.007e-02 9.093e+00 725.743 <2e-16 \*\*\*  
## sexM -9.395e-02 2.752e-03 1.056e+04 -34.138 <2e-16 \*\*\*  
## dayD2 -2.528e-01 4.330e-03 1.064e+04 -58.387 <2e-16 \*\*\*  
## sexM:dayD2 1.276e-02 5.533e-03 1.056e+04 2.306 0.0211 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexM dayD2   
## sexM -0.070   
## dayD2 -0.050 0.313   
## sexM:dayD2 0.035 -0.493 -0.647

Anova(SSP02\_test)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: pupa  
## Chisq Df Pr(>Chisq)   
## sex 1438.9486 1 < 2e-16 \*\*\*  
## day 5562.9346 1 < 2e-16 \*\*\*  
## sex:day 5.3168 1 0.02112 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##Question 3.2: Does this SSP vary in the DGRP flies? ### SSP01

# we want to compare effect of sex and day of starvation on pupal size. I am using df01, subset off data day 0 and day1  
  
model2<-lmer(pupa~sex\*day+(1|line)+(1|block), data=df01)  
model1<-lmer(pupa~sex\*day+(sex+day|line)+(1|block), data=df01)  
anova(model1,model2)

## Data: df01  
## Models:  
## model2: pupa ~ sex \* day + (1 | line) + (1 | block)  
## model1: pupa ~ sex \* day + (sex + day | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 7 -17635 -17582 8824.7 -17649   
## model1 12 -18490 -18399 9256.9 -18514 864.43 5 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Model 1 is better.

### Do a LRT

How many parameters for each models

(AIC(model1) - REMLcrit(model1))/2 # # of parameters the model "thinks" are being estimated

## [1] 12

(AIC(model2) - REMLcrit(model2))/2 # # of parameters the model "thinks" are being estimated

## [1] 7

So lme4/lmer is treating model 1 as having five more parameters than model2.

LR.model <- -as.numeric(REMLcrit(model1) - REMLcrit(model2))  
LR.model

## [1] 866.7934

nlevels(df01$line)

## [1] 0

pchisq(q = LR.model, df=5, lower=F)

## [1] 4.086886e-185

pchisq(q = LR.model, df=nlevels(df01$line), lower=F)

## [1] 0

### Parametric boostrap

Finally, we can conduct a parametric bootstrap to compare the two models.

### Finally using Bayesian Analysis

##NB: did not run before need to change the model

#prior.2 <-list(R=list(V=0.01, nu=0.002),   
 # G=list(G1=list(V=0.01\*diag(1), nu=0.002),  
 # G2=list(V=0.01\*diag(2), nu=0.002)))  
  
#model1M.MCMC <- MCMCglmm(pupa ~ 1 + sex, #is this right for SSP?  
 #random=~block + us(1 + sex):line,  
 #prior = prior.2, burnin = 5000, nitt = 20000, thin = 10,  
 #verbose = F, pr = T,  
 #data=df01)  
#summary(model1M.MCMC)

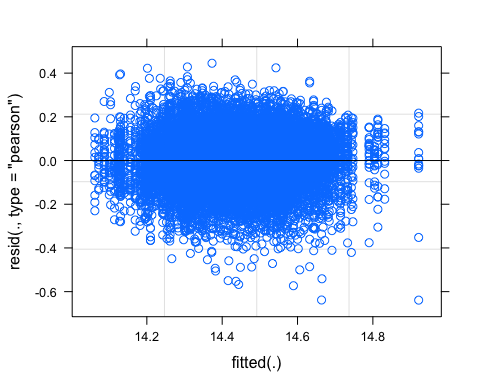
### Post model 1 fitting check

#### Residual distribution

res\_model1=residuals(model1)

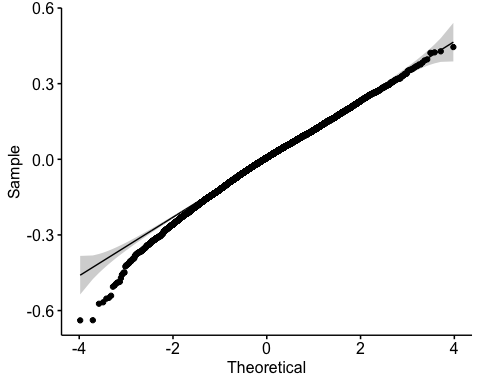
#### Model 1 residual distribution

plot(model1)



#### QQ plot

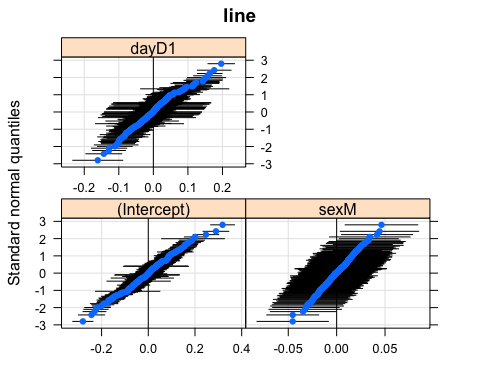
require(ggpubr)  
ggqqplot(res\_model1)



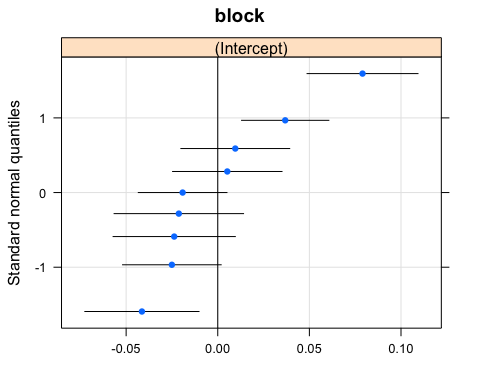
#### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block

 We have two plots, one for line and one for block

### SSP02

# we want to compare effect of sex and day of starvation on pupal size. I am using df02, subset of data by day 0 and day 2  
model2<-lmer(pupa~sex\*day+(1|line)+(1|block), data=df02)  
model1<-lmer(pupa~sex\*day+(sex\*day|line)+(1|block), data=df02)  
anova(model1,model2)

## Data: df02  
## Models:  
## model2: pupa ~ sex \* day + (1 | line) + (1 | block)  
## model1: pupa ~ sex \* day + (sex \* day | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 7 -14010 -13958 7011.7 -14024   
## model1 16 -14905 -14788 7468.5 -14937 913.55 9 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Model 1 is better here.

