Sperm size figures

jeremy

Wednesday, January 14, 2015

Sperm size figures

This is the output of an R markdown script - a fancy way of embedding r in a page so we can rerun/tune it as needed and then have the commands

I'm embedding, but hiding a bynch of stuff for readability - the actual code is in the .Rmd file.

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
## Warning: package 'survival' was built under R version 3.1.3
```

```
## Warning: package 'psych' was built under R version 3.1.3
```

Here's the tree from Kiontke et al 2011, figure 3 intial values I recieved had angaria and sp12 flipped - I've corrected it here Here's the tree:
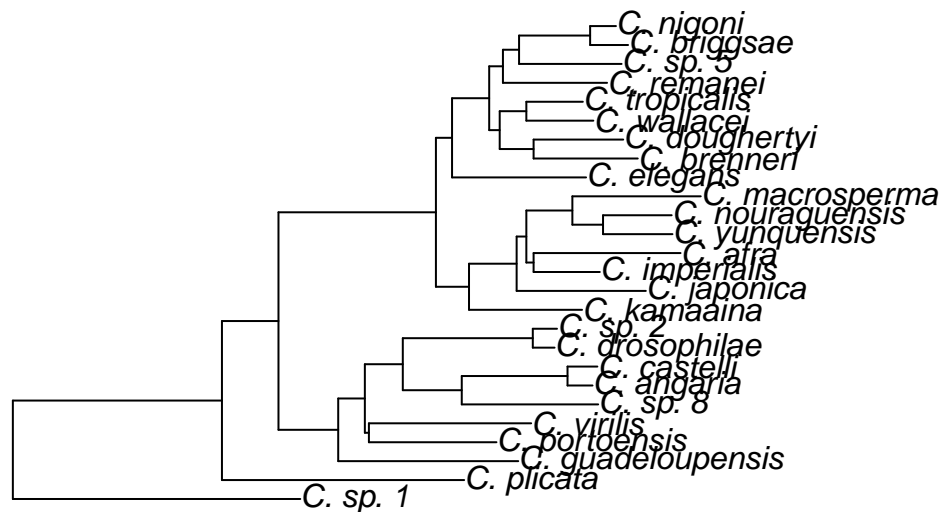
Now read in the compiled data - we have oocytesize, spermsize, primaryspermatocytesize, and oocytearea. The species names in the files are inconsistent - they are fixed now. For now, the oocyte area is using the

area formulation as made by the initial sperm size formula. All response variables are logged unless otherwise noted.

I have filtered out some of the data: Any oocytesize that is not stage -1, and any primaryspermatocyte data that is from hermaphrodites rather than males.

