DGRP Genetic Arch

Alexander Shingleton

9/8/2020

# Sept 22 2020

* Partial correlations
* Refine the questions, remove D2
* rerun lm on no clock pupae to see if the block effect is gone
* do wing and leg

#Pre-processing of data - n<10 groups removed - remove block effect when calculating SSD and SSP

# 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 starved flies - use D1
  + Question 2.1: Is SSD in starved flies the same as in fed flies? (do a lmer sex\*day on D0 and D1)
    - Does SSD increase or decrease when the flies are starved? Our hypothesis is that overall SSD should decrease
  + Question 2.2: Does the SSD in starved conditions vary the same way as in fed flies?
    - use a correlation and variance analysis

We have established that SSD0 and SSD1 are not the same. If they were the same, that would mean that the body size of male and female changed in the same way in different environment. We know that males and females do not change the same way so that means we have Sex Specific Plasticity.

-Section 3: Male and Female Plasticity - Question 3.1: Which sex is more plastic and does plasticity vary among lines? - Question 3.2: Which sex plasticity contributes to SSP.

[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: Isthe variation causing SSP (Question 3.4) the same as the variation caused by SSD in Question 1.3?]

To answer these questions, I will use the size of fed flies and flies starved for 1 day before pupariation.

# Data preparation

importing file and checking for column names

#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 a letter in front of factorial columns

#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,2)

## id line block day sex wing leg pupa group  
## 1 021-A-0-F-085 L21 B6 D0 F 14.23040 13.93253 14.70779 L21\_F\_D0  
## 2 021-A-0-F-086 L21 B6 D0 F 13.99867 13.65442 14.43095 L21\_F\_D0

### KEEP THIS HERE BUT RUN BEFORE ANALYSES

#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)

## Is there variation among blocks = OK

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)   
  
#how many Na in pupae  
sum(is.na(df\_sub$pupa)) #1970 Nas

## [1] 1970

#remove NA in wing column  
df\_sub<-df\_sub[!is.na(df\_sub$pupa), ]  
  
#subset fed and 1day starved flies  
df\_sub01<-subset(df\_sub, day=="D0"|day=="D1")  
  
#lm with block fixed effect  
blocktest01<-lm(pupa~block, data=df\_sub01)  
summary(blocktest01)

##   
## Call:  
## lm(formula = pupa ~ block, data = df\_sub01)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.64395 -0.12171 0.00728 0.12808 0.65787   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 14.480592 0.004798 3018.031 < 2e-16 \*\*\*  
## blockB11 -0.019641 0.007123 -2.757 0.00584 \*\*   
## blockB12 0.014971 0.006733 2.223 0.02620 \*   
## blockB13 -0.009908 0.007227 -1.371 0.17038   
## blockB4 -0.031259 0.006391 -4.892 1.01e-06 \*\*\*  
## blockB5 -0.029399 0.005959 -4.933 8.17e-07 \*\*\*  
## blockB6 -0.083454 0.006264 -13.322 < 2e-16 \*\*\*  
## blockB7 -0.092453 0.006147 -15.039 < 2e-16 \*\*\*  
## blockB8 -0.052912 0.006674 -7.928 2.38e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1809 on 15385 degrees of freedom  
## Multiple R-squared: 0.03598, Adjusted R-squared: 0.03547   
## F-statistic: 71.77 on 8 and 15385 DF, p-value: < 2.2e-16

Anova(blocktest01)

## Anova Table (Type II tests)  
##   
## Response: pupa  
## Sum Sq Df F value Pr(>F)   
## block 18.78 8 71.768 < 2.2e-16 \*\*\*  
## Residuals 503.29 15385   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#move residuals to dataframe  
df\_sub01$pupa\_noblock <- blocktest01$resid  
head(df\_sub01)

## # A tibble: 6 x 10  
## # Groups: group [1]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 021-A-0-F-085 L21 B6 D0 F 14.2 13.9 14.7 L21\_F\_D0 0.311   
## 2 021-A-0-F-086 L21 B6 D0 F 14.0 13.7 14.4 L21\_F\_D0 0.0338  
## 3 021-A-0-F-087 L21 B6 D0 F 13.9 13.6 14.5 L21\_F\_D0 0.0702  
## 4 021-A-0-F-088 L21 B6 D0 F 13.9 13.8 14.4 L21\_F\_D0 0.0243  
## 5 021-A-0-F-089 L21 B6 D0 F 14.2 13.9 14.7 L21\_F\_D0 0.265   
## 6 021-A-0-F-090 L21 B6 D0 F 14.1 13.8 14.6 L21\_F\_D0 0.219

#df\_sub01$pupanoblock is the column I will use when I want to calculate summary and indexes for correlation analyses  
  
#rerun lm on pupa\_noblock  
blocktest01\_2<-lm(pupa\_noblock~block, data=df\_sub01)  
summary(blocktest01\_2)

##   
## Call:  
## lm(formula = pupa\_noblock ~ block, data = df\_sub01)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.64395 -0.12171 0.00728 0.12808 0.65787   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.199e-16 4.798e-03 0 1  
## blockB11 1.190e-17 7.123e-03 0 1  
## blockB12 -8.512e-18 6.733e-03 0 1  
## blockB13 2.734e-16 7.227e-03 0 1  
## blockB4 -3.380e-16 6.391e-03 0 1  
## blockB5 1.069e-15 5.959e-03 0 1  
## blockB6 2.592e-16 6.264e-03 0 1  
## blockB7 1.940e-17 6.147e-03 0 1  
## blockB8 5.121e-19 6.674e-03 0 1  
##   
## Residual standard error: 0.1809 on 15385 degrees of freedom  
## Multiple R-squared: 5.867e-30, Adjusted R-squared: -0.00052   
## F-statistic: 1.128e-26 on 8 and 15385 DF, p-value: 1

Anova(blocktest01\_2) #block effect is gone!

## Anova Table (Type II tests)  
##   
## Response: pupa\_noblock  
## Sum Sq Df F value Pr(>F)  
## block 0.00 8 0 1  
## Residuals 503.29 15385

There is a difference between collecting blocks so we have to account for block as a random factor when using lmer, or use the adjusted values to calculate SSD and SSP values.

# Section 1: SSD in fed flies

* Question 1.1: Is there Sexual Size Dimorphism in the DGRP lines. A1.1: YES Plot box plot -> should I use adjusted value?
* Question 1.2: Do we see a genetic variation of SSD in the DGRP lines. A1.2: YES Plot mean female size and male size on a bar graph per line Plot SSD on a separate bar graph
* Question 1.3: Is the SSD genetic variation due to a genetic variation in male or in female size? FEMALE Plot correlation between SSD0 and female size, and SSD0 and male size in same Figure

## 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\_sub01 #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") #8500  
head(df0,2)

## # A tibble: 2 x 10  
## # Groups: group [1]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 021-A-0-F-085 L21 B6 D0 F 14.2 13.9 14.7 L21\_F\_D0 0.311   
## 2 021-A-0-F-086 L21 B6 D0 F 14.0 13.7 14.4 L21\_F\_D0 0.0338

#na.omit only if pupa has NA if not done before  
#df0<-na.omit(df0, cols="pupa")

### Linear Mixed Model on fed flies

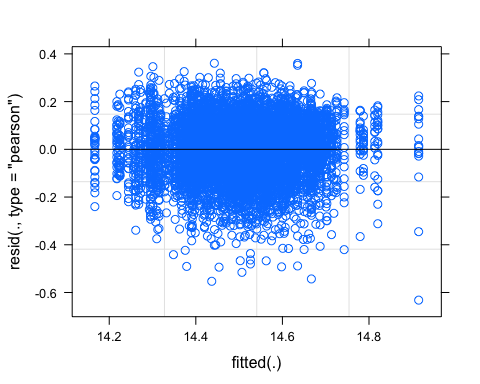
# Testing effect of sex in pupa size, with random effect for line and block.  
SSDtest<-lmer(pupa~sex+(1|line) +(1|block), REML=TRUE, data=df0)  
summary(SSDtest) #Number of obs: 8065, groups: line, 187; block, 9

## 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: -11807  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.4453 -0.6081 0.0870 0.6708 3.1077   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.010007 0.10003   
## block (Intercept) 0.002205 0.04696   
## Residual 0.013471 0.11607   
## Number of obs: 8500, groups: line, 187; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.456e+01 1.758e-02 1.003e+01 828.3 <2e-16 \*\*\*  
## sexM -9.443e-02 2.559e-03 8.326e+03 -36.9 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sexM -0.076

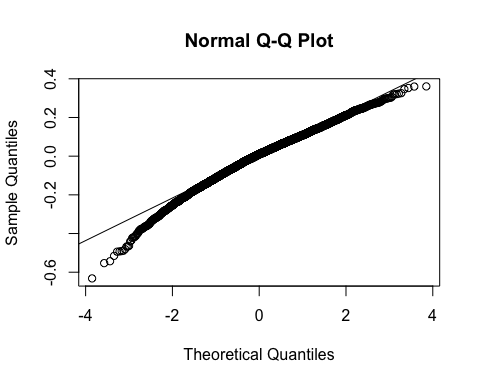
Anova(SSDtest)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: pupa  
## Chisq Df Pr(>Chisq)   
## sex 1361.9 1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