### Do a LRT

How many parameters for each models

(AIC(model1) - REMLcrit(model1))/2 # # of parameters the model "thinks" are being estimated

## [1] 16

(AIC(model2) - REMLcrit(model2))/2 # # of parameters the model "thinks" are being estimated

## [1] 7

So lme4/lmer is treating model 1 as having nine more parameters than model2.

LR.model <- -as.numeric(REMLcrit(model1) - REMLcrit(model2))  
LR.model

## [1] 916.3482

nlevels(df02$line)

## [1] 0

pchisq(q = LR.model, df=9, lower=F)

## [1] 1.856931e-191

pchisq(q = LR.model, df=nlevels(df02$line), lower=F)

## [1] 0

### Parametric boostrap

Finally, we can conduct a parametric bootstrap to compare the two models.

### Finally using Bayesian Analysis

#did not run because need to update model

#prior.2 <-list(R=list(V=0.01, nu=0.002),   
 # G=list(G1=list(V=0.01\*diag(1), nu=0.002),  
 # G2=list(V=0.01\*diag(2), nu=0.002)))  
  
#model1M.MCMC <- MCMCglmm(pupa ~ 1 + sex,   
 # random=~block + us(1 + sex):line,  
# prior = prior.2, burnin = 5000, nitt = 20000, thin = 10,  
 # verbose = F, pr = T,  
# data=df02)  
#summary(model1M.MCMC)

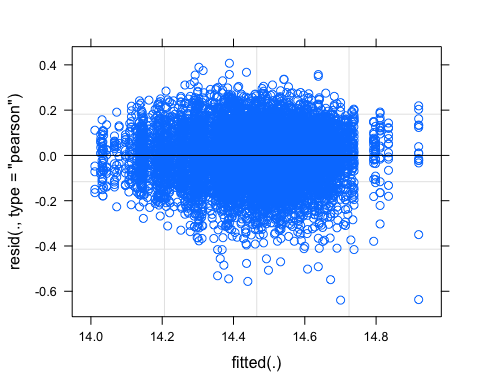
### Post model 1 fitting check

#### Residual distribution

res\_model1=residuals(model1)

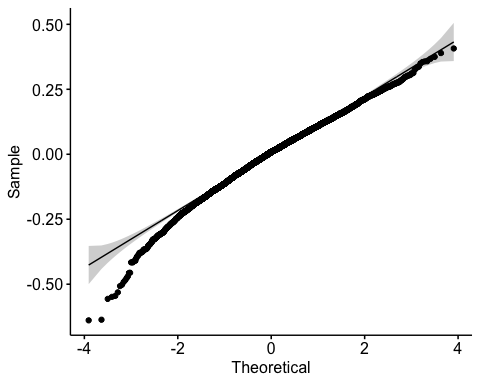
#### Model 1 residual distribution

plot(model1)



#### QQ plot

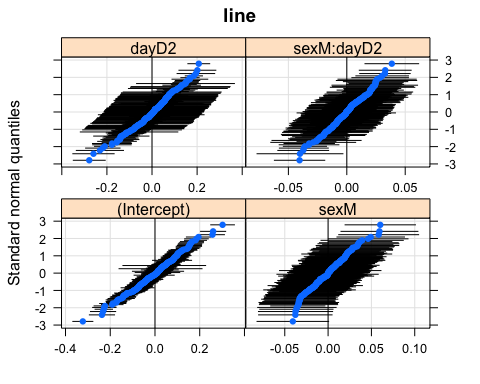
require(ggpubr)  
ggqqplot(res\_model1)



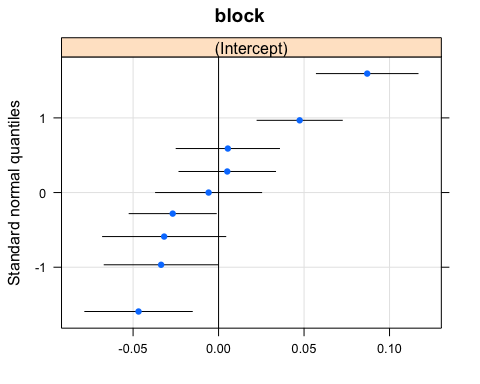
#### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block

 We have two plots, one for line and one for block

Answer 3.2: There is genetic variation in SSP if we compare fed flies and starved flies. BUT need to check with Alex if my model is correct.

##Question 3.3: Is the variation in SSP due to a variation in SSD in fed or starved flies? Look at the correlation between SSP01, SSD0 and SSD1 To calculate SSP, I first calculate the plasticity of female and male and subtract the plasticity female-male

#calculate plasticity female  
head(df1)

## # A tibble: 6 x 10  
## # Groups: group [5]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 317-B-1-F-7742 L317 B11 D1 F 13.7 13.4 13.8 L317\_F\_… -0.607  
## 2 911-B-1-M-096 L911 B7 D1 M 13.3 13.0 13.8 L911\_M\_… -0.537  
## 3 796-A-1-M-034 L796 B7 D1 M 13.6 13.6 13.8 L796\_M\_… -0.526  
## 4 105-A-1-M-7585 L105 B11 D1 M 13.5 13.4 13.8 L105\_M\_… -0.581  
## 5 105-B-1-M-7570 L105 B11 D1 M 13.5 13.2 13.8 L105\_M\_… -0.579  
## 6 861-B-1-M-101 L861 B7 D1 M 13.5 13.2 13.8 L861\_M\_… -0.524

df1F<-subset(df1, sex=="F")  
head(df1F)