We now have clean data, clustered into species, strains or individuals. Variables ending in 1 are in species, 2 are strains and 3 are individuals. The non-clustered data is as is.
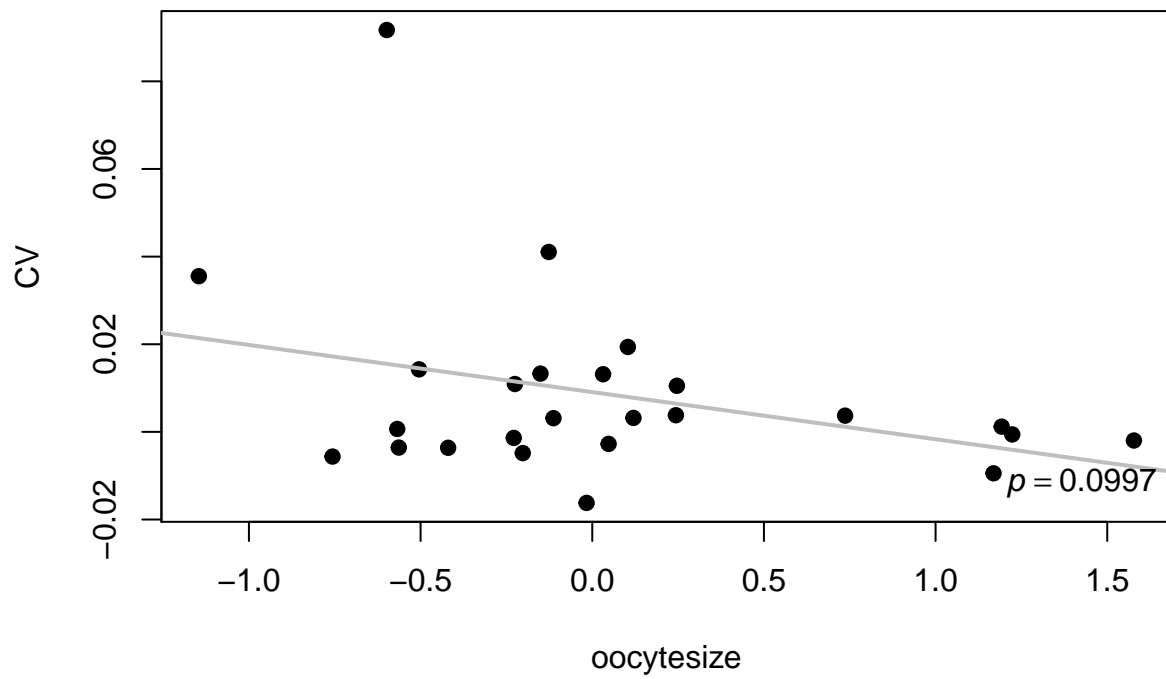
PIC plots - mean vs CV

Plotting the phylogenetic independent contrasts (PIC) of means of traits vs CV of traits. All means and CVs are grouped at strain level.

PIC of spermsize means vs CV

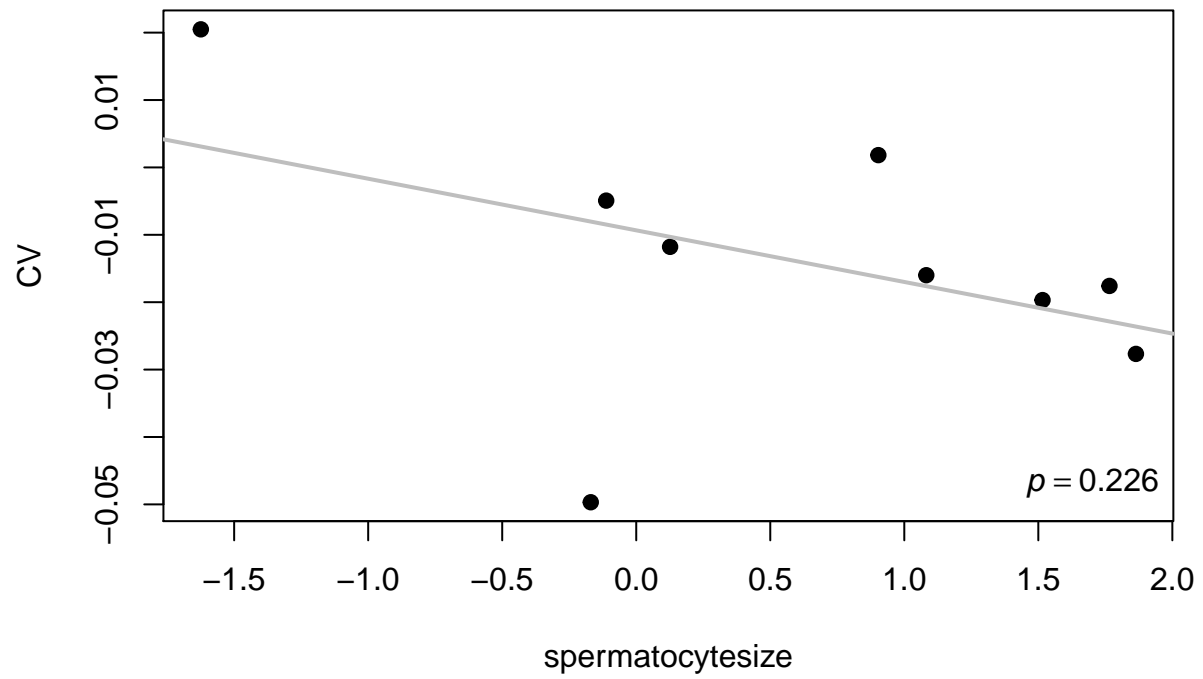
$p = 0.0599$
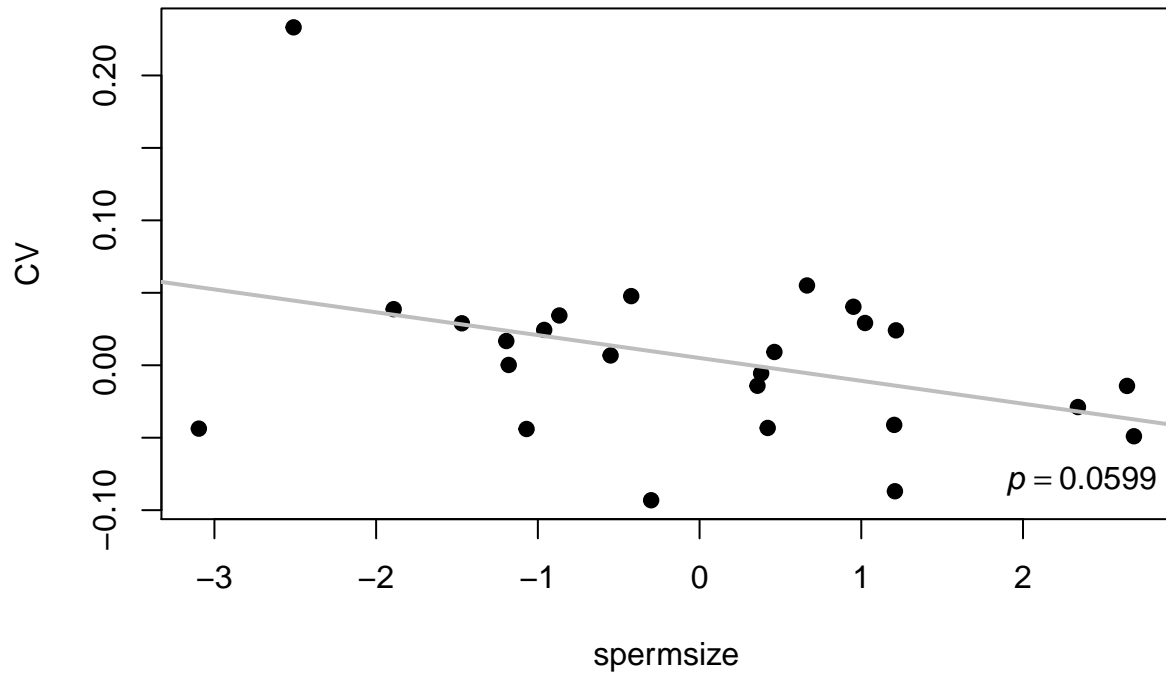
PIC of oocytesize means vs CV

$p = 0.0997$

PIC of spermatocytesize means vs CV

CV

spermatocytesize

$p = 0.226$

PIC of oocytesizearea means vs CV

$p = 0.331$
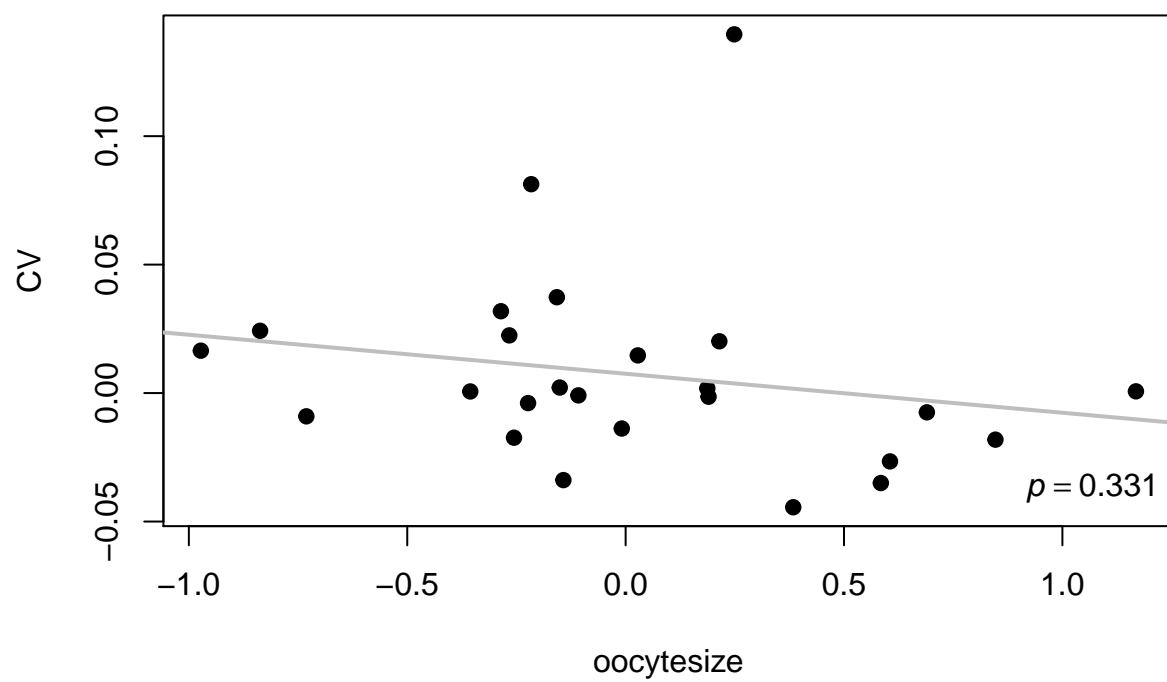
PIC cv bodyvolume male

CV

bodyvolume

$p = 0.782$
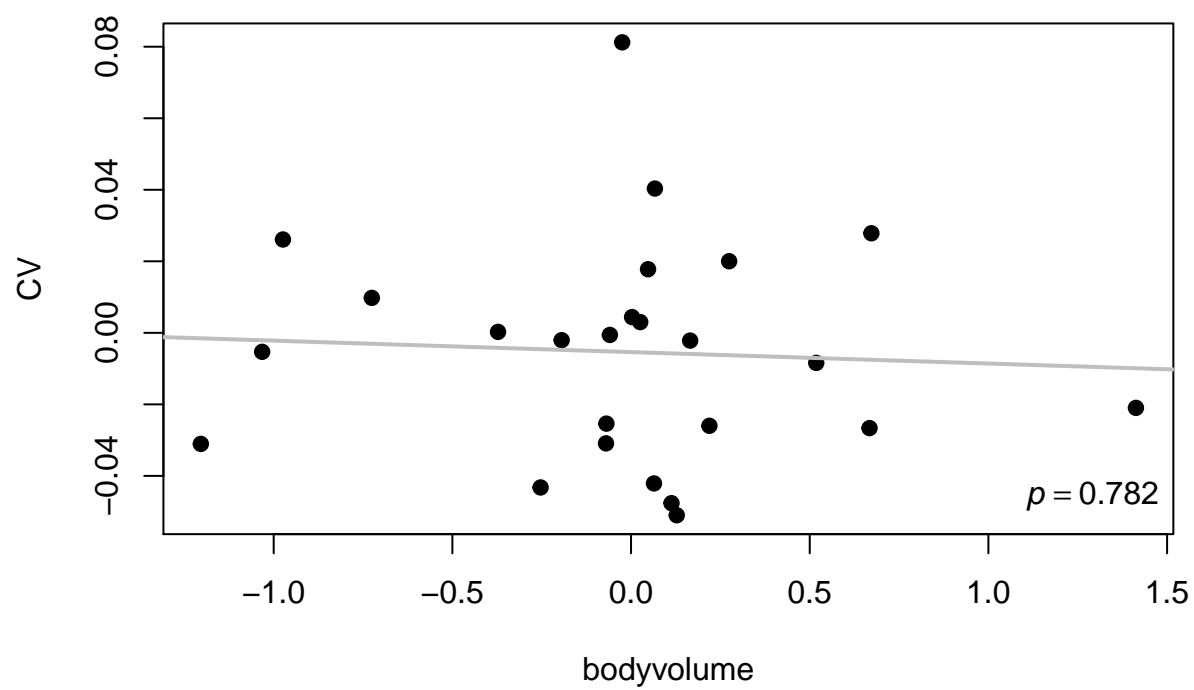
PIC cv bodyvolume female

CV

bodyvolume

$p = 0.126$
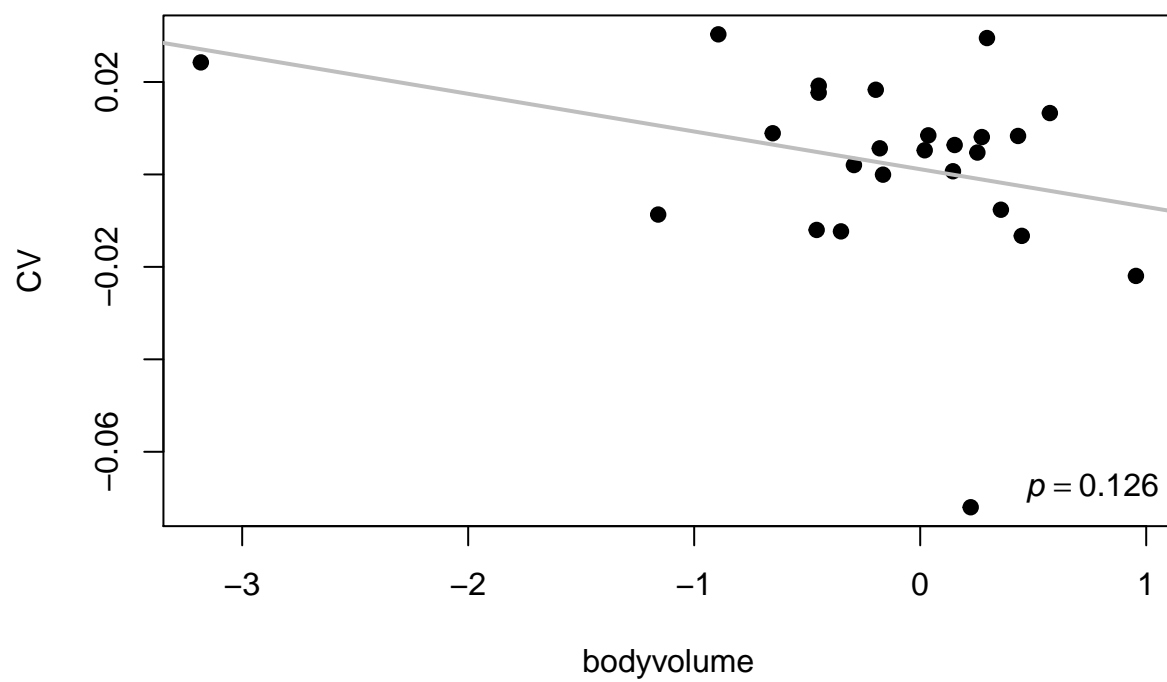
PIC cv bodylength male

$p = 0.0749$

CV
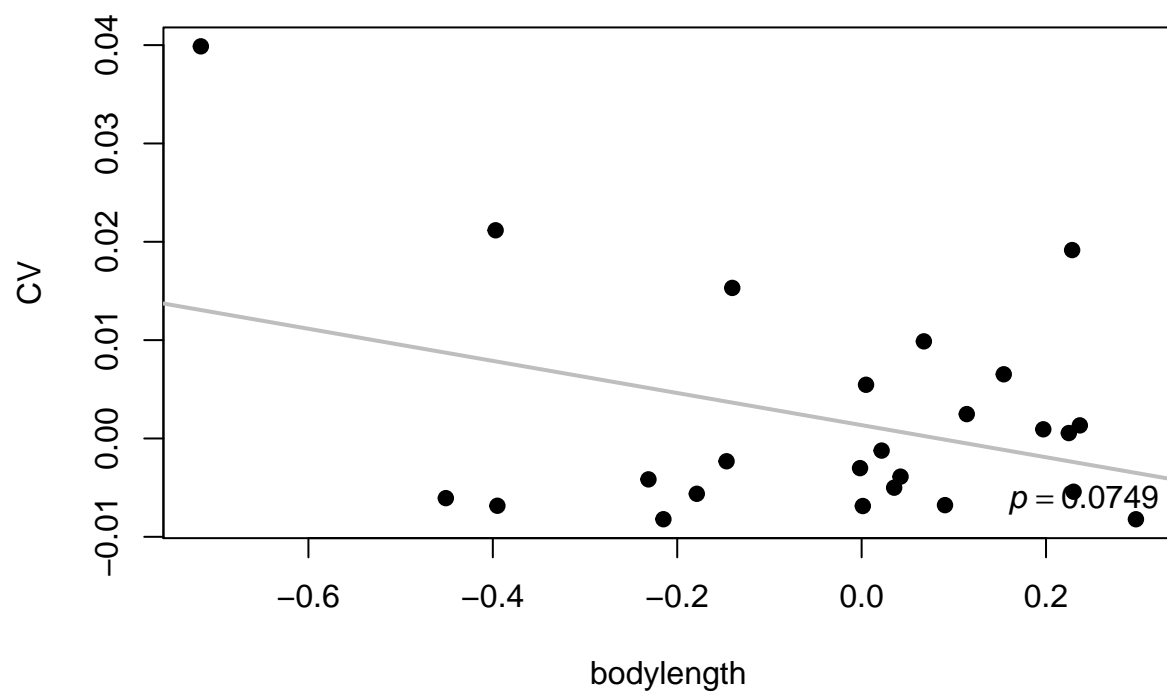
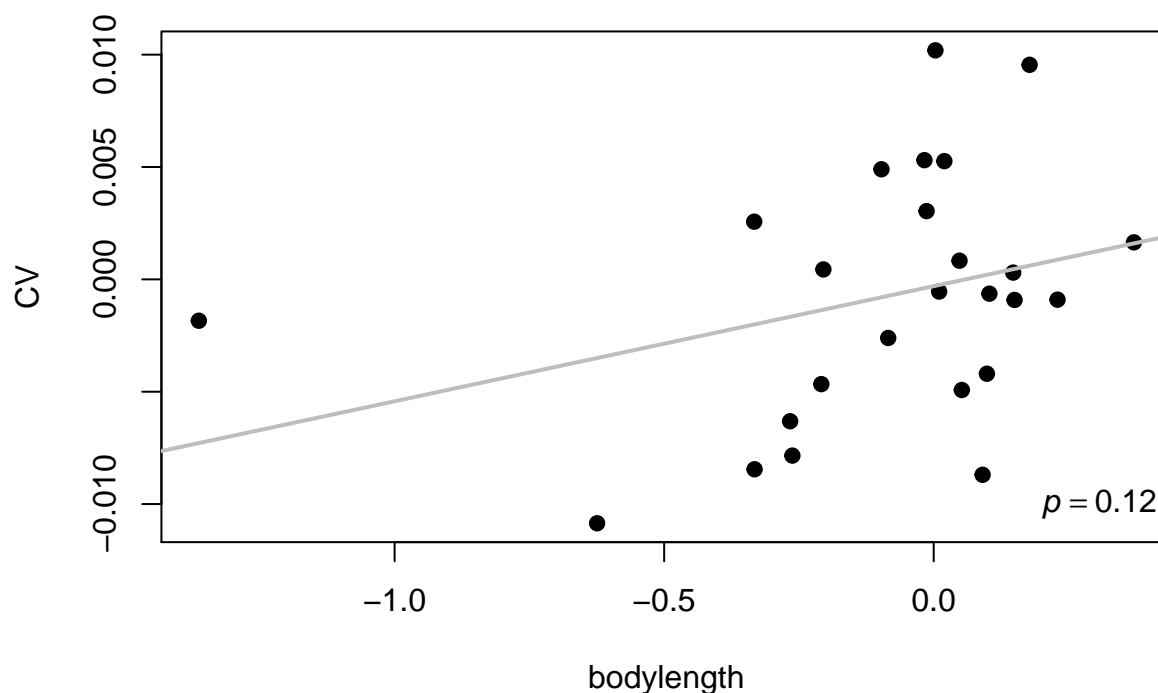bodylength

PIC cv bodylength female

table of all pairwise PIC correlations and p-values

tables with all pairwise PIC correlations. For the p-value table, those above diagonal are holm corrected