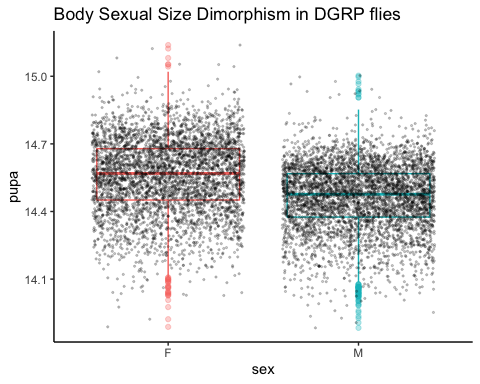
plot(SSDtest)



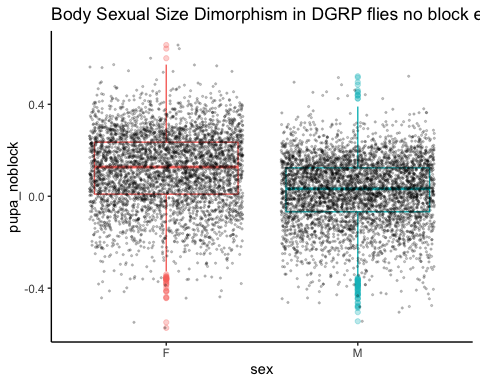
qqnorm(resid(SSDtest))  
qqline(resid(SSDtest))

 ### Box plot showing average difference in pupal size between male and female => ok Plot with block effect

# original pupa size  
(SSDtest\_plot <- ggplot(df0, aes(x = sex, y = pupa, colour = sex)) +  
 geom\_boxplot(alpha=0.3) +  
 theme\_classic() +  
 geom\_jitter(color="black", size=0.4, alpha=0.2)+  
 theme(legend.position = "none",  
 panel.spacing = unit(1, "lines")) + # adding space between panels  
 ggtitle("Body Sexual Size Dimorphism in DGRP flies")  
 )

 Plot adjusted no block effect

#removed block effect pupal size  
  
(SSDtest\_plot\_noblock <- ggplot(df0, aes(x = sex, y = pupa\_noblock, colour = sex)) +  
 geom\_boxplot(alpha=0.3) +  
 theme\_classic() +  
 geom\_jitter(color="black", size=0.4, alpha=0.2)+  
   
 theme(legend.position = "none",  
 panel.spacing = unit(1, "lines")) + # adding space between panels  
 ggtitle("Body Sexual Size Dimorphism in DGRP flies no block effect")  
 )



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 9.8828 9.8828 1 155.3 745.76 < 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 -11813 -11778 5911.7 -11823   
## model1 7 -11851 -11801 5932.3 -11865 41.191 2 1.136e-09 \*\*\*  
## ---  
## 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.

summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex + (sex | line) + (1 | block)  
## Data: df0  
##   
## REML criterion at convergence: -11848.8  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.5372 -0.6019 0.0826 0.6674 3.1766   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## line (Intercept) 0.0112218 0.10593   
## sexM 0.0009334 0.03055 -0.45  
## block (Intercept) 0.0021387 0.04625   
## Residual 0.0132521 0.11512   
## Number of obs: 8500, groups: line, 187; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 14.562972 0.017558 10.585965 829.42 <2e-16 \*\*\*  
## sexM -0.094663 0.003466 155.296229 -27.31 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sexM -0.190

In the model summary we see that the fitted value is smaller for males, so SSD is female biased even when considering genetic variation of SSD.

#### 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] 41.76099

nlevels(df$line)

## [1] 0

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

## [1] 8.545078e-10

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

## [1] 0

#### Parametric boostrap

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

#### Finally using Bayesian Analysis

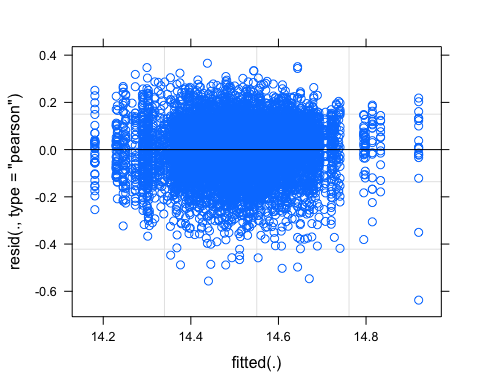
#prior.2 <-list(R=list(V=0.01, nu=0.002),   
 # G=list(G1=list(V=0.01\*diag(1), nu=0.002),  
 # G2=list(V=0.01\*diag(2), nu=0.002)))  
  
#model1M.MCMC <- MCMCglmm(pupa ~ 1 + sex,   
 # random=~block + us(1 + sex):line,  
 #prior = prior.2, burnin = 5000, nitt = 20000, thin = 10,  
 #verbose = F, pr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

#### Post model 1 fitting check

##### Residual distribution

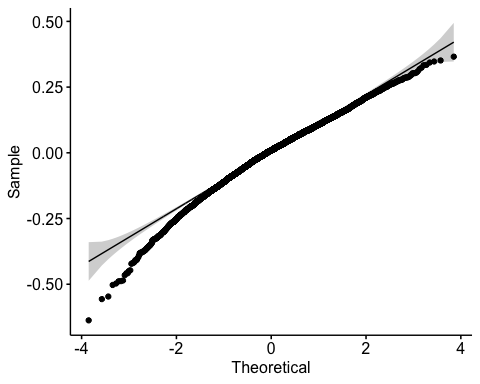
##### Model 1 residual distribution

plot(model1)



##### QQ plot

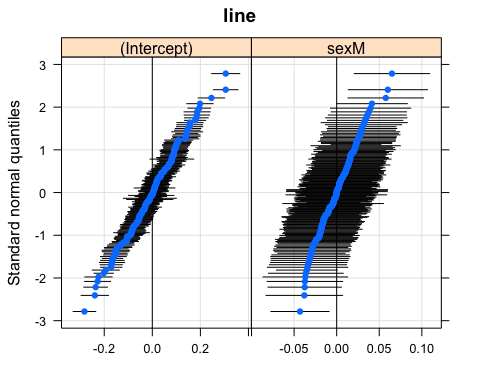
res\_model1=residuals(model1)  
ggqqplot(res\_model1)



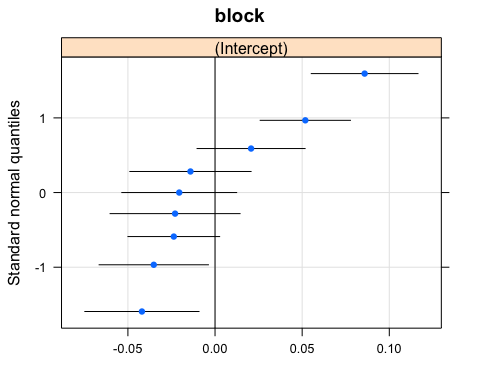
##### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block



### Plot mean size female and male per line + SSD per line

Calculate mean size per line

#calculate means for each group using pupa\_noblock  
head(df0,2)

## # A tibble: 2 x 10  
## # Groups: group [1]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 021-A-0-F-085 L21 B6 D0 F 14.2 13.9 14.7 L21\_F\_D0 0.311   
## 2 021-A-0-F-086 L21 B6 D0 F 14.0 13.7 14.4 L21\_F\_D0 0.0338

df0\_mean<-aggregate(df0[, 10], list(df0$group), mean)  
head(df0\_mean,2)

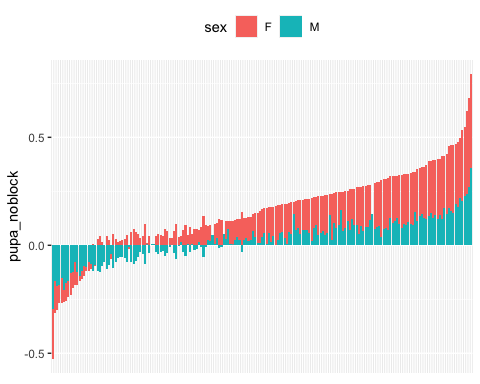
## Group.1 pupa\_noblock  
## 1 L100\_F\_D0 0.23636559  
## 2 L100\_M\_D0 0.09779092

#re-add line, day and sex columns  
df0\_mean<-df0\_mean %>%  
 separate(Group.1, c("line", "sex","day"), "\_")  
head(df0\_mean,2)

## line sex day pupa\_noblock  
## 1 L100 F D0 0.23636559  
## 2 L100 M D0 0.09779092

Plot Mean size by line for fed flies

meanpupa0\_plot<-ggplot(df0\_mean, aes(x=reorder(line,pupa\_noblock), y=pupa\_noblock)) +  
 geom\_col(aes(fill = sex)) +  
 #facet\_wrap(~ day) +  
 #ggtitle("Mean pupa size per sex for each line in fed flies, without block effect")  
theme(legend.direction = "horizontal",legend.position = "top",axis.title.x = element\_blank(),axis.text.x=element\_blank(),  
 axis.ticks.x=element\_blank())  
meanpupa0\_plot