## # A tibble: 6 x 10  
## # Groups: group [5]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 317-B-1-F-7742 L317 B11 D1 F 13.7 13.4 13.8 L317\_F\_… -0.607  
## 2 595-B-1-F-104 L595 B6 D1 F 13.7 13.2 13.9 L595\_F\_… -0.522  
## 3 584-B-1-F-7029 L584 B11 D1 F 13.9 13.5 13.9 L584\_F\_… -0.558  
## 4 595-B-1-F-088 L595 B6 D1 F 13.6 13.2 13.9 L595\_F\_… -0.496  
## 5 505-A-1-F-648 L505 B12 D1 F 14.1 13.8 13.9 L505\_F\_… -0.592  
## 6 392-B-1-F-084 L392 B6 D1 F 13.9 13.7 13.9 L392\_F\_… -0.495

df1F\_mean<-aggregate(df1F[, 10], list(df1F$line), mean)  
colnames(df1F\_mean)<-c("line","pupaFmean\_1")  
head(df1F\_mean)

## line pupaFmean\_1  
## 1 L100 0.08515325  
## 2 L101 -0.03520331  
## 3 L105 -0.08479509  
## 4 L109 -0.15244216  
## 5 L129 0.04989262  
## 6 L136 -0.09580059

df0F<-subset(df0, sex=="F")  
df0F\_mean<-aggregate(df0F[, 10], list(df0F$line), mean)  
colnames(df0F\_mean)<-c("line","pupaFmean\_0")  
head(df0F\_mean)

## line pupaFmean\_0  
## 1 L100 0.256259113  
## 2 L101 0.109003255  
## 3 L105 -0.003621951  
## 4 L109 0.063180232  
## 5 L136 0.184305984  
## 6 L138 0.207235102

df2F<-subset(df2, sex=="F")  
df2F\_mean<-aggregate(df2F[, 10], list(df2F$line), mean)  
colnames(df2F\_mean)<-c("line","pupaFmean\_2")  
head(df2F\_mean)

## line pupaFmean\_2  
## 1 L153 0.208349100  
## 2 L158 0.019870621  
## 3 L195 -0.324207734  
## 4 L229 0.025244072  
## 5 L256 -0.051082781  
## 6 L26 -0.002092585

pupa\_mean\_F01<-merge(x=df0F\_mean, y=df1F\_mean, by.x="line", by.y="line") #cannot merge all three treatment together or we loose data because less data at Day 2  
head(pupa\_mean\_F01) #151

## line pupaFmean\_0 pupaFmean\_1  
## 1 L100 0.256259113 0.085153254  
## 2 L101 0.109003255 -0.035203315  
## 3 L105 -0.003621951 -0.084795090  
## 4 L109 0.063180232 -0.152442163  
## 5 L136 0.184305984 -0.095800591  
## 6 L138 0.207235102 -0.009467483

#for SSP02  
pupa\_mean\_F02<-merge(x=df0F\_mean, y=df2F\_mean, by.x="line", by.y="line")  
head(pupa\_mean\_F02) #66 lines

## line pupaFmean\_0 pupaFmean\_2  
## 1 L158 0.4372961 0.019870621  
## 2 L195 0.2585795 -0.324207734  
## 3 L229 0.2342914 0.025244072  
## 4 L256 0.1506149 -0.051082781  
## 5 L26 0.2132552 -0.002092585  
## 6 L28 0.2039808 0.092029050

pupa\_mean\_F01$PF01<-pupa\_mean\_F01$pupaFmean\_0-pupa\_mean\_F01$pupaFmean\_1 #calculating plast female 01  
pupa\_mean\_F02$PF02<-pupa\_mean\_F02$pupaFmean\_0-pupa\_mean\_F02$pupaFmean\_2 #calculating plast female 02  
  
plastF01<-pupa\_mean\_F01   
plastF02<-pupa\_mean\_F02

#calculate plasticity male  
head(df1)

## # A tibble: 6 x 10  
## # Groups: group [5]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 317-B-1-F-7742 L317 B11 D1 F 13.7 13.4 13.8 L317\_F\_… -0.607  
## 2 911-B-1-M-096 L911 B7 D1 M 13.3 13.0 13.8 L911\_M\_… -0.537  
## 3 796-A-1-M-034 L796 B7 D1 M 13.6 13.6 13.8 L796\_M\_… -0.526  
## 4 105-A-1-M-7585 L105 B11 D1 M 13.5 13.4 13.8 L105\_M\_… -0.581  
## 5 105-B-1-M-7570 L105 B11 D1 M 13.5 13.2 13.8 L105\_M\_… -0.579  
## 6 861-B-1-M-101 L861 B7 D1 M 13.5 13.2 13.8 L861\_M\_… -0.524

df1M<-subset(df1, sex=="M")  
df1M\_mean<-aggregate(df1M[, 10], list(df1M$line), mean)  
colnames(df1M\_mean)<-c("line","pupaMmean\_1")  
head(df1M\_mean)

## line pupaMmean\_1  
## 1 L100 0.06014172  
## 2 L101 -0.10079021  
## 3 L105 -0.25153763  
## 4 L109 -0.25892157  
## 5 L129 -0.03625669  
## 6 L136 -0.14308417

df0M<-subset(df0, sex=="M")  
df0M\_mean<-aggregate(df0M[, 10], list(df0M$line), mean)  
colnames(df0M\_mean)<-c("line","pupaMmean\_0")  
head(df0M\_mean)

## line pupaMmean\_0  
## 1 L100 0.11768444  
## 2 L101 0.08961544  
## 3 L105 -0.05341221  
## 4 L109 -0.02019754  
## 5 L136 0.06867247  
## 6 L138 0.12057467

df2M<-subset(df2, sex=="M")  
df2M\_mean<-aggregate(df2M[, 10], list(df2M$line), mean)  
colnames(df2M\_mean)<-c("line","pupaMmean\_2")  
head(df2M\_mean)