```
## Call:corr.test(x = totaldata)
## Correlation matrix
##                   spermsize oocytesizearea oocytesize malebodyarea
## spermsize              1.00           0.27       0.44         0.59
## oocytesizearea         0.27           1.00       0.96         0.58
## oocytesize             0.44           0.96       1.00         0.65
## malebodyarea           0.59           0.58       0.65         1.00
## femalebodyarea         0.36           0.39       0.51         0.81
## malebodylength         0.36           0.53       0.57         0.73
## femalebodylength       0.33           0.24       0.34         0.47
## malebodywidth          0.60           0.52       0.60         0.97
## femalebodywidth        0.27           0.35       0.44         0.73
## eggsize               -0.49           0.15      -0.01         0.21
## primaryspermatocyte    0.40           0.15       0.31         0.26
##                   femalebodyarea malebodylength femalebodylength
## spermsize                   0.36           0.36             0.33
## oocytesizearea              0.39           0.53             0.24
## oocytesize                  0.51           0.57             0.34
## malebodyarea                0.81           0.73             0.47
## femalebodyarea              1.00           0.61             0.59
## malebodylength              0.61           1.00             0.84
```

```
## femalebodylength                    0.59                0.84                    1.00
## malebodywidth                       0.77                0.53                    0.27
## femalebodywidth                     0.90                0.29                    0.18
## eggsize                             0.24                0.10                   -0.06
## primaryspermatocyte                 0.24               -0.03                   -0.04
##                       malebodywidth femalebodywidth eggsize
## spermsize                      0.60                0.27  -0.49
## oocytesizearea                 0.52                0.35   0.15
## oocytesize                     0.60                0.44  -0.01
## malebodyarea                   0.97                0.73   0.21
## femalebodyarea                 0.77                0.90   0.24
## malebodylength                 0.53                0.29   0.10
## femalebodylength               0.27                0.18  -0.06
## malebodywidth                  1.00                0.80   0.19
## femalebodywidth                0.80                1.00   0.32
## eggsize                        0.19                0.32   1.00
## primaryspermatocyte            0.28                0.22  -0.11
##                       primaryspermatocyte
## spermsize                            0.40
## oocytesizearea                       0.15
## oocytesize                           0.31
## malebodyarea                         0.26
## femalebodyarea                       0.24
## malebodylength                      -0.03
## femalebodylength                    -0.04
## malebodywidth                        0.28
## femalebodywidth                      0.22
## eggsize                             -0.11
## primaryspermatocyte                  1.00
## Sample Size
##                       spermsize oocytesizearea oocytesize malebodyarea
## spermsize                    25             25         25           25
## oocytesizearea               25             25         25           25
## oocytesize                   25             25         25           25
## malebodyarea                 25             25         25           25
## femalebodyarea               25             25         25           25
## malebodylength               25             25         25           25
## femalebodylength             25             25         25           25
## malebodywidth                25             25         25           25
## femalebodywidth              25             25         25           25
## eggsize                      24             24         24           24
## primaryspermatocyte           9              9          9            9
##                       femalebodyarea malebodylength femalebodylength
## spermsize                         25             25               25
## oocytesizearea                    25             25               25
## oocytesize                        25             25               25
## malebodyarea                      25             25               25
## femalebodyarea                    25             25               25
## malebodylength                    25             25               25
## femalebodylength                  25             25               25
## malebodywidth                     25             25               25
## femalebodywidth                   25             25               25
## eggsize                           24             24               24
## primaryspermatocyte                9              9                9
```

```
##                     malebodywidth femalebodywidth eggsize
## spermsize                      25              25      24
## oocytesizearea                 25              25      24
## oocytesize                     25              25      24
## malebodyarea                   25              25      24
## femalebodyarea                 25              25      24
## malebodylength                 25              25      24
## femalebodylength               25              25      24
## malebodywidth                  25              25      24
## femalebodywidth                25              25      24
## eggsize                        24              24      24
## primaryspermatocyte            9               9       9
##                     primaryspermatocyte
## spermsize                            9
## oocytesizearea                       9
## oocytesize                           9
## malebodyarea                         9
## femalebodyarea                       9
## malebodylength                       9
## femalebodylength                     9
## malebodywidth                        9
## femalebodywidth                      9
## eggsize                              9
## primaryspermatocyte                  9
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                     spermsize oocytesizearea oocytesize malebodyarea
## spermsize                0.00           1.00       0.90         0.08
## oocytesizearea           0.20           0.00       0.00         0.09
## oocytesize               0.03           0.00       0.00         0.02
## malebodyarea             0.00           0.00       0.00         0.00
## femalebodyarea           0.07           0.05       0.01         0.00
## malebodylength           0.08           0.01       0.00         0.00
## femalebodylength         0.10           0.25       0.10         0.02
## malebodywidth            0.00           0.01       0.00         0.00
## femalebodywidth          0.20           0.09       0.03         0.00
## eggsize                  0.02           0.49       0.98         0.33
## primaryspermatocyte      0.28           0.71       0.42         0.50
##                     femalebodyarea malebodylength femalebodylength
## spermsize                     1.00           1.00             1.00
## oocytesizearea                1.00           0.23             1.00
## oocytesize                    0.34           0.12             1.00
## malebodyarea                  0.00           0.00             0.57
## femalebodyarea                0.00           0.06             0.08
## malebodylength                0.00           0.00             0.00
## femalebodylength              0.00           0.00             0.00
## malebodywidth                 0.00           0.01             0.20
## femalebodywidth               0.00           0.15             0.38
## eggsize                       0.26           0.65             0.77
## primaryspermatocyte           0.54           0.94             0.93
##                     malebodywidth femalebodywidth eggsize
## spermsize                    0.07            1.00    0.52
## oocytesizearea               0.27            1.00    1.00
## oocytesize                   0.07            0.90    1.00
## malebodyarea                 0.00            0.00    1.00
```

11

```
## femalebodyarea                0.00        0.00   1.00
## malebodylength                0.26        1.00   1.00
## femalebodylength              1.00        1.00   1.00
## malebodywidth                 0.00        0.00   1.00
## femalebodywidth               0.00        0.00   1.00
## eggsize                       0.37        0.12   0.00
## primaryspermatocyte           0.47        0.57   0.77
##                   primaryspermatocyte
## spermsize                            1
## oocytesizearea                       1
## oocytesize                           1
## malebodyarea                         1
## femalebodyarea                       1
## malebodylength                       1
## femalebodylength                     1
## malebodywidth                        1
## femalebodywidth                      1
## eggsize                              1
## primaryspermatocyte                  0
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```
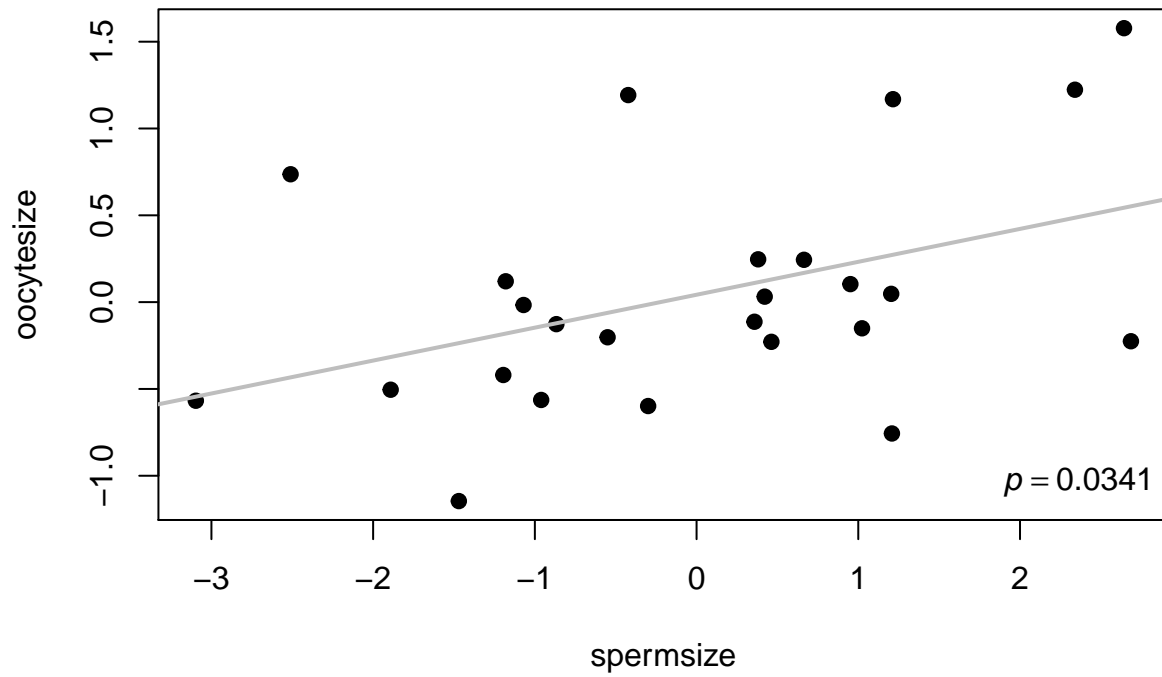
PIC of spermsize vs oocytesize