Calculate SSD index

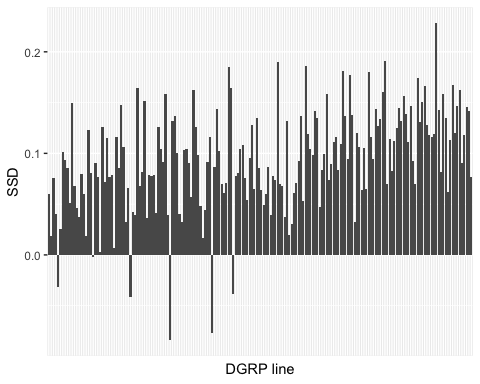
#calculate SSD0  
#separating males and females to put the values in columns  
df0\_mean\_F<-subset(df0\_mean, sex=="F")  
df0\_mean\_M<-subset(df0\_mean, sex=="M")  
  
df0\_mean\_2<-merge(x=df0\_mean\_F, y=df0\_mean\_M, by.x="line", by.y="line")  
head(df0\_mean\_2)

## line sex.x day.x pupa\_noblock.x sex.y day.y pupa\_noblock.y  
## 1 L100 F D0 0.23636559 M D0 0.09779092  
## 2 L101 F D0 0.08156526 M D0 0.06470751  
## 3 L105 F D0 -0.04813763 M D0 -0.09408385  
## 4 L109 F D0 0.04641259 M D0 -0.03249400  
## 5 L136 F D0 0.17411774 M D0 0.05794346  
## 6 L138 F D0 0.17581337 M D0 0.09222124

#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  
  
SSD0<-df0\_mean\_2 #move back to df0\_mean  
SSD0$SSD<- SSD0$pupaF - SSD0$pupaM #since we established that females are larger than males in general, SSD is female-male sizes  
head(SSD0)#182 lines

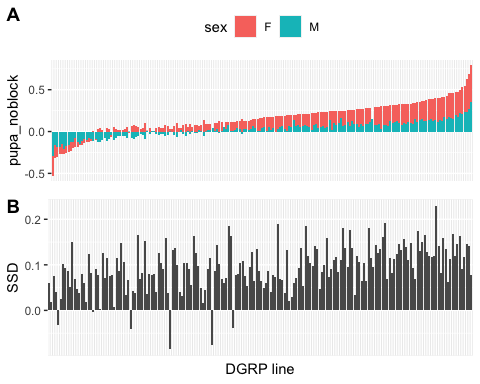
## line pupaF pupaM SSD  
## 1 L100 0.23636559 0.09779092 0.13857467  
## 2 L101 0.08156526 0.06470751 0.01685776  
## 3 L105 -0.04813763 -0.09408385 0.04594622  
## 4 L109 0.04641259 -0.03249400 0.07890658  
## 5 L136 0.17411774 0.05794346 0.11617428  
## 6 L138 0.17581337 0.09222124 0.08359213

SSD0\_plot<-ggplot(SSD0, aes(x=reorder(line,pupaF), y=SSD)) +  
 geom\_col(aes()) +  
 theme(axis.text.x=element\_blank(),  
 axis.ticks.x=element\_blank())+  
 labs(x = "DGRP line")  
  
SSD0\_plot



#### Arrange plots together

ggarrange(meanpupa0\_plot, SSD0\_plot,   
 labels = c("A", "B"),  
 ncol = 1, nrow = 2)

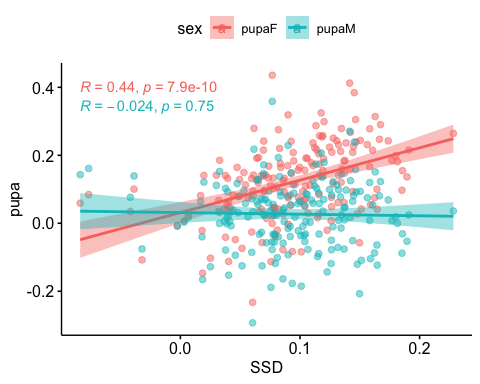


## Question 1.3: Is SSD genetic variation due to the genetic variation in male or in female size?

Correlation test and plot

#reshape data  
SSD0\_gather<-gather(SSD0,key="sex",value="pupa",pupaF:pupaM)  
#plot  
sp <- ggscatter(SSD0\_gather, x = "SSD", y = "pupa",  
 color = "sex",  
 add = "reg.line", conf.int = TRUE, alpha=0.5)  
sp + ggpubr::stat\_cor(aes(color = sex)) #stat\_cor uses the Pearson correlation

## `geom\_smooth()` using formula 'y ~ x'



In fed flies, SSD covaries with female size.

# Section 2: SSD in starved flies

We found that in fed flies, SSD is female biased and it varies genetically. We finally found out that SSD covaries with female size.

How is SSD in starved conditions?

* Section 2: SSD in starved flies - use D1
  + Question 2.1: Is SSD in starved flies the same as in fed flies? (do a lmer sex\*day on D0 and D1)
    - Does SSD increase or decrease when the flies are starved? Our hypothesis is that overall SSD should decrease
  + Question 2.2: Does the SSD in starved conditions vary the same way as in fed flies?
    - use a correlation and variance analysis

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

### 1 day starvation data preparation = OK

#fisrt subset Day1 data from df\_sub  
df1<-subset(df\_sub01, day=="D1")   
#na.omit only if pupa has NA  
df1<-df1[!is.na(df1$pupa), ] #just in case, but already removed previously  
df1<-df1%>%  
group\_by(group) %>%  
filter(n() >=10)   
length(unique(df1$line)) #174 lines left

## [1] 174

### Model fit for SSD

# Testing effect of sex in pupa size, with random effect for line and block.  
SSD1test<-lmer(pupa~sex+(1|line) +(1|block), REML=TRUE, data=df1)  
summary(SSD1test)

## 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: -7497.8  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.6775 -0.6418 0.0128 0.6666 3.3590   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## line (Intercept) 0.011182 0.10575   
## block (Intercept) 0.001546 0.03932   
## Residual 0.017728 0.13315   
## Number of obs: 6754, groups: line, 174; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.439e+01 1.581e-02 8.509e+00 910.70 <2e-16 \*\*\*  
## sexM -6.419e-02 3.403e-03 6.658e+03 -18.86 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## sexM -0.118

Anova(SSD1test)

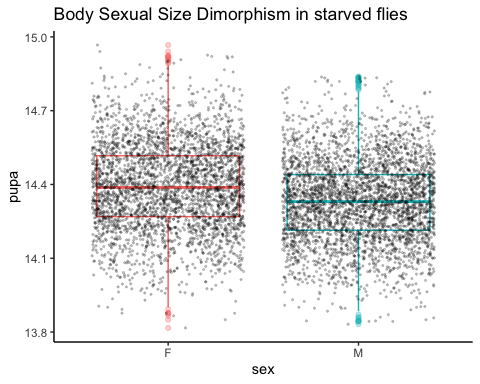
## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: pupa  
## Chisq Df Pr(>Chisq)   
## sex 355.75 1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

SSD still exists at 1 day starvation

### Plots

Plot with block effect

# original pupa size  
(SSDtest\_plot <- ggplot(df1, aes(x = sex, y = pupa, colour = sex)) +  
 geom\_boxplot(alpha=0.3) +  
 theme\_classic() +  
 geom\_jitter(color="black", size=0.4, alpha=0.2)+  
 theme(legend.position = "none",  
 panel.spacing = unit(1, "lines")) + # adding space between panels  
 ggtitle("Body Sexual Size Dimorphism in starved flies")  
 )

 Plot adjusted no block effect

#removed block effect pupal size  
(SSDtest\_plot\_noblock <- ggplot(df1, aes(x = sex, y = pupa\_noblock, colour = sex)) +  
 geom\_boxplot(alpha=0.3) +  
 theme\_classic() +  
 geom\_jitter(color="black", size=0.4, alpha=0.2)+  
   
 theme(legend.position = "none",  
 panel.spacing = unit(1, "lines")) + # adding space between panels  
 ggtitle("Body Sexual Size Dimorphism in starved flies no block effect")  
 )



HERE plot of mean size male and female for starved flies

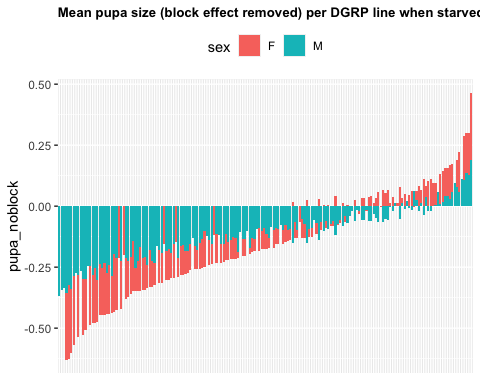
#calculate means for each group using pupa\_noblock  
df1\_mean<-aggregate(df1[, 10], list(df1$group), mean)  
head(df1\_mean,2)

## Group.1 pupa\_noblock  
## 1 L100\_F\_D1 0.04734688  
## 2 L100\_M\_D1 0.03662500

#re-add line, day and sex columns  
df1\_mean<-df1\_mean %>%  
 separate(Group.1, c("line", "sex","day"), "\_")  
head(df1\_mean,2)