## line pupaMmean\_2  
## 1 L100 -0.23292373  
## 2 L153 0.08699819  
## 3 L158 -0.07201030  
## 4 L177 -0.35313746  
## 5 L189 -0.30536655  
## 6 L229 -0.17136726

pupa\_mean\_M01<-merge(x=df0M\_mean, y=df1M\_mean, by.x="line", by.y="line") #cannot merge all three treatment together or we loose data because less data at Day 2  
head(pupa\_mean\_F01) #151

## line pupaFmean\_0 pupaFmean\_1 PF01  
## 1 L100 0.256259113 0.085153254 0.17110586  
## 2 L101 0.109003255 -0.035203315 0.14420657  
## 3 L105 -0.003621951 -0.084795090 0.08117314  
## 4 L109 0.063180232 -0.152442163 0.21562240  
## 5 L136 0.184305984 -0.095800591 0.28010657  
## 6 L138 0.207235102 -0.009467483 0.21670258

pupa\_mean\_M02<-merge(x=df0M\_mean, y=df2M\_mean, by.x="line", by.y="line")  
head(pupa\_mean\_M02) #66 lines

## line pupaMmean\_0 pupaMmean\_2  
## 1 L100 0.11768444 -0.23292373  
## 2 L153 0.17209388 0.08699819  
## 3 L158 0.29326145 -0.07201030  
## 4 L177 -0.02499460 -0.35313746  
## 5 L189 0.08990092 -0.30536655  
## 6 L229 0.10297417 -0.17136726

pupa\_mean\_M01$PM01<-pupa\_mean\_M01$pupaMmean\_0-pupa\_mean\_M01$pupaMmean\_1 #calculating SSP01  
pupa\_mean\_M02$PM02<-pupa\_mean\_M02$pupaMmean\_0-pupa\_mean\_M02$pupaMmean\_2 #calculating SSP02  
  
plastM01<-pupa\_mean\_M01   
plastM02<-pupa\_mean\_M02

#SSP01  
  
pupa\_SSP01<-merge(x=plastF01, y= plastM01, by.x="line", by.y="line") #merging male and female datasets  
head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.256259113 0.085153254 0.17110586 0.11768444 0.06014172 0.05754272  
## 2 L101 0.109003255 -0.035203315 0.14420657 0.08961544 -0.10079021 0.19040565  
## 3 L105 -0.003621951 -0.084795090 0.08117314 -0.05341221 -0.25153763 0.19812542  
## 4 L109 0.063180232 -0.152442163 0.21562240 -0.02019754 -0.25892157 0.23872403  
## 5 L136 0.184305984 -0.095800591 0.28010657 0.06867247 -0.14308417 0.21175665  
## 6 L138 0.207235102 -0.009467483 0.21670258 0.12057467 -0.04318888 0.16376355

length(unique(pupa\_SSP01$line)) #146 lines left

## [1] 146

#SSP02  
pupa\_SSP02<-merge(x=plastF02, y= plastM02, by.x="line", by.y="line") #merging male and female datasets  
head(pupa\_SSP02)

## line pupaFmean\_0 pupaFmean\_2 PF02 pupaMmean\_0 pupaMmean\_2 PM02  
## 1 L158 0.43729608 0.019870621 0.4174255 0.293261448 -0.072010303 0.36527175  
## 2 L229 0.23429138 0.025244072 0.2090473 0.102974168 -0.171367258 0.27434143  
## 3 L256 0.15061487 -0.051082781 0.2016977 0.090507203 -0.174299175 0.26480638  
## 4 L26 0.21325524 -0.002092585 0.2153478 0.103814444 -0.104164760 0.20797920  
## 5 L28 0.20398076 0.092029050 0.1119517 0.145080084 -0.003598201 0.14867829  
## 6 L313 0.08642449 -0.117978446 0.2044029 -0.002395298 -0.096055488 0.09366019

length(unique(pupa\_SSP02$line)) #55 lines left

## [1] 55

#calculate SSP from the plasticity.  
pupa\_SSP01$SSP01<-pupa\_SSP01$PF01-pupa\_SSP01$PM01  
  
pupa\_SSP02$SSP02<-pupa\_SSP02$PF02-pupa\_SSP02$PM02  
  
  
#calculating SSDs again to add them to the dataframe  
pupa\_SSP01$SSD0<-pupa\_SSP01$pupaFmean\_0-pupa\_SSP01$pupaMmean\_0  
pupa\_SSP01$SSD1<-pupa\_SSP01$pupaFmean\_1-pupa\_SSP01$pupaMmean\_1  
  
  
pupa\_SSP02$SSD0<-pupa\_SSP02$pupaFmean\_0-pupa\_SSP02$pupaMmean\_0  
pupa\_SSP02$SSD2<-pupa\_SSP02$pupaFmean\_2-pupa\_SSP02$pupaMmean\_2

#### Correlation between SSP and SSDs

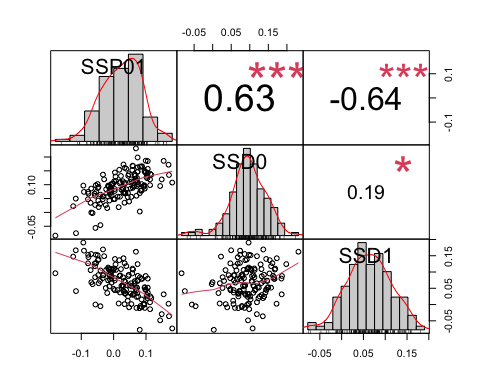
##### SSP01, SSD0 and SSD1

Is SSD0 or 1 covarying with SSP01?