PIC of spermsize vs spermatocytesize
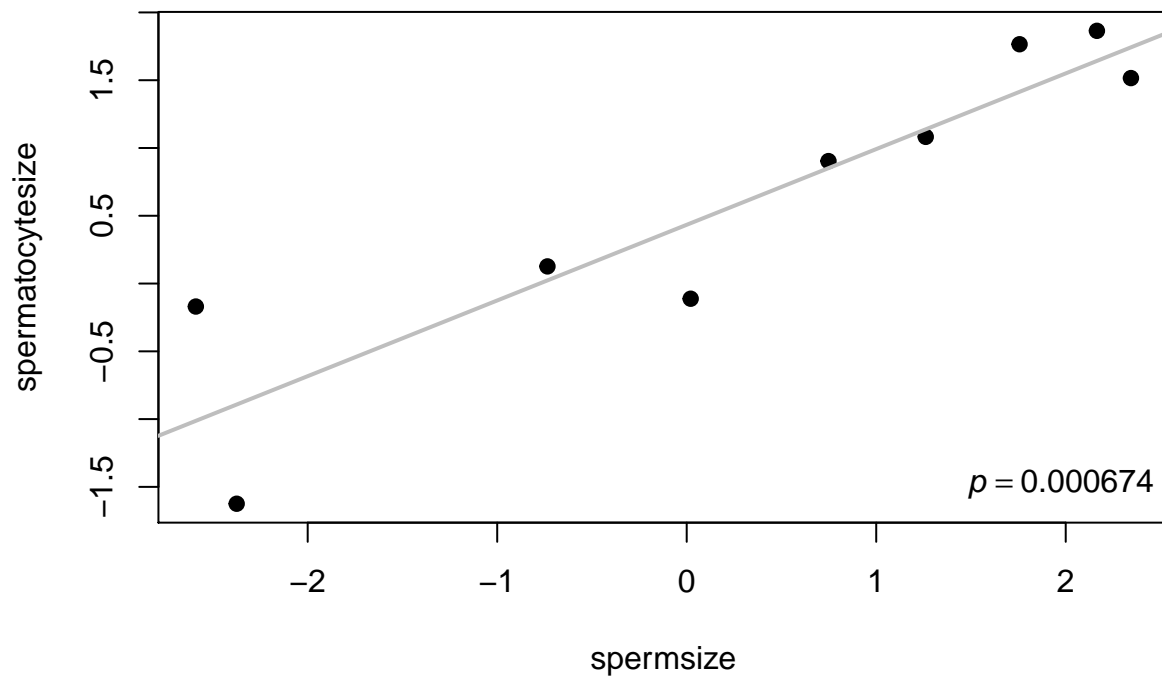
PIC of spermatocytesize vs oocytesize

$p = 0.128$

spermatocyte

oocytesize

PIC of spermsize vs oocytearea

$p = 0.58$

oocytearea

spermsize

PIC of spermatocytesize vs oocytearea

spermatocyte

oocytearea

$p = 0.42$
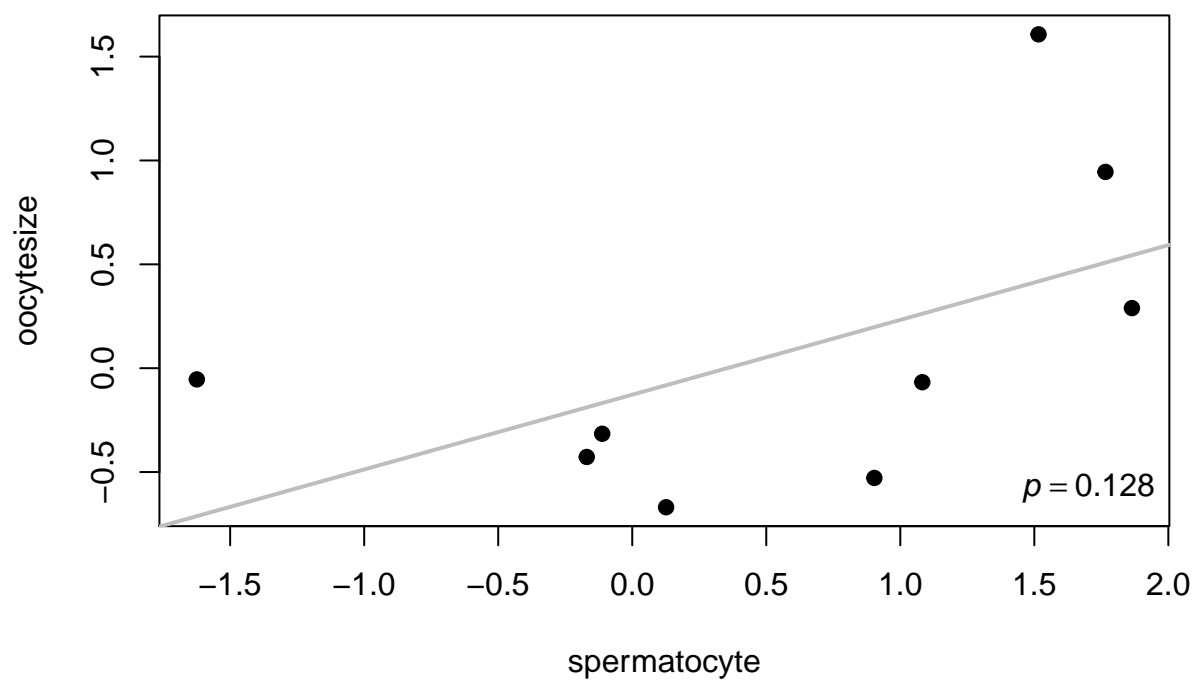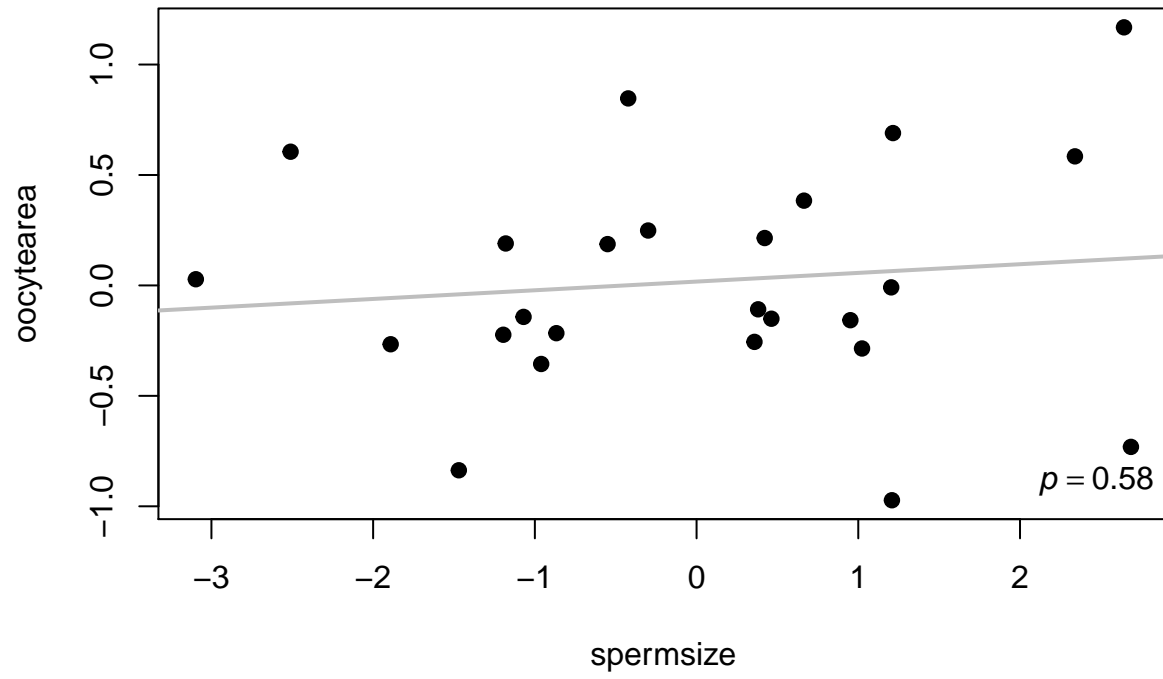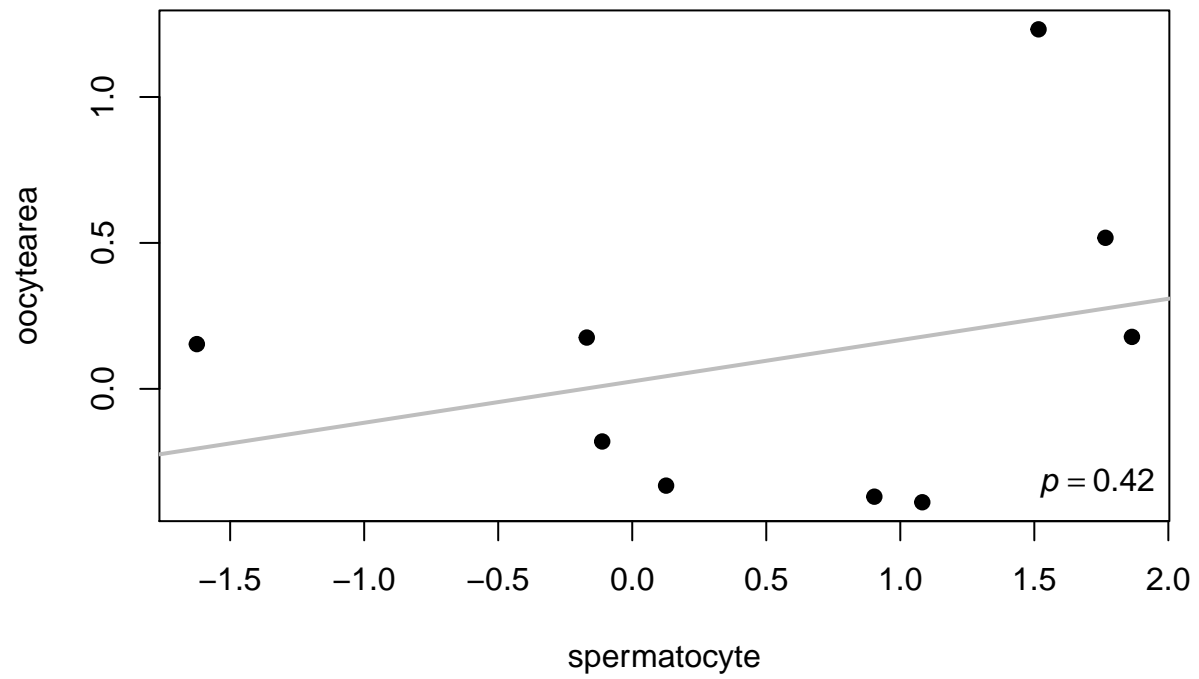
PIC of oocytesize vs oocytearea