## line sex day pupa\_noblock  
## 1 L100 F D1 0.04734688  
## 2 L100 M D1 0.03662500

meanpupa1\_plot<-ggplot(df1\_mean, aes(x=reorder(line,pupa\_noblock), y=pupa\_noblock)) +  
 geom\_col(aes(fill = sex)) +  
 #facet\_wrap(~ day) +  
 #ggtitle("Mean pupa size per sex for each line in fed flies, without block effect")  
theme(legend.direction = "horizontal",legend.position = "top",axis.title.x = element\_blank(),axis.text.x=element\_blank(),  
 axis.ticks.x=element\_blank(),plot.title = element\_text(size = 10, face = "bold"))  
meanpupa1\_plot+ggtitle("Mean pupa size (block effect removed) per DGRP line when starved, for each sex", )



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

Use df\_sub01

model2<-lmer(pupa~sex\*day+(sex+day|line)+(1|block), data=df\_sub01)  
model1<-lmer(pupa~sex\*day+(sex\*day|line)+(1|block), data=df\_sub01)  
anova(model1,model2)

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

## Data: df\_sub01  
## Models:  
## model2: pupa ~ sex \* day + (sex + day | line) + (1 | block)  
## model1: pupa ~ sex \* day + (sex \* day | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 12 -19489 -19397 9756.3 -19513   
## model1 16 -19494 -19372 9763.3 -19526 13.923 4 0.007544 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex \* day + (sex \* day | line) + (1 | block)  
## Data: df\_sub01  
##   
## REML criterion at convergence: -19492.7  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.2044 -0.6176 0.0517 0.6581 3.5543   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## line (Intercept) 0.0112675 0.10615   
## sexM 0.0007463 0.02732 -0.49   
## dayD1 0.0061539 0.07845 -0.28 0.32   
## sexM:dayD1 0.0009538 0.03088 0.32 -0.46 -0.52  
## block (Intercept) 0.0015922 0.03990   
## Residual 0.0151956 0.12327   
## Number of obs: 15394, groups: line, 195; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 14.566230 0.015663 10.940499 929.966 < 2e-16 \*\*\*  
## sexM -0.094440 0.003440 152.019321 -27.451 < 2e-16 \*\*\*  
## dayD1 -0.179327 0.006792 178.941747 -26.404 < 2e-16 \*\*\*  
## sexM:dayD1 0.030618 0.004900 170.094684 6.248 3.22e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexM dayD1   
## sexM -0.218   
## dayD1 -0.167 0.339   
## sexM:dayD1 0.129 -0.593 -0.515

Using summary model 1, I can calculate the means for Female 0, Male 0, Female 1 and Male 1.

# figure out how to extract intercept values  
F0 = 14.566230  
M0 = 14.566230 -0.094440   
F1 = 14.566230 -0.179327  
M1 = 14.566230 -0.094440 -0.179327 + 0.030618   
print("SSD0")

## [1] "SSD0"

F0-M0

## [1] 0.09444

print("SSD1")

## [1] "SSD1"

F1-M1

## [1] 0.063822

SSD1<SSD0 and it is significantly different

##### Calculating SSD in both conditions

#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 0.04734688 M D1 0.03662500  
## 2 L101 F D1 -0.07406721 M D1 -0.14679362  
## 3 L105 F D1 -0.15055976 M D1 -0.28742581  
## 4 L109 F D1 -0.16955959 M D1 -0.27347100  
## 5 L129 F D1 0.03332566 M D1 -0.05734916  
## 6 L136 F D1 -0.10598883 M D1 -0.15490167

#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 0.04734688 0.03662500 0.01072188  
## 2 L101 -0.07406721 -0.14679362 0.07272640  
## 3 L105 -0.15055976 -0.28742581 0.13686606  
## 4 L109 -0.16955959 -0.27347100 0.10391141  
## 5 L129 0.03332566 -0.05734916 0.09067482  
## 6 L136 -0.10598883 -0.15490167 0.04891283

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

## [1] 143

#combining SSDs values without discarding rows, which would happen if I merged the dataframes (we would end up with 59 lines for all SSD2)  
head(SSD0)

## line pupaF pupaM SSD  
## 1 L100 0.23636559 0.09779092 0.13857467  
## 2 L101 0.08156526 0.06470751 0.01685776  
## 3 L105 -0.04813763 -0.09408385 0.04594622  
## 4 L109 0.04641259 -0.03249400 0.07890658  
## 5 L136 0.17411774 0.05794346 0.11617428  
## 6 L138 0.17581337 0.09222124 0.08359213

SSD0<-SSD0[,c(1,4)]  
SSD0$day<-"D0"  
head(SSD0)

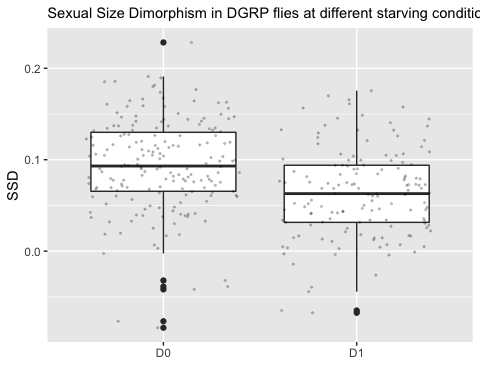
## line SSD day  
## 1 L100 0.13857467 D0  
## 2 L101 0.01685776 D0  
## 3 L105 0.04594622 D0  
## 4 L109 0.07890658 D0  
## 5 L136 0.11617428 D0  
## 6 L138 0.08359213 D0

SSD1<-df1\_mean[,c(1,4)]  
SSD1$day<-"D1"  
names(SSD1)[names(SSD1) == "SSD1"] <- "SSD"  
  
SSD\_all<-rbind(SSD0,SSD1)  
head(SSD\_all)

## line SSD day  
## 1 L100 0.13857467 D0  
## 2 L101 0.01685776 D0  
## 3 L105 0.04594622 D0  
## 4 L109 0.07890658 D0  
## 5 L136 0.11617428 D0  
## 6 L138 0.08359213 D0

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

## 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 ### Model fit

#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 4.1056 4.1056 1 156.37 233.98 < 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 -7503.9 -7469.8 3756.9 -7513.9   
## model1 7 -7517.2 -7469.5 3765.6 -7531.2 17.289 2 0.0001761 \*\*\*  
## ---  
## 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] 17.63734

nlevels(df1$line)

## [1] 0

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

## [1] 0.0001479447

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.

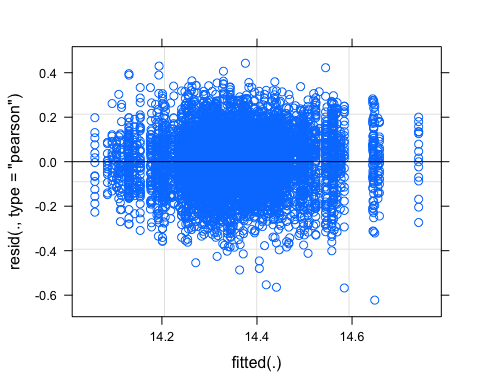
#### Finally using Bayesian Analysis

#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

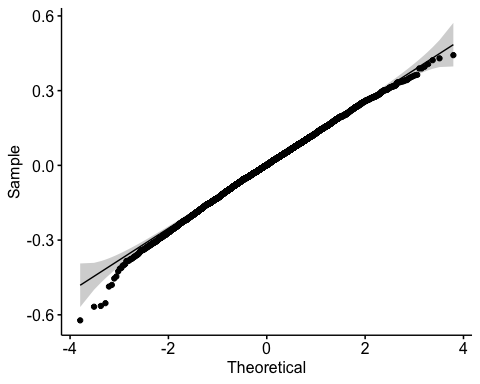
##### Model 1 residual distribution

plot(model1)



##### QQ plot

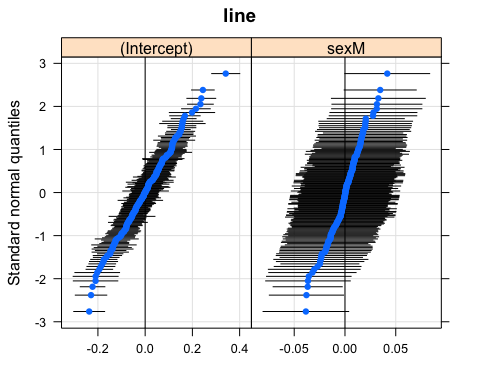
res\_model1=residuals(model1)  
ggqqplot(res\_model1)



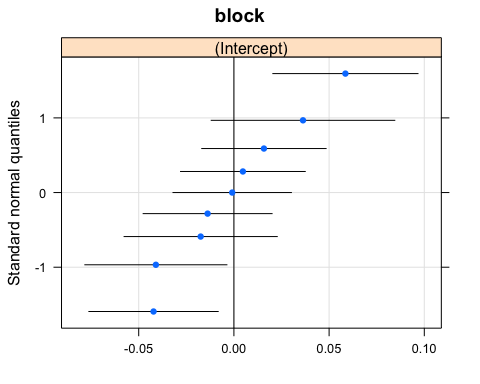
##### Random effect plot

qqmath(ranef(model1))

## $line



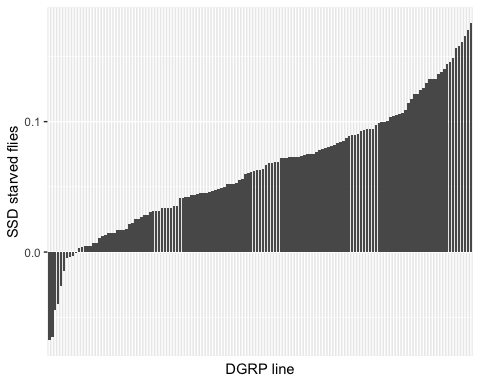
##   
## $block



SSD in starved flies varies genetically.