#use pupa\_SSP01   
head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.256259113 0.085153254 0.17110586 0.11768444 0.06014172 0.05754272  
## 2 L101 0.109003255 -0.035203315 0.14420657 0.08961544 -0.10079021 0.19040565  
## 3 L105 -0.003621951 -0.084795090 0.08117314 -0.05341221 -0.25153763 0.19812542  
## 4 L109 0.063180232 -0.152442163 0.21562240 -0.02019754 -0.25892157 0.23872403  
## 5 L136 0.184305984 -0.095800591 0.28010657 0.06867247 -0.14308417 0.21175665  
## 6 L138 0.207235102 -0.009467483 0.21670258 0.12057467 -0.04318888 0.16376355  
## SSP01 SSD0 SSD1  
## 1 0.11356314 0.13857467 0.02501154  
## 2 -0.04619908 0.01938781 0.06558689  
## 3 -0.11695228 0.04979026 0.16674254  
## 4 -0.02310163 0.08337778 0.10647941  
## 5 0.06834993 0.11563351 0.04728358  
## 6 0.05293904 0.08666043 0.03372139

SSP01<-pupa\_SSP01[c(1,8:10)]  
chart.Correlation(SSP01[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

 The highest correlation is between SSD1 and SSP01, and it is a negative correlation, which means that the higher SSP is, the more SSD decreases, at lower food condition, one of the two sexes decreases in size more than it increases in size in fed flies.

#### Correlation SSP01, female Day 1 and male Day 1 size

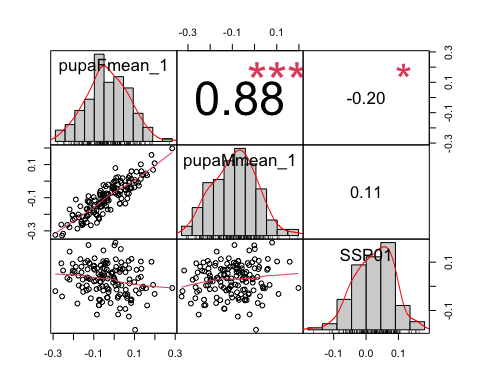
head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.256259113 0.085153254 0.17110586 0.11768444 0.06014172 0.05754272  
## 2 L101 0.109003255 -0.035203315 0.14420657 0.08961544 -0.10079021 0.19040565  
## 3 L105 -0.003621951 -0.084795090 0.08117314 -0.05341221 -0.25153763 0.19812542  
## 4 L109 0.063180232 -0.152442163 0.21562240 -0.02019754 -0.25892157 0.23872403  
## 5 L136 0.184305984 -0.095800591 0.28010657 0.06867247 -0.14308417 0.21175665  
## 6 L138 0.207235102 -0.009467483 0.21670258 0.12057467 -0.04318888 0.16376355  
## SSP01 SSD0 SSD1  
## 1 0.11356314 0.13857467 0.02501154  
## 2 -0.04619908 0.01938781 0.06558689  
## 3 -0.11695228 0.04979026 0.16674254  
## 4 -0.02310163 0.08337778 0.10647941  
## 5 0.06834993 0.11563351 0.04728358  
## 6 0.05293904 0.08666043 0.03372139

SSP01\_size<-pupa\_SSP01[,c(1,3,6,8)]  
head(SSP01\_size)

## line pupaFmean\_1 pupaMmean\_1 SSP01  
## 1 L100 0.085153254 0.06014172 0.11356314  
## 2 L101 -0.035203315 -0.10079021 -0.04619908  
## 3 L105 -0.084795090 -0.25153763 -0.11695228  
## 4 L109 -0.152442163 -0.25892157 -0.02310163  
## 5 L136 -0.095800591 -0.14308417 0.06834993  
## 6 L138 -0.009467483 -0.04318888 0.05293904

chart.Correlation(SSP01\_size[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

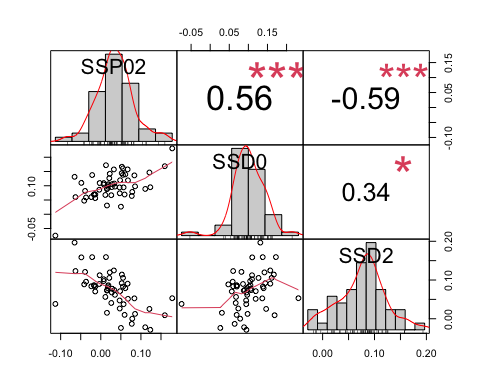
 There is almost no correlation between SSP01 and male and weak correlation with female size in Day 1 starvation. That means that SSP01 covaries negatively with SSD01,(when the difference in plasticity is higher, then sexual size dimorphism decreases) We assumed that this covariation was created by the variation in female size in Day1, but it does not explain it very strongly?

##### Do we have the same trend in SSP02, SSD0 and SSD2?

head(pupa\_SSP02)

## line pupaFmean\_0 pupaFmean\_2 PF02 pupaMmean\_0 pupaMmean\_2 PM02  
## 1 L158 0.43729608 0.019870621 0.4174255 0.293261448 -0.072010303 0.36527175  
## 2 L229 0.23429138 0.025244072 0.2090473 0.102974168 -0.171367258 0.27434143  
## 3 L256 0.15061487 -0.051082781 0.2016977 0.090507203 -0.174299175 0.26480638  
## 4 L26 0.21325524 -0.002092585 0.2153478 0.103814444 -0.104164760 0.20797920  
## 5 L28 0.20398076 0.092029050 0.1119517 0.145080084 -0.003598201 0.14867829  
## 6 L313 0.08642449 -0.117978446 0.2044029 -0.002395298 -0.096055488 0.09366019  
## SSP02 SSD0 SSD2  
## 1 0.052153707 0.14403463 0.09188092  
## 2 -0.065294122 0.13131721 0.19661133  
## 3 -0.063108726 0.06010767 0.12321639  
## 4 0.007368619 0.10944079 0.10207217  
## 5 -0.036726573 0.05890068 0.09562725  
## 6 0.110742750 0.08881979 -0.02192296

SSP02<-pupa\_SSP02[c(1,8:10)]  
chart.Correlation(SSP02[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

 SSP02 correlates the most with SSD2 and negatively, same as we found between SSP01 and SSD1.