$p = 4.56\mathrm{e}{-08}$

oocytesize

oocytearea
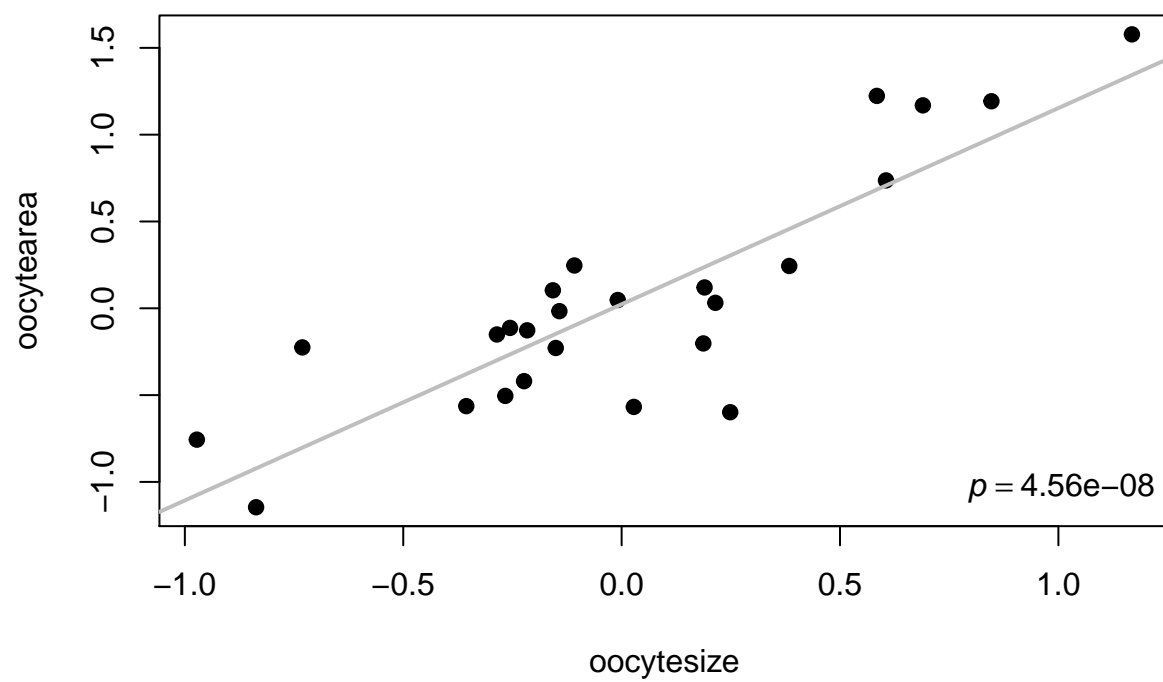
Body area male vs female

$p = 0.000121$
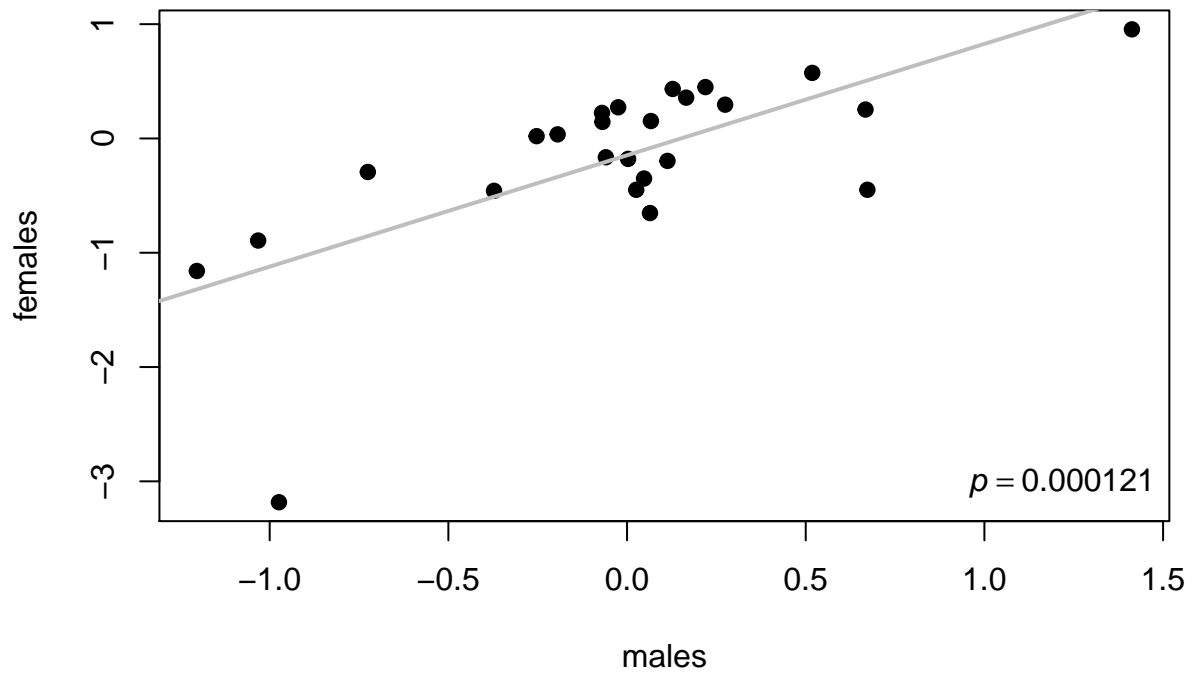
Body area female vs oocytesize

females

oocytes

$p = 0.0517$
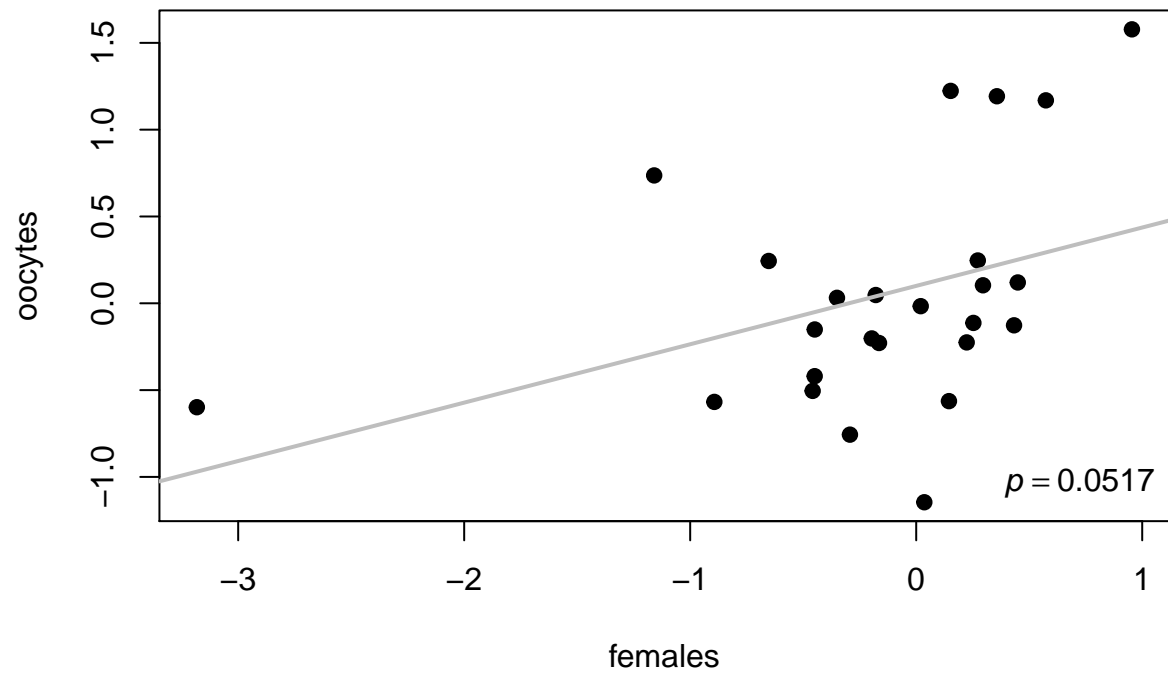
Body area male vs spermsize