### Plot SSD in starved flies by line

SSD1\_plot<-ggplot(SSD1, aes(x=reorder(line,SSD), y=SSD)) +  
 geom\_col(aes()) +  
 theme(axis.text.x=element\_blank(),  
 axis.ticks.x=element\_blank())+  
 labs(x = " DGRP line ",y="SSD starved flies")  
SSD1\_plot



### Covariation with SSD in fed flies?

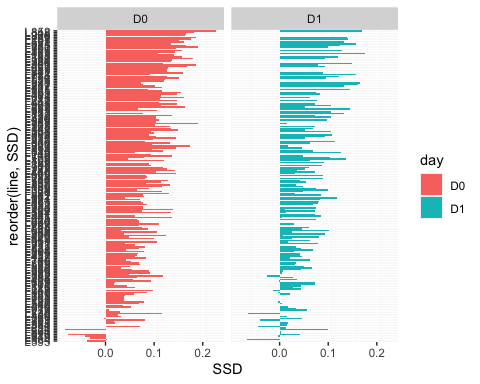
Does SSD in fed flies covary with SSD in old flies?

To know, this I am looking a the correlation between SSD values in the two food conditions, across lines. Plot of SSD values between two food conditions

head(SSD\_all,2)

## line SSD day  
## 1 L100 0.13857467 D0  
## 2 L101 0.01685776 D0

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

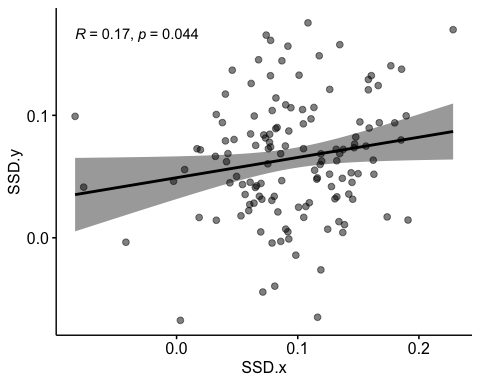
 Correlation between SSDs. Hypothesis: if SSD changes with environment, which is what we expect, we will not see a correlation between SSD0 and SSD1 and/or SSD2

#reshape data  
head(SSD\_all)

## line SSD day  
## 1 L100 0.13857467 D0  
## 2 L101 0.01685776 D0  
## 3 L105 0.04594622 D0  
## 4 L109 0.07890658 D0  
## 5 L136 0.11617428 D0  
## 6 L138 0.08359213 D0

x= SSD\_all[SSD\_all$day =="D0", ]  
y= SSD\_all[SSD\_all$day =="D1", ]  
  
SSD\_merged<-merge(x=SSD\_all[SSD\_all$day =="D0", ], y=SSD\_all[SSD\_all$day =="D1", ], by.x="line", by.y="line")   
#plot  
sp2<- ggscatter(SSD\_merged, x="SSD.x" ,y="SSD.y",  
 add = "reg.line", conf.int = TRUE, alpha=0.5)  
sp2 + ggpubr::stat\_cor(aes())#stat\_cor uses the Pearson correlation

## `geom\_smooth()` using formula 'y ~ x'

 Correlation between SSD1 and SSD0, baarely significant and rsquare very low.

It seems that SSD1 may vary differently, does female covariation with SSD0 changes in starved conditions?

Correlation SSD1 and female and male size

SSD1\_gather<-gather(df1\_mean, key="sex",value="pupa",pupaF1:pupaM1)  
head(df1\_mean)

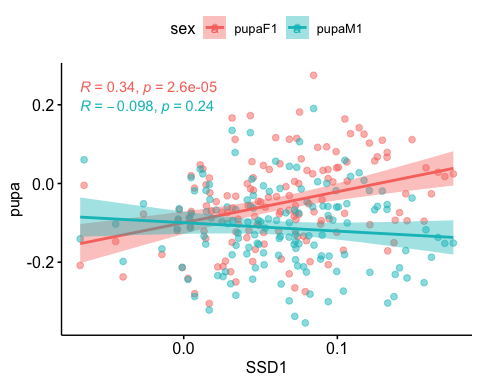
## line pupaF1 pupaM1 SSD1  
## 1 L100 0.04734688 0.03662500 0.01072188  
## 2 L101 -0.07406721 -0.14679362 0.07272640  
## 3 L105 -0.15055976 -0.28742581 0.13686606  
## 4 L109 -0.16955959 -0.27347100 0.10391141  
## 5 L129 0.03332566 -0.05734916 0.09067482  
## 6 L136 -0.10598883 -0.15490167 0.04891283

head(SSD1\_gather)

## line SSD1 sex pupa  
## 1 L100 0.01072188 pupaF1 0.04734688  
## 2 L101 0.07272640 pupaF1 -0.07406721  
## 3 L105 0.13686606 pupaF1 -0.15055976  
## 4 L109 0.10391141 pupaF1 -0.16955959  
## 5 L129 0.09067482 pupaF1 0.03332566  
## 6 L136 0.04891283 pupaF1 -0.10598883

sp3<- ggscatter(SSD1\_gather, x="SSD1" ,y="pupa",  
 color = "sex",  
 add = "reg.line", conf.int = TRUE, alpha=0.5)  
sp3 + ggpubr::stat\_cor(aes(color=sex))#stat\_cor uses the Pearson correlation

## `geom\_smooth()` using formula 'y ~ x'

 SSD1 also covaries with female size in starved conditions

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

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

SSD changes in different food conditions, that means that there is a specific plasticity between sexes, or Sex Specific Plasticity. How does it look like and which sex is responsible for this difference in plasticity that is linked to SSD?

#Section 3: Origin of SSP - Question 3.1: Confirming that SSP exists. - 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? ### SSP: between fed and 1 day starved flies

We showed in Question 2.X that SSD changed when changing food conditions. Let’s view again the summary of the Model.

model1<-lmer(pupa~sex\*day+(sex\*day|line)+(1|block), data=df\_sub01)  
summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: pupa ~ sex \* day + (sex \* day | line) + (1 | block)  
## Data: df\_sub01  
##   
## REML criterion at convergence: -19492.7  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.2044 -0.6176 0.0517 0.6581 3.5543   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## line (Intercept) 0.0112675 0.10615   
## sexM 0.0007463 0.02732 -0.49   
## dayD1 0.0061539 0.07845 -0.28 0.32   
## sexM:dayD1 0.0009538 0.03088 0.32 -0.46 -0.52  
## block (Intercept) 0.0015922 0.03990   
## Residual 0.0151956 0.12327   
## Number of obs: 15394, groups: line, 195; block, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 14.566230 0.015663 10.940499 929.966 < 2e-16 \*\*\*  
## sexM -0.094440 0.003440 152.019321 -27.451 < 2e-16 \*\*\*  
## dayD1 -0.179327 0.006792 178.941747 -26.404 < 2e-16 \*\*\*  
## sexM:dayD1 0.030618 0.004900 170.094684 6.248 3.22e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexM dayD1   
## sexM -0.218   
## dayD1 -0.167 0.339   
## sexM:dayD1 0.129 -0.593 -0.515

##Question 3.2: Does this SSP vary in the DGRP flies? ### Model fit comparison - previously done to compare SSD0 and SSD1

# 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+(sex+day|line)+(1|block), data=df\_sub01) #there is randomness of sex and also the effect of day  
model1<-lmer(pupa~sex\*day+(sex\*day|line)+(1|block), data=df\_sub01) #basic model  
anova(model1,model2)

## Data: df\_sub01  
## Models:  
## model2: pupa ~ sex \* day + (sex + day | line) + (1 | block)  
## model1: pupa ~ sex \* day + (sex \* day | line) + (1 | block)  
## npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)   
## model2 12 -19489 -19397 9756.3 -19513   
## model1 16 -19494 -19372 9763.3 -19526 13.923 4 0.007544 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Anova(model1) #there is SSP

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: pupa  
## Chisq Df Pr(>Chisq)   
## sex 869.840 1 < 2.2e-16 \*\*\*  
## day 731.895 1 < 2.2e-16 \*\*\*  
## sex:day 39.044 1 4.145e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 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] 12

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] 14.1722

nlevels(df\_sub01$line)

## [1] 0

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

## [1] 0.01455181

pchisq(q = LR.model, df=nlevels(df\_sub01$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)