If SSPs covary wit starved SSDs, that means that the variation in starved flies is more important than the variation in fed flies.

### Question 3.4: If it is due to SSD in either fed or starved flies, is it because the females vary more or the males vary more in size?

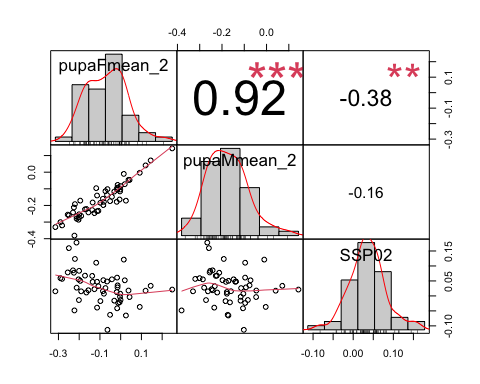
SSP02 covaries negatively with SSD2, which sex at Day 2 contributes to SSD2 variation?

head(pupa\_SSP02)

## line pupaFmean\_0 pupaFmean\_2 PF02 pupaMmean\_0 pupaMmean\_2  
## 1 L158 0.43729608 0.0198706211 0.41742546 0.293261448 -0.072010303  
## 2 L229 0.23429138 0.0252440717 0.20904730 0.102974168 -0.171367258  
## 3 L256 0.15061487 -0.0510827809 0.20169765 0.090507203 -0.174299175  
## 4 L26 0.21325524 -0.0020925854 0.21534782 0.103814444 -0.104164760  
## 5 L28 0.20398076 0.0920290502 0.11195171 0.145080084 -0.003598201  
  
## PM02 SSP02 SSD0 SSD2  
## 1 0.36527175 0.052153707 0.14403463 0.0918809238  
## 2 0.27434143 -0.065294122 0.13131721 0.1966113298  
## 3 0.26480638 -0.063108726 0.06010767 0.1232163941  
## 4 0.20797920 0.007368619 0.10944079 0.1020721748  
## 5 0.14867829 -0.036726573 0.05890068 0.0956272516  
SSP02size<-pupa\_SSP02[,c(1,3,6,8)]  
head(SSP02size)

## line pupaFmean\_2 pupaMmean\_2 SSP02  
## 1 L158 0.019870621 -0.072010303 0.052153707  
## 2 L229 0.025244072 -0.171367258 -0.065294122  
## 3 L256 -0.051082781 -0.174299175 -0.063108726  
## 4 L26 -0.002092585 -0.104164760 0.007368619  
## 5 L28 0.092029050 -0.003598201 -0.036726573  
## 6 L313 -0.117978446 -0.096055488 0.110742750

chart.Correlation(SSP02size[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)



### Question 3.5: Is the variation causing SSP (Question 3.4) the same as the variation caused by SSD in Question 1.3?

We saw that we only see a covariation between SSP and SSD starved, and female starved.

In Question 1.3, we found that SSD0 (in fed flies) is correlated with fed female size, but that fed female size variation does not correlate with SSPs.

If there is no direct correlation between SSP and SSD in fed flies, let’s check how the variation of plasticity between each sexes covaries with SSP.

### Question 3.6: Do we see a correlation between SSP and plasticity

#### Correlation SSP01, plasticity female, plasticity male

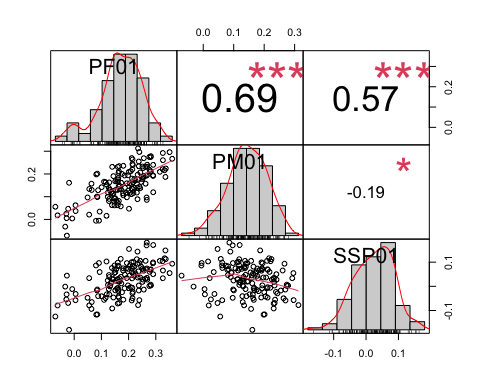
head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.256259113 0.085153254 0.17110586 0.11768444 0.06014172 0.05754272  
## 2 L101 0.109003255 -0.035203315 0.14420657 0.08961544 -0.10079021 0.19040565  
## 3 L105 -0.003621951 -0.084795090 0.08117314 -0.05341221 -0.25153763 0.19812542  
## 4 L109 0.063180232 -0.152442163 0.21562240 -0.02019754 -0.25892157 0.23872403  
## 5 L136 0.184305984 -0.095800591 0.28010657 0.06867247 -0.14308417 0.21175665  
## 6 L138 0.207235102 -0.009467483 0.21670258 0.12057467 -0.04318888 0.16376355  
## SSP01 SSD0 SSD1  
## 1 0.11356314 0.13857467 0.02501154  
## 2 -0.04619908 0.01938781 0.06558689  
## 3 -0.11695228 0.04979026 0.16674254  
## 4 -0.02310163 0.08337778 0.10647941  
## 5 0.06834993 0.11563351 0.04728358  
## 6 0.05293904 0.08666043 0.03372139

SSP01\_plast<-pupa\_SSP01[,c(1,4,7,8)]  
head(SSP01\_plast)

## line PF01 PM01 SSP01  
## 1 L100 0.17110586 0.05754272 0.11356314  
## 2 L101 0.14420657 0.19040565 -0.04619908  
## 3 L105 0.08117314 0.19812542 -0.11695228  
## 4 L109 0.21562240 0.23872403 -0.02310163  
## 5 L136 0.28010657 0.21175665 0.06834993  
## 6 L138 0.21670258 0.16376355 0.05293904

chart.Correlation(SSP01\_plast[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)

 SSP01 is correlated with female plasticity.