$p = 0.00205$

sperms

males
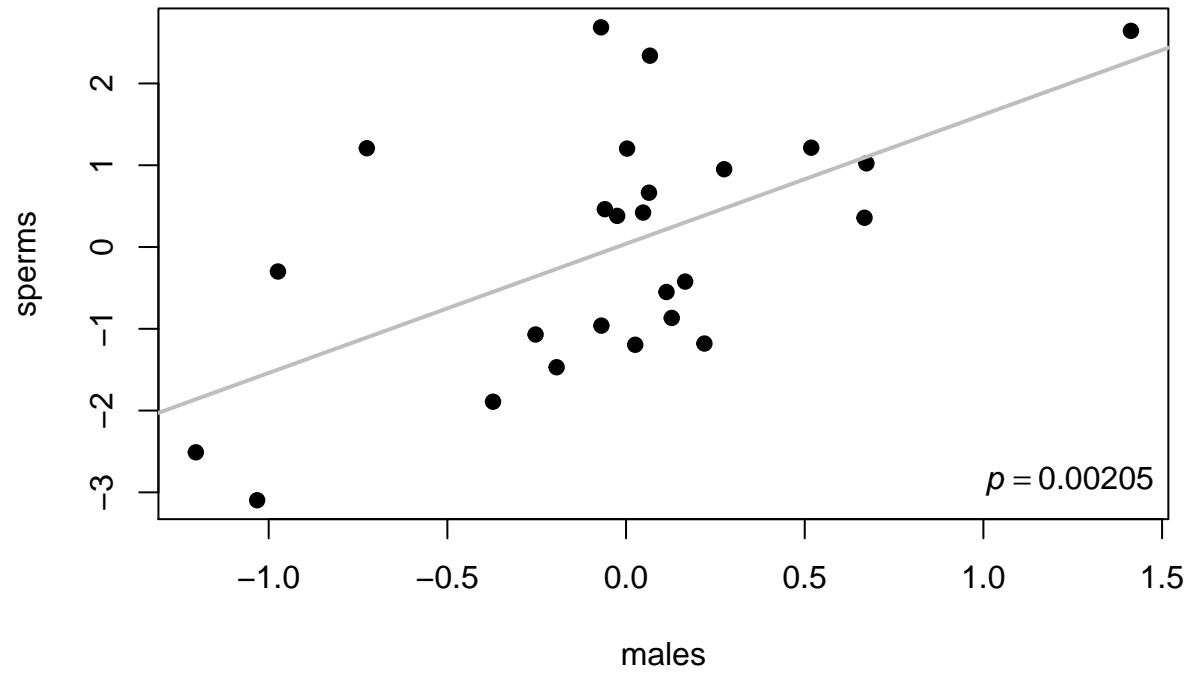
Body area male vs spermatocytesize
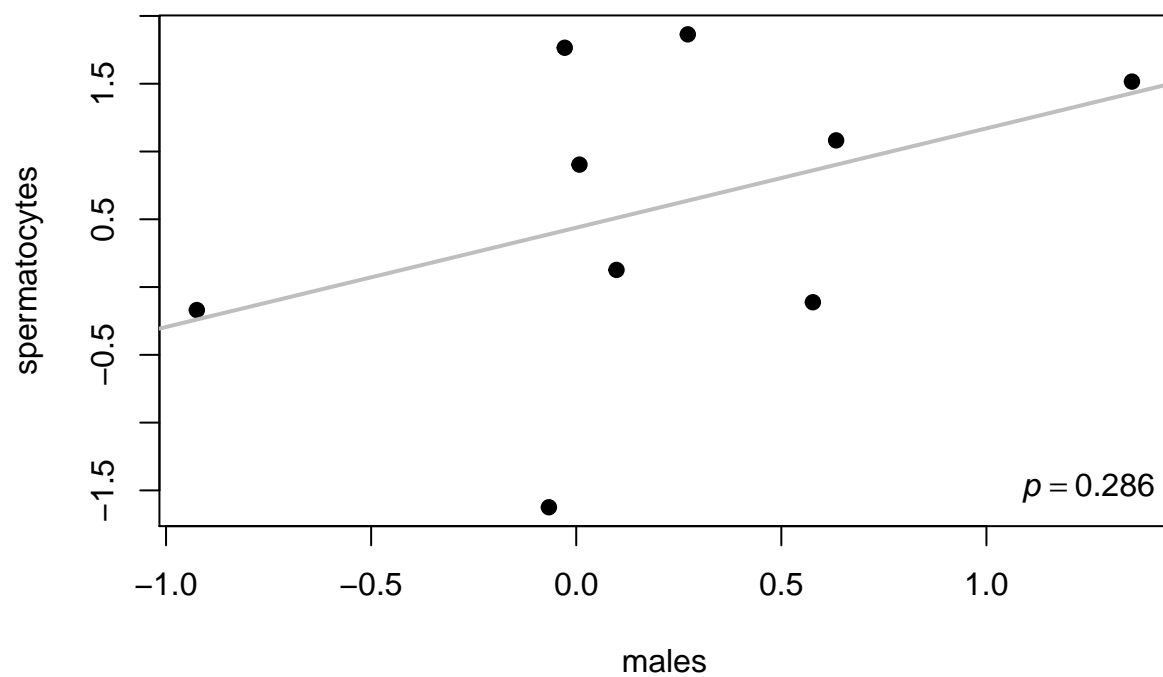
Body length male vs female

$p = 3.02\text{e}{-}09$
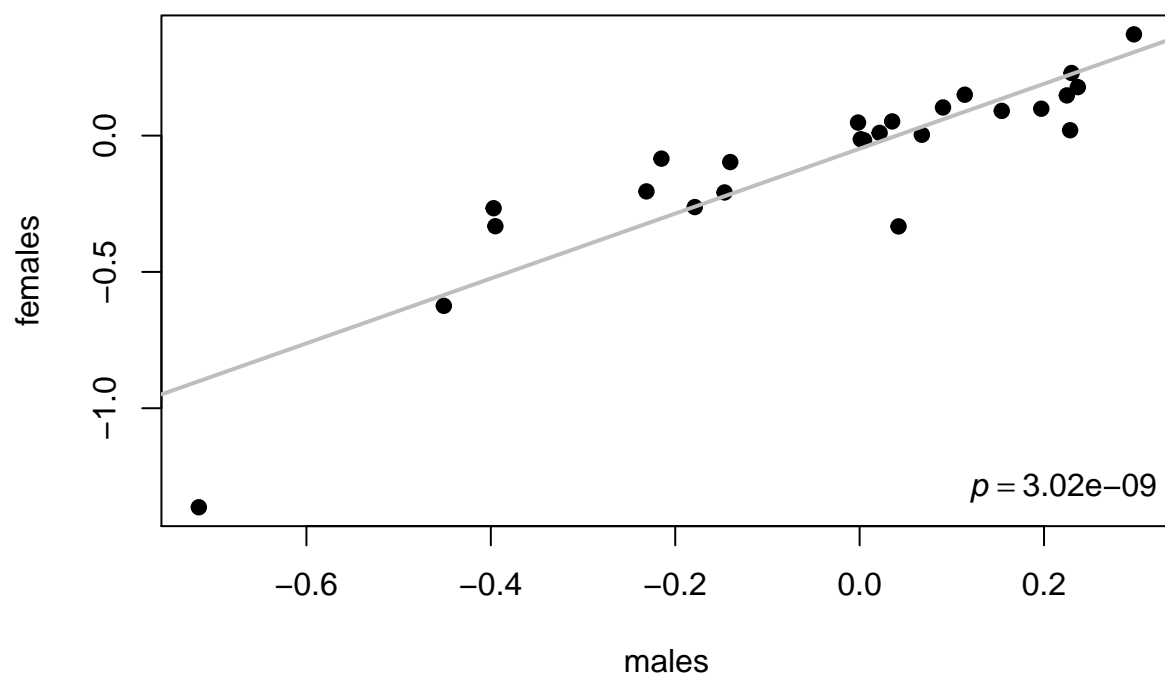
Body length female vs oocytesize
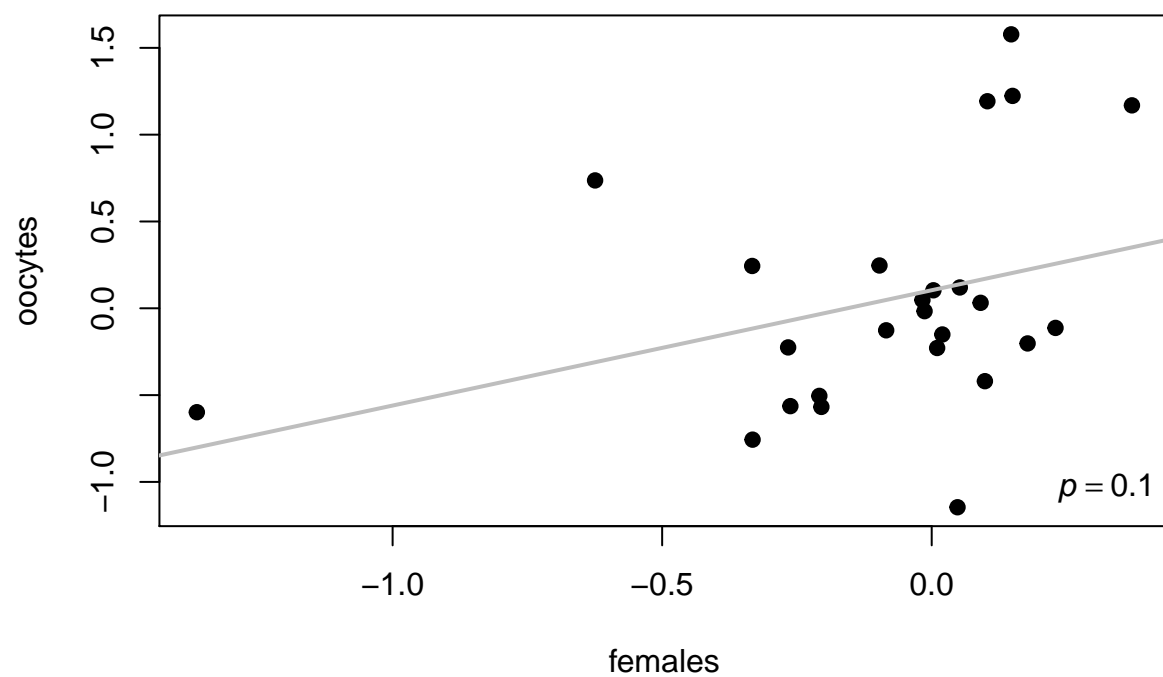
Body length male vs spermsize