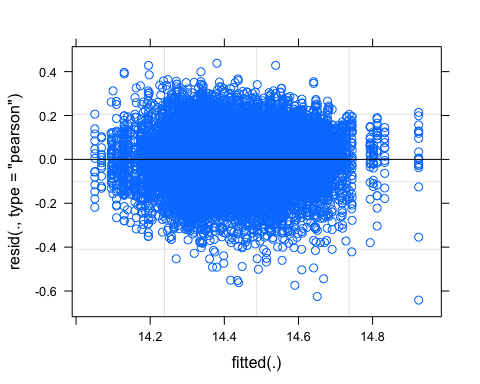
#### Post model 1 fitting check

#### Residual distribution

res\_model1=residuals(model1)

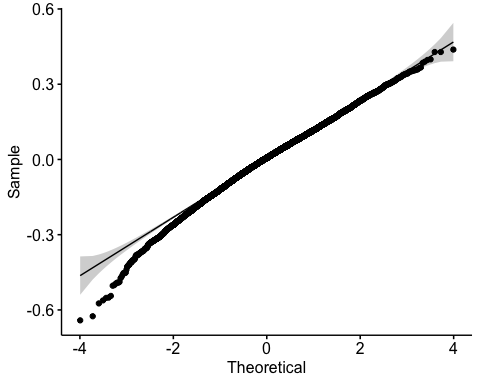
#### Model 1 residual distribution

plot(model1)



#### QQ plot

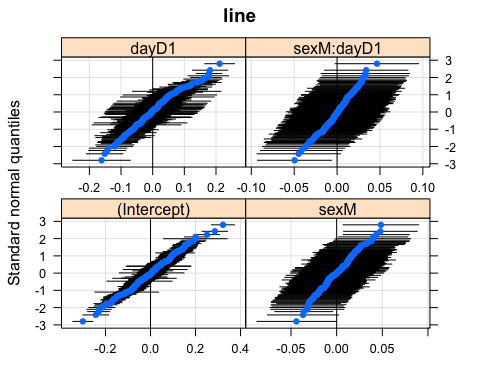
require(ggpubr)  
ggqqplot(res\_model1)



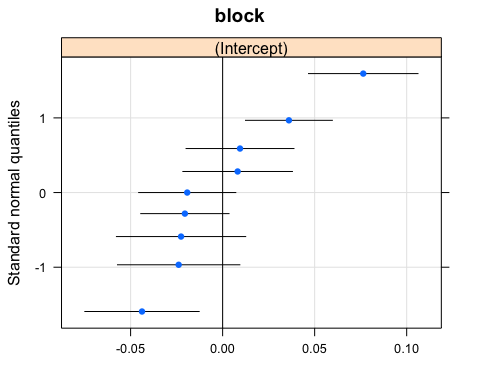
#### Random effect plot

qqmath(ranef(model1))

## $line



##   
## $block



Answer 3.2: There is genetic variation in SSP if we compare fed flies and starved flies.

### Plot of SSP per line

#### Calculate SSP

To calculate SSP, I first calculate the plasticity of female and male and subtract the plasticity female-male

#calculate plasticity female  
  
df1F<-subset(df1, sex=="F")  
df1F\_mean<-aggregate(df1F[, 10], list(df1F$line), mean) #using pupa\_noblock  
colnames(df1F\_mean)<-c("line","pupaFmean\_1")  
  
df0F<-subset(df0, sex=="F")  
df0F\_mean<-aggregate(df0F[, 10], list(df0F$line), mean)  
colnames(df0F\_mean)<-c("line","pupaFmean\_0")  
  
pupa\_mean\_F01<-merge(x=df0F\_mean, y=df1F\_mean, by.x="line", by.y="line")   
head(pupa\_mean\_F01,2) #140 lines

## line pupaFmean\_0 pupaFmean\_1  
## 1 L100 0.23636559 0.04734688  
## 2 L101 0.08156526 -0.07406721

pupa\_mean\_F01$PF01<-pupa\_mean\_F01$pupaFmean\_0-pupa\_mean\_F01$pupaFmean\_1 #calculating plast female 01  
plastF01<-pupa\_mean\_F01

#calculate plasticity male  
head(df1)

## # A tibble: 6 x 10  
## # Groups: group [1]  
## id line block day sex wing leg pupa group pupa\_noblock  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl>  
## 1 021-A-1-F-024 L21 B6 D1 F 14.1 13.8 14.6 L21\_F\_D1 0.219   
## 2 021-A-1-F-026 L21 B6 D1 F 14.3 14.0 14.8 L21\_F\_D1 0.371   
## 3 021-A-1-F-029 L21 B6 D1 F 14.1 13.7 14.4 L21\_F\_D1 0.0103   
## 4 021-A-1-F-031 L21 B6 D1 F 14.1 NA 14.6 L21\_F\_D1 0.169   
## 5 021-A-1-F-034 L21 B6 D1 F 13.8 13.4 14.1 L21\_F\_D1 -0.299   
## 6 021-A-1-F-035 L21 B6 D1 F 14.0 13.6 14.4 L21\_F\_D1 -0.00168

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 0.03662500  
## 2 L101 -0.14679362  
## 3 L105 -0.28742581  
## 4 L109 -0.27347100  
## 5 L129 -0.05734916  
## 6 L136 -0.15490167

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 0.09779092  
## 2 L101 0.06470751  
## 3 L105 -0.09408385  
## 4 L109 -0.03249400  
## 5 L136 0.05794346  
## 6 L138 0.09222124

pupa\_mean\_M01<-merge(x=df0M\_mean, y=df1M\_mean, by.x="line", by.y="line")  
head(pupa\_mean\_F01) #151

## line pupaFmean\_0 pupaFmean\_1 PF01  
## 1 L100 0.23636559 0.04734688 0.1890187  
## 2 L101 0.08156526 -0.07406721 0.1556325  
## 3 L105 -0.04813763 -0.15055976 0.1024221  
## 4 L109 0.04641259 -0.16955959 0.2159722  
## 5 L136 0.17411774 -0.10598883 0.2801066  
## 6 L138 0.17581337 -0.04629073 0.2221041

pupa\_mean\_M01$PM01<-pupa\_mean\_M01$pupaMmean\_0-pupa\_mean\_M01$pupaMmean\_1 #calculating SSP01  
plastM01<-pupa\_mean\_M01

#SSP  
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 0.23636559 0.04734688 0.1890187 0.09779092 0.03662500 0.06116592  
## 2 L101 0.08156526 -0.07406721 0.1556325 0.06470751 -0.14679362 0.21150112  
## 3 L105 -0.04813763 -0.15055976 0.1024221 -0.09408385 -0.28742581 0.19334197  
## 4 L109 0.04641259 -0.16955959 0.2159722 -0.03249400 -0.27347100 0.24097700  
## 5 L136 0.17411774 -0.10598883 0.2801066 0.05794346 -0.15490167 0.21284513  
## 6 L138 0.17581337 -0.04629073 0.2221041 0.09222124 -0.06743099 0.15965223

length(unique(pupa\_SSP01$line)) #133 lines left

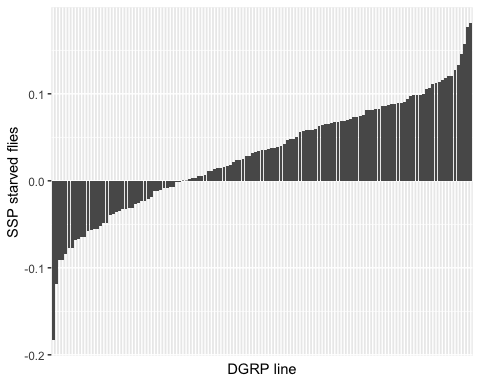
## [1] 133

#calculate SSP from the plasticity.  
pupa\_SSP01$SSP01<-pupa\_SSP01$PF01-pupa\_SSP01$PM01  
  
#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  
head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.23636559 0.04734688 0.1890187 0.09779092 0.03662500 0.06116592  
## 2 L101 0.08156526 -0.07406721 0.1556325 0.06470751 -0.14679362 0.21150112  
## 3 L105 -0.04813763 -0.15055976 0.1024221 -0.09408385 -0.28742581 0.19334197  
## 4 L109 0.04641259 -0.16955959 0.2159722 -0.03249400 -0.27347100 0.24097700  
## 5 L136 0.17411774 -0.10598883 0.2801066 0.05794346 -0.15490167 0.21284513  
## 6 L138 0.17581337 -0.04629073 0.2221041 0.09222124 -0.06743099 0.15965223  
## SSP01 SSD0 SSD1  
## 1 0.12785279 0.13857467 0.01072188  
## 2 -0.05586865 0.01685776 0.07272640  
## 3 -0.09091984 0.04594622 0.13686606  
## 4 -0.02500483 0.07890658 0.10391141  
## 5 0.06726144 0.11617428 0.04891283  
## 6 0.06245187 0.08359213 0.02114025

#### Plot

SSP01\_plot<-ggplot(pupa\_SSP01, aes(x=reorder(line,SSP01), y=SSP01)) +  
 geom\_col(aes()) +  
 theme(axis.text.x=element\_blank(),  
 axis.ticks.x=element\_blank())+  
 labs(x = " DGRP line ",y="SSP starved flies")  
SSP01\_plot



We saw that SSD variation in both food conditions is correlated with female size variation. Do we then have SSP variation due to female size variation or possibly plasticity in one of the two sexes?