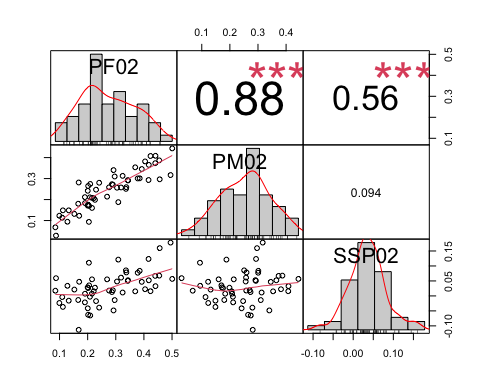
head(pupa\_SSP02)

## line pupaFmean\_0 pupaFmean\_2 PF02 pupaMmean\_0 pupaMmean\_2 PM02  
## 1 L158 0.43729608 0.019870621 0.4174255 0.293261448 -0.072010303 0.36527175  
## 2 L229 0.23429138 0.025244072 0.2090473 0.102974168 -0.171367258 0.27434143  
## 3 L256 0.15061487 -0.051082781 0.2016977 0.090507203 -0.174299175 0.26480638  
## 4 L26 0.21325524 -0.002092585 0.2153478 0.103814444 -0.104164760 0.20797920  
## 5 L28 0.20398076 0.092029050 0.1119517 0.145080084 -0.003598201 0.14867829  
## 6 L313 0.08642449 -0.117978446 0.2044029 -0.002395298 -0.096055488 0.09366019  
## SSP02 SSD0 SSD2  
## 1 0.052153707 0.14403463 0.09188092  
## 2 -0.065294122 0.13131721 0.19661133  
## 3 -0.063108726 0.06010767 0.12321639  
## 4 0.007368619 0.10944079 0.10207217  
## 5 -0.036726573 0.05890068 0.09562725  
## 6 0.110742750 0.08881979 -0.02192296

SSP02\_plast<-pupa\_SSP02[,c(1,4,7,8)]  
head(SSP02\_plast)

## line PF02 PM02 SSP02  
## 1 L158 0.4174255 0.36527175 0.052153707  
## 2 L229 0.2090473 0.27434143 -0.065294122  
## 3 L256 0.2016977 0.26480638 -0.063108726  
## 4 L26 0.2153478 0.20797920 0.007368619  
## 5 L28 0.1119517 0.14867829 -0.036726573  
## 6 L313 0.2044029 0.09366019 0.110742750

chart.Correlation(SSP02\_plast[2:4],  
 method="pearson",  
 histogram=TRUE,  
 pch=16)



SSP02 also covaries with female plasticity. In both cases, we see that SSPs are correlated with a variation in female plasticity.

Now, I want to see if the variance of female plasticity is significantly different than the variance of male plasticity.

For plasticity 0-1

#dataset   
  
head(SSP01\_plast)

## line PF01 PM01 SSP01  
## 1 L100 0.17110586 0.05754272 0.11356314  
## 2 L101 0.14420657 0.19040565 -0.04619908  
## 3 L105 0.08117314 0.19812542 -0.11695228  
## 4 L109 0.21562240 0.23872403 -0.02310163  
## 5 L136 0.28010657 0.21175665 0.06834993  
## 6 L138 0.21670258 0.16376355 0.05293904

#one tail F test  
  
var.test(SSP01\_plast$PF01, SSP01\_plast$PM01, alternative = "greater")

##   
## F test to compare two variances  
##   
## data: SSP01\_plast$PF01 and SSP01\_plast$PM01  
## F = 1.4386, num df = 145, denom df = 145, p-value = 0.01463  
## alternative hypothesis: true ratio of variances is greater than 1  
## 95 percent confidence interval:  
## 1.093735 Inf  
## sample estimates:  
## ratio of variances   
## 1.43863

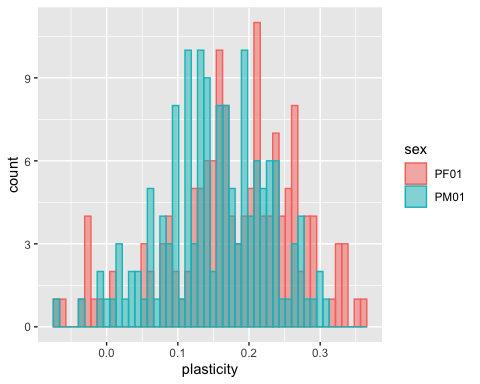
Yes the female plasticity varies more than male plasticity

Histogram of female and male plasticity

SSP01\_plast

## line PF01 PM01 SSP01  
## 1 L100 0.171105859 0.057542724 0.1135631358  
## 2 L101 0.144206569 0.190405651 -0.0461990817  
## 3 L105 0.081173139 0.198125420 -0.1169522814  
## 4 L109 0.215622395 0.238724029 -0.0231016333  
## 5 L136 0.280106575 0.211756645 0.0683499294  
plasticity01<-SSP01\_plast %>% gather(sex, plasticity, PF01:PM01)  
plasticity01

## line SSP01 sex plasticity  
## 1 L100 0.1135631358 PF01 0.171105859  
## 2 L101 -0.0461990817 PF01 0.144206569  
## 3 L105 -0.1169522814 PF01 0.081173139  
## 4 L109 -0.0231016333 PF01 0.215622395  
## 5 L136 0.0683499294 PF01 0.280106575  
ggplot(plasticity01, aes(x=plasticity, fill=sex, color=sex)) +  
 geom\_histogram(position="identity", alpha=0.5, bins=50)



#dataset   
  
head(SSP02\_plast)

## line PF02 PM02 SSP02  
## 1 L158 0.4174255 0.36527175 0.052153707  
## 2 L229 0.2090473 0.27434143 -0.065294122  
## 3 L256 0.2016977 0.26480638 -0.063108726  
## 4 L26 0.2153478 0.20797920 0.007368619  
## 5 L28 0.1119517 0.14867829 -0.036726573  
## 6 L313 0.2044029 0.09366019 0.110742750

#one tail F test  
  
var.test(SSP02\_plast$PF02, SSP02\_plast$PM02, alternative = "greater")