$p = 0.217$

Body length male vs spermatocytesize

$p = 0.697$

spermatocytes

males
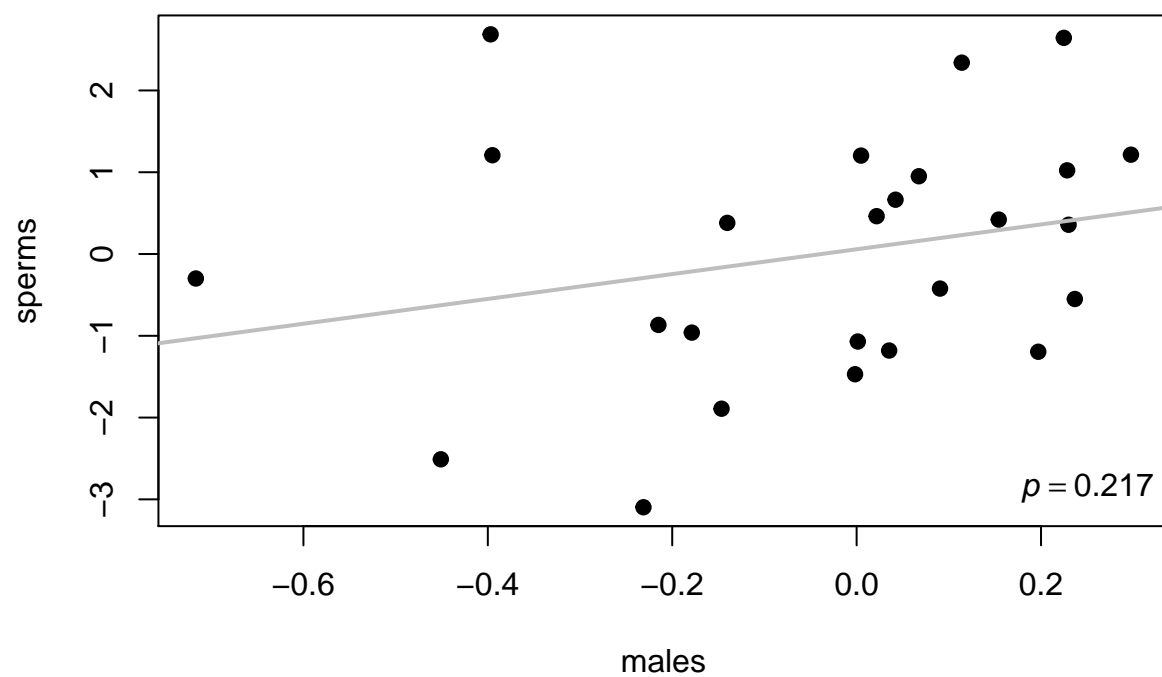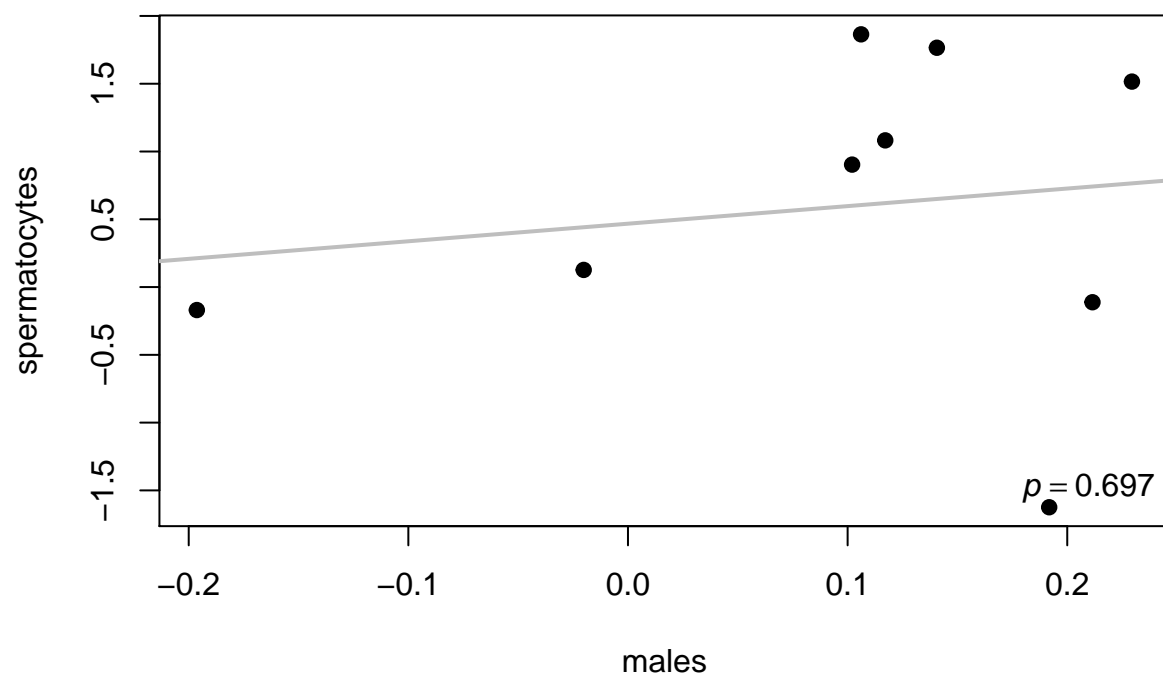
eggsize vs female length

$p = 0.132$

x-axis: egg size

y-axis: length
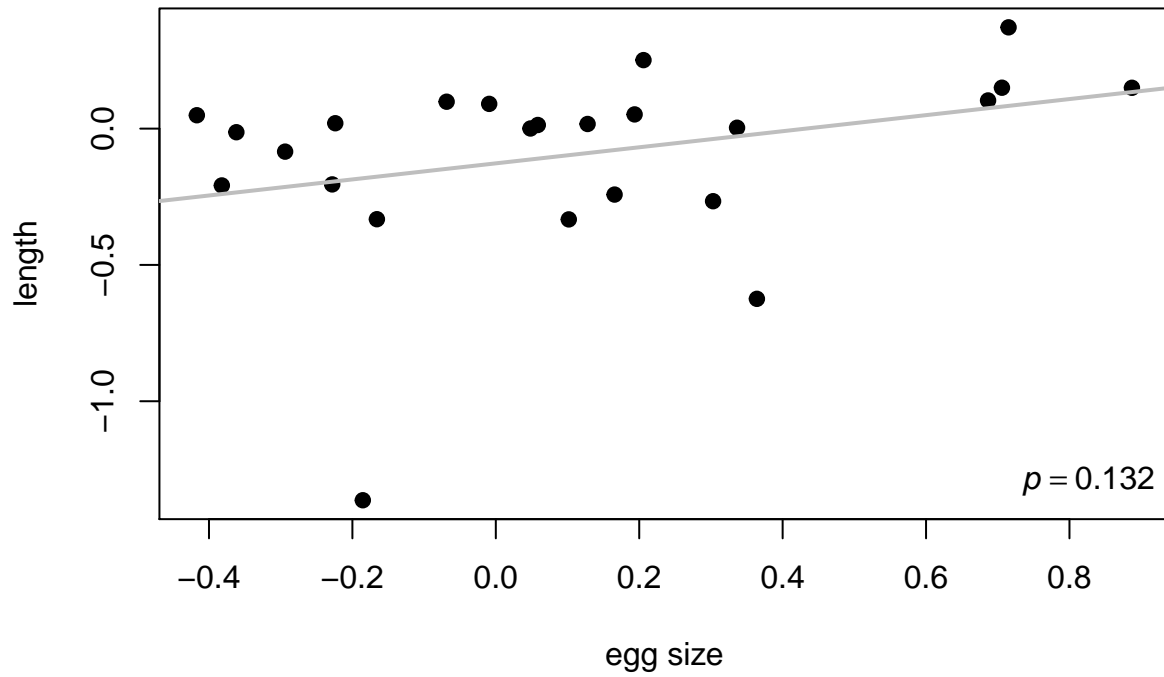
eggsize vs female volume

$p = 0.0455$

volume

egg size

eggsize vs sperm size

sperm size

egg size

$p = 0.00848$

eggsize vs oocyte size
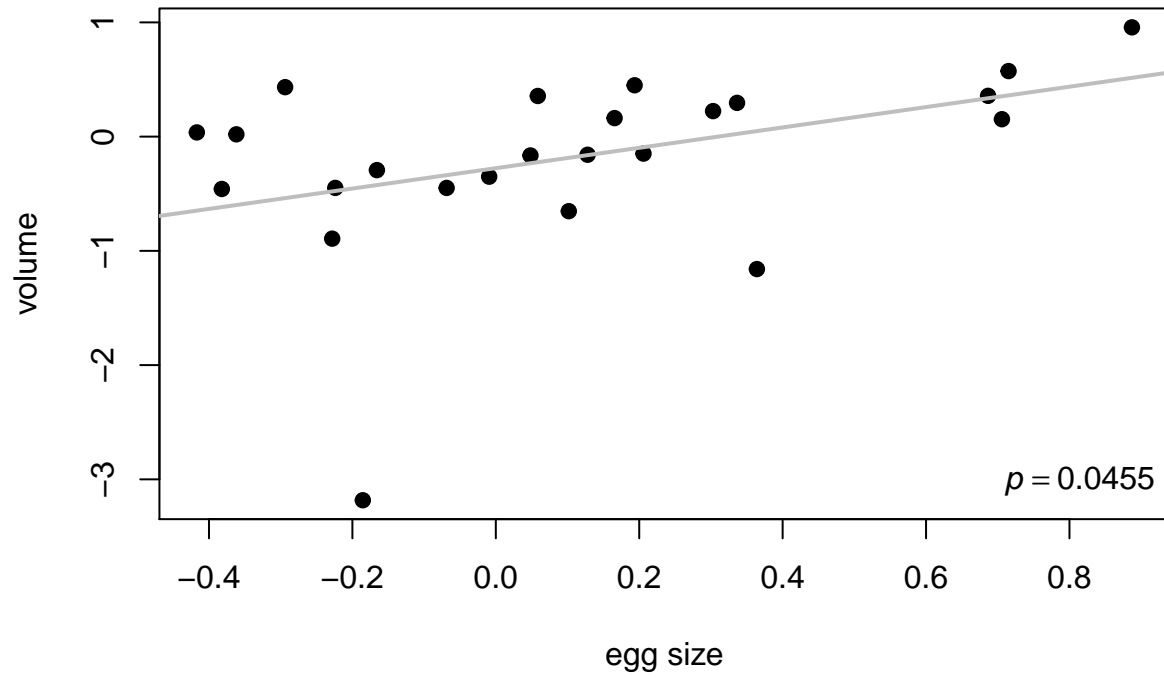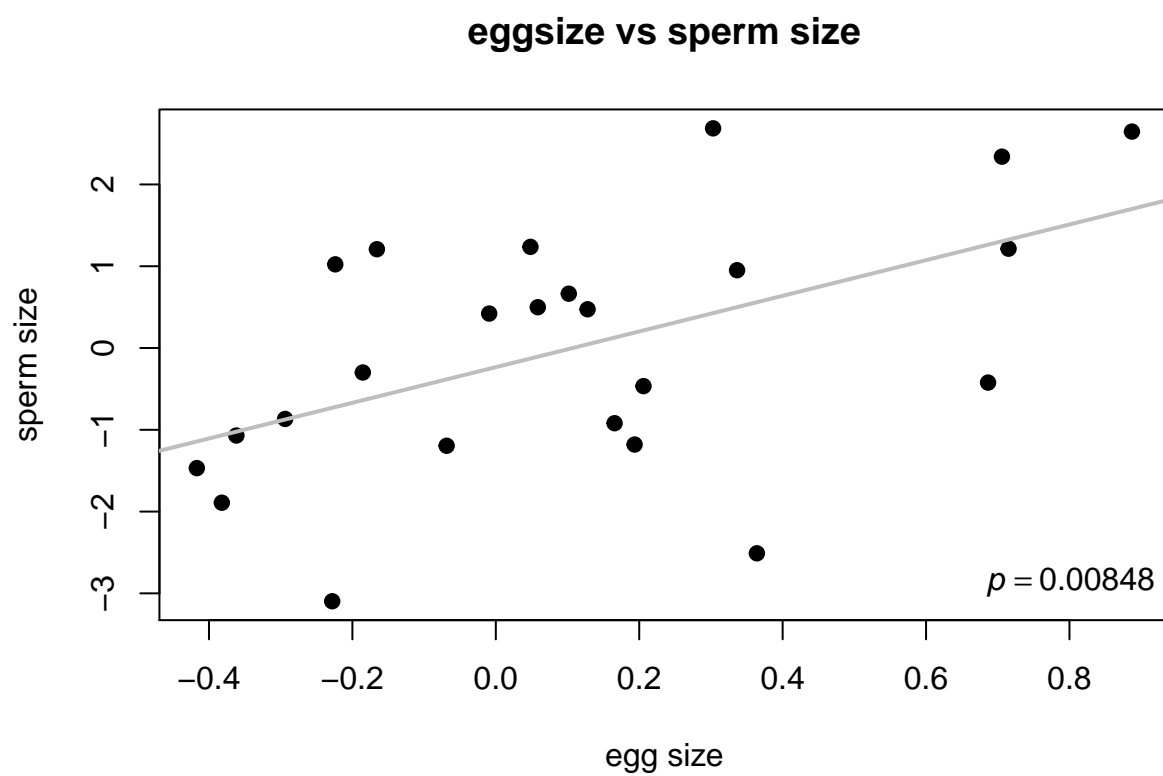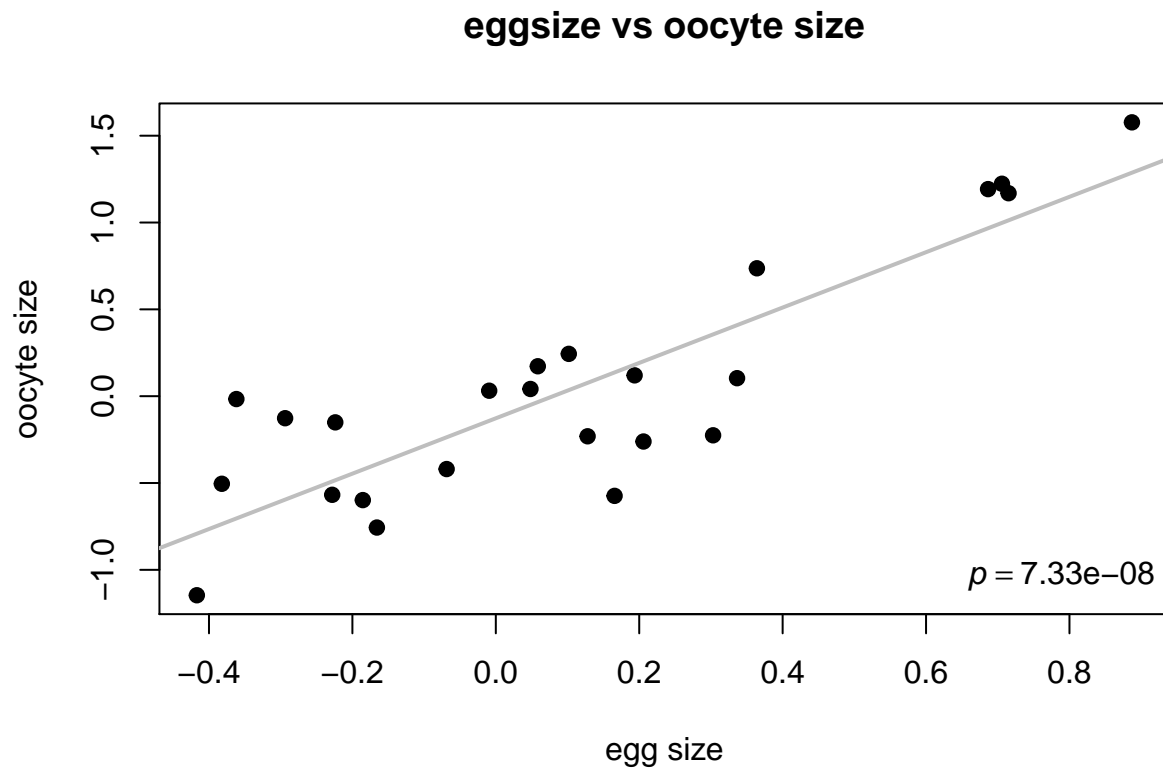
PGLS - testing for correlation with evolutionary models