## Question 3.3: Is SSP genetic variation correlated with female size in either fed or starved conditions?

#### Correlation between SSP and female size in both conditions

Prepare a dataset with line, SSP, fed female size and starved female size

head(df1\_mean\_F)

## line sex day pupa\_noblock  
## 1 L100 F D1 0.04734688  
## 3 L101 F D1 -0.07406721  
## 5 L105 F D1 -0.15055976  
## 7 L109 F D1 -0.16955959  
## 9 L129 F D1 0.03332566  
## 11 L136 F D1 -0.10598883

head(df0\_mean\_F)

## line sex day pupa\_noblock  
## 1 L100 F D0 0.23636559  
## 3 L101 F D0 0.08156526  
## 5 L105 F D0 -0.04813763  
## 7 L109 F D0 0.04641259  
## 9 L136 F D0 0.17411774  
## 11 L138 F D0 0.17581337

head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.23636559 0.04734688 0.1890187 0.09779092 0.03662500 0.06116592  
## 2 L101 0.08156526 -0.07406721 0.1556325 0.06470751 -0.14679362 0.21150112  
## 3 L105 -0.04813763 -0.15055976 0.1024221 -0.09408385 -0.28742581 0.19334197  
## 4 L109 0.04641259 -0.16955959 0.2159722 -0.03249400 -0.27347100 0.24097700  
## 5 L136 0.17411774 -0.10598883 0.2801066 0.05794346 -0.15490167 0.21284513  
## 6 L138 0.17581337 -0.04629073 0.2221041 0.09222124 -0.06743099 0.15965223  
## SSP01 SSD0 SSD1  
## 1 0.12785279 0.13857467 0.01072188  
## 2 -0.05586865 0.01685776 0.07272640  
## 3 -0.09091984 0.04594622 0.13686606  
## 4 -0.02500483 0.07890658 0.10391141  
## 5 0.06726144 0.11617428 0.04891283  
## 6 0.06245187 0.08359213 0.02114025

merge1<-merge(x=df1\_mean\_F,y=df0\_mean\_F,by.x="line",by.y="line")  
SSP\_female01<-merge(x=merge1,y=pupa\_SSP01,by.x="line",by.y="line")  
head(SSP\_female01)

## line sex.x day.x pupa\_noblock.x sex.y day.y pupa\_noblock.y pupaFmean\_0  
## 1 L100 F D1 0.04734688 F D0 0.23636559 0.23636559  
## 2 L101 F D1 -0.07406721 F D0 0.08156526 0.08156526  
## 3 L105 F D1 -0.15055976 F D0 -0.04813763 -0.04813763  
## 4 L109 F D1 -0.16955959 F D0 0.04641259 0.04641259  
## 5 L136 F D1 -0.10598883 F D0 0.17411774 0.17411774  
## 6 L138 F D1 -0.04629073 F D0 0.17581337 0.17581337  
## pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01 SSP01  
## 1 0.04734688 0.1890187 0.09779092 0.03662500 0.06116592 0.12785279  
## 2 -0.07406721 0.1556325 0.06470751 -0.14679362 0.21150112 -0.05586865  
## 3 -0.15055976 0.1024221 -0.09408385 -0.28742581 0.19334197 -0.09091984  
## 4 -0.16955959 0.2159722 -0.03249400 -0.27347100 0.24097700 -0.02500483  
## 5 -0.10598883 0.2801066 0.05794346 -0.15490167 0.21284513 0.06726144  
## 6 -0.04629073 0.2221041 0.09222124 -0.06743099 0.15965223 0.06245187  
## SSD0 SSD1  
## 1 0.13857467 0.01072188  
## 2 0.01685776 0.07272640  
## 3 0.04594622 0.13686606  
## 4 0.07890658 0.10391141  
## 5 0.11617428 0.04891283  
## 6 0.08359213 0.02114025

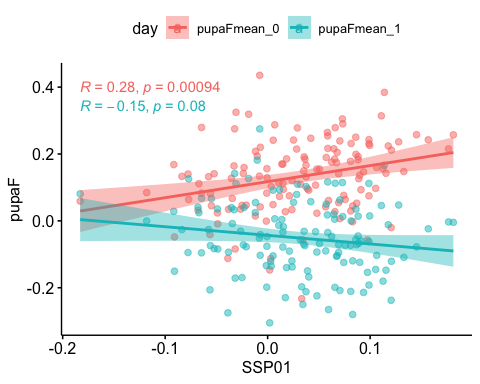
SSP\_female01<-SSP\_female01[,c(1,8,9,14)]  
SSP\_female01