##   
## F test to compare two variances  
##   
## data: SSP02\_plast$PF02 and SSP02\_plast$PM02  
## F = 1.4446, num df = 54, denom df = 54, p-value = 0.08986  
## alternative hypothesis: true ratio of variances is greater than 1  
## 95 percent confidence interval:  
## 0.9196404 Inf  
## sample estimates:  
## ratio of variances   
## 1.444648

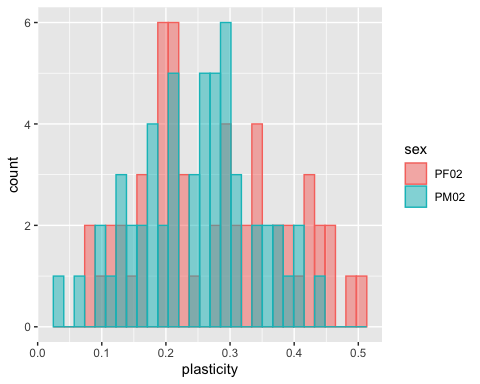
In both cases, plasticity if female does not vary more than plasticity of male.

head(SSP02\_plast)

## line PF02 PM02 SSP02  
## 1 L158 0.41742546 0.36527175 0.052153707  
## 2 L229 0.20904730 0.27434143 -0.065294122  
## 3 L256 0.20169765 0.26480638 -0.063108726  
## 4 L26 0.21534782 0.20797920 0.007368619  
## 5 L28 0.11195171 0.14867829 -0.036726573

plasticity02<-SSP02\_plast %>% gather(sex, plasticity, PF02:PM02)  
head(plasticity02)

## line SSP02 sex plasticity  
## 1 L158 0.052153707 PF02 0.41742546  
## 2 L229 -0.065294122 PF02 0.20904730  
## 3 L256 -0.063108726 PF02 0.20169765  
## 4 L26 0.007368619 PF02 0.21534782  
## 5 L28 -0.036726573 PF02 0.11195171  
ggplot(plasticity02, aes(x=plasticity, fill=sex, color=sex)) +  
 geom\_histogram(position="identity", alpha=0.5, bins=30)

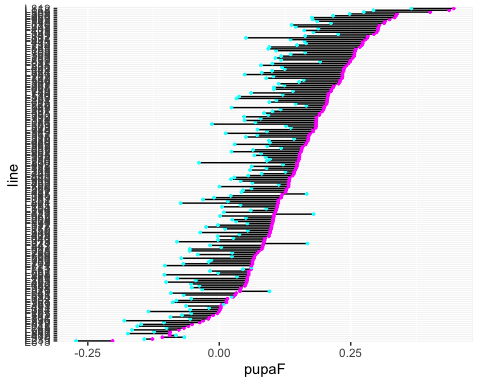


#plots ##Compare size male and female fed flies across lines: plotting SSD in decreasing number

head(df0\_mean)

## line pupaF pupaM SSD0  
## 1 L100 0.256259113 0.11768444 0.13857467  
## 2 L101 0.109003255 0.08961544 0.01938781  
## 3 L105 -0.003621951 -0.05341221 0.04979026  
## 4 L109 0.063180232 -0.02019754 0.08337778  
## 5 L136 0.184305984 0.06867247 0.11563351  
## 6 L138 0.207235102 0.12057467 0.08666043

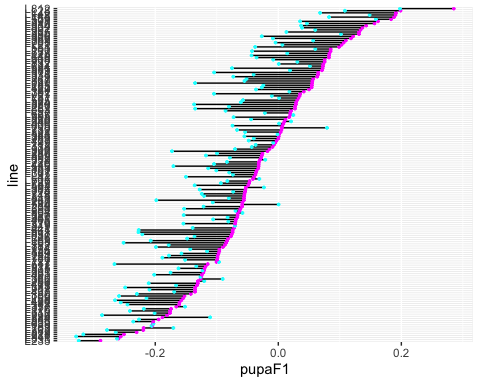
df0\_mean$line <- factor(df0\_mean$line, levels = df0\_mean$line[order(df0\_mean$pupaF)])  
  
ggplot(df0\_mean,aes(x=pupaF, xend=pupaM, y=line, group=line))+  
geom\_dumbbell(size=0.5, color="black",  
colour\_x = "magenta", colour\_xend = "cyan",  
dot\_guide=F, dot\_guide\_size=0.5)



head(df1\_mean)

## line pupaF1 pupaM1 SSD1  
## 1 L100 0.08515325 0.06014172 0.02501154  
## 2 L101 -0.03520331 -0.10079021 0.06558689  
## 3 L105 -0.08479509 -0.25153763 0.16674254  
## 4 L109 -0.15244216 -0.25892157 0.10647941  
## 5 L129 0.04989262 -0.03625669 0.08614931  
## 6 L136 -0.09580059 -0.14308417 0.04728358

df1\_mean$line <- factor(df1\_mean$line, levels = df1\_mean$line[order(df1\_mean$pupaF1)])  
  
ggplot(df1\_mean,aes(x=pupaF1, xend=pupaM1, y=line, group=line))+  
geom\_dumbbell(size=0.5, color="black",  
colour\_x = "magenta", colour\_xend = "cyan",  
dot\_guide=F, dot\_guide\_size=0.5)



head(df2\_mean)

## line pupaF2 pupaM2 SSD2  
## 1 L153 0.208349100 0.086998187 0.12135091  
## 2 L158 0.019870621 -0.072010303 0.09188092  
## 3 L229 0.025244072 -0.171367258 0.19661133  
## 4 L256 -0.051082781 -0.174299175 0.12321639  
## 5 L26 -0.002092585 -0.104164760 0.10207217  
## 6 L28 0.092029050 -0.003598201 0.09562725

df2\_mean$line <- factor(df2\_mean$line, levels = df2\_mean$line[order(df2\_mean$pupaF2)])  
  
ggplot(df2\_mean,aes(x=pupaF2, xend=pupaM2, y=line, group=line))+  
geom\_dumbbell(size=0.5, color="black",  
colour\_x = "magenta", colour\_xend = "cyan",  
dot\_guide=F, dot\_guide\_size=0.5)