## line pupaFmean\_0 pupaFmean\_1 SSP01  
## 1 L100 0.2363655921 0.0473468776 0.1278527943  
## 2 L101 0.0815652643 -0.0740672146 -0.0558686452  
## 3 L105 -0.0481376305 -0.1505597578 -0.0909198410  
## 4 L109 0.0464125851 -0.1695595913 -0.0250048263  
## 5 L136 0.1741177402 -0.1059888347 0.0672614443  
## 6 L138 0.1758133673 -0.0462907336 0.0624518735  
## 7 L142 0.2788247161 0.1124320287 0.0659022309  
## 8 L149 0.2794139668 0.0679115568 -0.0644893290  
## 9 L176 0.0828895412 -0.1982529409 -0.0008910385  
## 10 L177 0.0763457344 -0.1244081884 0.0742911684  
## 11 L181 0.1113136125 0.0242835147 -0.0670982610  
## 12 L189 0.1314918418 -0.2119017336 0.0373946669  
## 13 L208 0.1021898679 -0.1353714469 0.0034327735  
## 14 L21 0.0591409097 0.0808869981 -0.1829300339  
## 15 L217 0.1330860433 -0.0261915862 -0.0836394967  
## 16 L227 0.2901414923 0.0127545094 0.0733151294  
## 17 L228 0.0608474691 -0.2653055214 0.0594349230  
## 18 L235 -0.1469117117 -0.3050290200 0.0020211166  
## 19 L237 0.1846093452 -0.0739131532 0.0363380568  
## 20 L239 -0.0089159393 -0.0842144573 -0.0486717965  
## 21 L256 0.1422637606 -0.1475450716 0.1154715094  
## 22 L26 0.2046802310 -0.1040041960 0.0677984366  
## 23 L28 0.1934488635 0.1816477711 -0.0101456889  
## 24 L303 0.1696067568 -0.0783018399 0.0148094652  
## 25 L306 0.3248422400 0.1115368697 -0.0309038369  
## 26 L307 0.2166794450 -0.0348015091 0.0587374068  
## 27 L309 0.2541790583 0.1164395614 0.0586236607  
## 28 L31 0.2257913060 0.0147863211 0.0995469125  
## 29 L313 0.0527884581 -0.0710717813 0.0869420534  
## 30 L315 0.2780299108 0.1099232491 0.0290328066  
## 31 L317 -0.0840519095 -0.1711664889 -0.0316905628  
## 32 L318 0.1550998845 -0.0678521247 0.0505956574  
## 33 L319 0.1699361095 0.0040459332 0.0138640741  
## 34 L32 0.0930740780 -0.0602446815 0.0910124759  
## 35 L320 0.1560828371 -0.0758596639 -0.0007259109  
## 36 L324 0.2115433722 -0.1208765527 0.1207626974  
## 37 L335 -0.0574342086 -0.1540911299 0.0974476182  
## 38 L338 0.1274572772 0.0396371437 -0.0353482572  
## 39 L340 0.2154813374 0.0196831750 0.0646397914  
## 40 L350 0.1364722330 -0.1239943144 0.0898896950  
## 41 L352 0.0273926840 -0.2295924842 0.0172499049  
## 42 L354 0.0731422634 0.0933248888 0.0115755201  
## 43 L357 0.1562696750 -0.1803805964 0.1126090927  
## 44 L358 0.0297146277 -0.2109861978 0.0811035687  
## 45 L360 0.1853729332 -0.1397210812 0.0184610056  
## 46 L362 0.1663762274 0.1028552793 0.0374149924  
## 47 L365 0.1475061980 -0.1090329667 0.1328633332  
## 48 L367 0.2079577780 0.0548463973 -0.0118552595  
## 49 L371 0.1596627981 0.0399780043 -0.0230112811  
## 50 L373 0.0749443096 -0.1919744114 0.0988716162  
## 51 L374 0.0146461055 0.0246370097 -0.0120422051  
## 52 L375 0.0335037106 -0.1199013363 0.0236379945  
## 53 L379 0.1408615543 -0.0830239146 -0.0521969080  
## 54 L380 0.1682918130 0.0285698523 -0.0915736530  
## 55 L383 0.2424640338 -0.0236574734 0.1567484668  
## 56 L386 0.1129364283 -0.0135708607 0.0108931560  
## 57 L391 0.1918692716 0.0140621159 0.0882765013  
## 58 L392 0.2576810075 -0.0776938159 0.1452745859  
## 59 L397 0.0641479503 0.1254143180 -0.0770750535  
## 60 L399 0.2201599408 -0.0412588480 -0.0071044831  
## 61 L405 0.0748907420 -0.0904731938 0.0211830705  
## 62 L406 0.2241024617 -0.0116287497 0.1177057368  
## 63 L427 0.0942656440 -0.0173276538 0.0153584897  
## 64 L437 0.1201569398 -0.1377788422 0.0238805800  
## 65 L440 0.1092393634 -0.0567382314 0.0319390266  
## 66 L441 0.1417398575 -0.0718712342 -0.0237165261  
## 67 L461 0.0971777060 -0.1453528465 0.1054129423  
## 68 L491 -0.0007700785 -0.1337019078 -0.0013100097  
## 69 L492 0.0274898194 -0.1988097498 0.0656245067  
## 70 L502 0.1408244894 -0.0714930598 0.0693170279  
## 71 L505 0.1260761921 -0.0669311325 0.0889042001  
## 72 L508 0.0348953070 -0.0307148830 -0.0263689111  
## 73 L509 0.3842939852 0.1665144960 0.1138743292  
## 74 L513 -0.2326105929 -0.1973014308 0.0332041170  
## 75 L530 0.1990326530 0.0354862026 0.0861552722  
## 76 L535 0.1320993535 -0.0260328746 -0.0576288908  
## 77 L551 0.2147583711 0.0789377133 0.0283039925  
## 78 L555 0.2314263122 -0.0001527161 0.0814284740  
## 79 L59 0.0918663722 -0.1676834450 0.0401181596  
## 80 L595 -0.0025975567 -0.2800337679 0.0829707919  
## 81 L596 0.0422140117 -0.0898969205 0.0566650994  
## 82 L627 0.0845980292 -0.0004378710 -0.1180391746  
## 83 L639 0.1464313412 -0.2139444489 0.0935961703  
## 84 L646 0.0320321532 0.0411129805 -0.0679430805  
## 85 L69 0.1546964465 -0.0933623225 0.0480424188  
## 86 L703 -0.0200607693 -0.2375398589 0.1205299810  
## 87 L705 0.2412686552 0.0917426851 0.0054309435  
## 88 L714 0.2502704250 -0.0559475048 0.0859834540  
## 89 L716 0.1493816346 -0.0695283755 0.0351695850  
## 90 L721 0.0546097855 -0.0942183571 0.0693023701  
## 91 L727 0.1779416121 -0.0733357249 0.0809467360  
## 92 L73 0.0237411310 0.0357352055 -0.0490329133  
## 93 L730 0.2573488620 -0.0046010000 0.1812220834  
## 94 L732 0.1848079400 0.0050458830 0.0653718645  
## 95 L737 0.2390990337 0.0626137870 0.0704494220  
## 96 L748 0.1595811084 -0.1031895645 0.1064438549  
## 97 L75 0.0961651719 -0.1495231450 0.0389014292  
## 98 L757 0.2153271480 -0.0031793307 0.1765053543  
## 99 L774 0.0486131256 -0.1046407313 0.0052357487  
## 100 L776 0.1826383995 -0.0451714711 0.0882290457  
## 101 L783 0.1914249194 -0.0312691916 0.0355861385  
## 102 L786 0.0418572815 -0.1030638313 -0.0072597778  
## 103 L787 0.0446667724 -0.1195024232 0.0830378226  
## 104 L790 0.0469805318 -0.0774350882 -0.0209499157  
## 105 L796 -0.0410976144 -0.2063130010 -0.0562716480  
## 106 L802 0.3189421900 0.1266672205 -0.0185472824  
## 107 L805 0.1118217024 -0.0332620260 0.0148699278  
## 108 L808 0.1898392513 0.0501041652 0.0571413226  
## 109 L810 0.2752462464 0.1416586119 -0.0322121180  
## 110 L812 0.4354588179 0.2750758734 -0.0077282250  
## 111 L818 0.0829175977 -0.0956458510 -0.0645909293  
## 112 L819 -0.0568381621 -0.2501603575 0.0343942207  
## 113 L821 0.1068982859 -0.0324285716 0.0467564850  
## 114 L822 0.1643850264 0.0332886534 -0.0328318357  
## 115 L832 0.2416905255 0.0554649961 0.0252681528  
## 116 L837 0.0451604635 0.0665439719 -0.0548852063  
## 117 L843 0.0639962193 -0.0729432201 0.0756999382  
## 118 L849 0.0347057927 -0.0937486428 -0.0382099285  
## 119 L85 0.2470770823 0.0574073974 0.0420898092  
## 120 L850 0.0671225626 -0.0080634205 -0.0345030181  
## 121 L852 0.2643669877 0.0177730034 0.0582991717  
## 122 L853 0.1338193839 -0.0515356543 -0.0080546702  
## 123 L855 0.0229263384 -0.2242715355 0.0480717268  
## 124 L857 0.1400547084 -0.0947375898 -0.0777077663  
## 125 L859 0.3141106644 0.0459769278 0.1109516146  
## 126 L884 0.0066431900 -0.2082734123 0.0705285253  
## 127 L887 -0.1128853749 -0.1210407736 0.0029949546  
## 128 L890 0.2308915430 0.1121200701 0.0988193193  
## 129 L892 0.1288729739 -0.0469099289 -0.0005834532  
## 130 L907 0.3050137834 0.1146349453 0.0729415747  
## 131 L91 0.2428560866 -0.0621108285 0.0988473051  
## 132 L911 -0.1123190879 -0.2750342057 -0.0388221692  
## 133 L93 0.2872816575 0.0061588744 0.0069246390

##### Plot

#reshape data  
SSP\_female01<-gather(SSP\_female01,key="day",value="pupaF",pupaFmean\_0:pupaFmean\_1)  
#plot  
sp4 <- ggscatter(SSP\_female01, x = "SSP01", y = "pupaF",  
 color = "day",  
 add = "reg.line", conf.int = TRUE, alpha=0.5)  
sp4 + ggpubr::stat\_cor(aes(color = day)) #stat\_cor uses the Pearson correlation

## `geom\_smooth()` using formula 'y ~ x'



## Question 3.4: Is SSP genetic variation correlated male or female plasticity?

#### Correlation between SSP and plasticity

Prepare a dataset with line, SSP, fed female and male plasticity

head(pupa\_SSP01)

## line pupaFmean\_0 pupaFmean\_1 PF01 pupaMmean\_0 pupaMmean\_1 PM01  
## 1 L100 0.23636559 0.04734688 0.1890187 0.09779092 0.03662500 0.06116592  
## 2 L101 0.08156526 -0.07406721 0.1556325 0.06470751 -0.14679362 0.21150112  
## 3 L105 -0.04813763 -0.15055976 0.1024221 -0.09408385 -0.28742581 0.19334197  
## 4 L109 0.04641259 -0.16955959 0.2159722 -0.03249400 -0.27347100 0.24097700  
## 5 L136 0.17411774 -0.10598883 0.2801066 0.05794346 -0.15490167 0.21284513  
## 6 L138 0.17581337 -0.04629073 0.2221041 0.09222124 -0.06743099 0.15965223  
## SSP01 SSD0 SSD1  
## 1 0.12785279 0.13857467 0.01072188  
## 2 -0.05586865 0.01685776 0.07272640  
## 3 -0.09091984 0.04594622 0.13686606  
## 4 -0.02500483 0.07890658 0.10391141  
## 5 0.06726144 0.11617428 0.04891283  
## 6 0.06245187 0.08359213 0.02114025

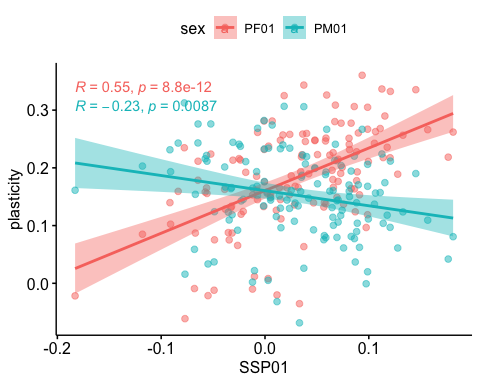
SSP\_plast<-pupa\_SSP01[,c(1,4,7,8)]  
head(SSP\_plast)

## line PF01 PM01 SSP01  
## 1 L100 0.1890187 0.06116592 0.12785279  
## 2 L101 0.1556325 0.21150112 -0.05586865  
## 3 L105 0.1024221 0.19334197 -0.09091984  
## 4 L109 0.2159722 0.24097700 -0.02500483  
## 5 L136 0.2801066 0.21284513 0.06726144  
## 6 L138 0.2221041 0.15965223 0.06245187

##### Plot

#reshape data  
SSP\_plast<-gather(SSP\_plast,key="sex",value="plasticity",PF01:PM01)  
#plot  
sp5 <- ggscatter(SSP\_plast, x = "SSP01", y = "plasticity",  
 color = "sex",  
 add = "reg.line", conf.int = TRUE, alpha=0.5)  
sp5 + ggpubr::stat\_cor(aes(color = sex)) #stat\_cor uses the Pearson correlation

## `geom\_smooth()` using formula 'y ~ x'

 There is a correlation between SSP and female plasticity.