Output are summary tables from fits. The OU model won't fit the spermatocyte fits due to overparameterisation.

```
## 
## Call:
## pgls(formula = meansperm ~ meanooarea, data = compsperm, lambda = "ML")
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9643 -0.5988  0.4323  0.6861  1.8263
## 
## Branch length transformations:
## 
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.000
##    lower bound : 0.000, p = 1
##    upper bound : 1.000, p = 0.11287
##    95.0% CI   : (NA, NA)
## delta  [Fix]  : 1.000
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.09037    6.57159 -0.9268   0.3812
```

```
## meanooarea    1.46171    0.94975  1.5390    0.1624
##
## Residual standard error: 1.057 on 8 degrees of freedom
## Multiple R-squared: 0.2284,  Adjusted R-squared: 0.132
## F-statistic: 2.369 on 1 and 8 DF,  p-value: 0.1624


##
## Call:
## pgls(formula = meansperm ~ meanoo, data = compsperm, lambda = "ML")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4586 -1.1960  0.6374  1.0989  1.4686
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.895
##    lower bound : 0.000, p = 1
##    upper bound : 1.000, p = 0.79696
##    95.0% CI   : (NA, NA)
## delta  [Fix]  : 1.000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11.31303    6.94067 -1.6300  0.14176
## meanoo        1.50685    0.65983  2.2837  0.05177 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.315 on 8 degrees of freedom
## Multiple R-squared: 0.3946,  Adjusted R-squared: 0.319
## F-statistic: 5.215 on 1 and 8 DF,  p-value: 0.05177


##
## Call:
## pgls(formula = meansperm ~ meanprimsp, data = compsperm, lambda = "ML")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8244 -0.3185 -0.1789  0.1091  0.5614
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.000
##    lower bound : 0.000, p = 1
##    upper bound : 1.000, p = 0.0066141
##    95.0% CI   : (NA, 0.708)
## delta  [Fix]  : 1.000
##
## Coefficients:
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -2.63590    0.94589 -2.7867 0.0236793 *
```

```
## meanprimsp    1.21417     0.17131   7.0873 0.0001033 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.446 on 8 degrees of freedom
## Multiple R-squared: 0.8626,  Adjusted R-squared: 0.8454
## F-statistic: 50.23 on 1 and 8 DF,  p-value: 0.0001033


## Generalized least squares fit by REML
##   Model: meansperm ~ meanooarea
##   Data: fullpglsframe
##        AIC      BIC    logLik
##   54.95884 59.67106 -23.47942
##
## Correlation Structure: corMartins
##  Formula: ~1
##  Parameter estimate(s):
##    alpha
## 6.841132
##
## Coefficients:
##                 Value Std.Error     t-value p-value
## (Intercept) -3.044212  3.398811 -0.8956698  0.3793
## meanooarea   1.005070  0.498035  2.0180725  0.0549
##
##  Correlation:
##            (Intr)
## meanooarea -0.999
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.7775727 -0.6837636 -0.2886598  0.4237510  2.2862305
##
## Residual standard error: 0.6334179
## Degrees of freedom: 26 total; 24 residual


## Generalized least squares fit by REML
##   Model: meansperm ~ meanoo
##   Data: fullpglsframe
##        AIC      BIC    logLik
##   50.30285 55.01506 -21.15142
##
## Correlation Structure: corMartins
##  Formula: ~1
##  Parameter estimate(s):
##    alpha
## 5.469326
##
## Coefficients:
##                 Value Std.Error    t-value p-value
## (Intercept) -7.068653  3.346328 -2.112361  0.0452
## meanoo       1.074887  0.329803  3.259179  0.0033
##
##  Correlation:
```

```
##        (Intr)
## meanoo -0.999
##
## Standardized residuals:
##        Min          Q1          Med          Q3          Max
## -1.9084685 -0.7013875 -0.3474741  0.5038124  2.4034645
##
## Residual standard error: 0.5843052
## Degrees of freedom: 26 total; 24 residual
```

Some phylogenetic signal methods - lambda

```
## Loading required package: parallel
```

```
##       Length Class       Mode
## lik   1      bm          function
## bnd   2      data.frame  list
## res 400      -none-      numeric
## opt   8      -none-      list
```

```
##       Length Class       Mode
## lik   1      bm          function
## bnd   2      data.frame  list
## res 400      -none-      numeric
## opt   8      -none-      list
```

```
##       Length Class       Mode
## lik   1      bm          function
## bnd   2      data.frame  list
## res 400      -none-      numeric
## opt   8      -none-      list
```

```
##       Length Class       Mode
## lik   1      bm          function
## bnd   2      data.frame  list
## res 400      -none-      numeric
## opt   8      -none-      list
```

Blombergs K

```
##           K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.3738911         2.289904              3.949588          0.026
##   PIC.variance.Z
## 1      -1.463943
```

```
##           K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.670685         0.4572932               1.21194          0.001
##   PIC.variance.Z
## 1      -2.091775
```

```
## [1] "Dropping tips from the tree because they are not present in the data:"
##  [1] "C. sp. 1"       "C. portoensis"  "C. virilis"
##  [4] "C. sp. 8"       "C. angaria"     "C. castelli"
##  [7] "C. sp. 2"       "C. kamaaina"    "C. imperialis"
## [10] "C. afra"        "C. nouraguensis" "C. doughertyi"
## [13] "C. wallacei"    "C. tropicalis"  "C. sp. 5"
## [16] "C. nigoni"
```

```
##          K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 1.218587         1.696502              2.598305          0.023
##   PIC.variance.Z
## 1      -1.929473
```

```
##           K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.5877787         0.2622437             0.6053382          0.004
##   PIC.variance.Z
## 1      -1.784902
```

Morans I

```
## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##     obs = res$obs, alter = alter, names = test.names)
##
## Number of tests:   1
##
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs    Std.Obs    Alter Pvalue
## 1    x -0.09396389 -1.414458 greater   0.94
##
## other elements: adj.method call
```

```
## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##     obs = res$obs, alter = alter, names = test.names)
##
## Number of tests:   1
##
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs    Std.Obs    Alter Pvalue
## 1    x -0.01730539 0.6084684 greater   0.217
##
## other elements: adj.method call
```

```
## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##     obs = res$obs, alter = alter, names = test.names)
```

```
## 
## Number of tests:   1
## 
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs     Std.Obs    Alter Pvalue
## 1    x -0.1482137 -0.9242992 greater  0.838
## 
## other elements: adj.method call


## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##     obs = res$obs, alter = alter, names = test.names)
## 
## Number of tests:   1
## 
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs    Std.Obs    Alter Pvalue
## 1    x -0.03082416 0.2634354 greater  0.347
## 
## other elements: adj.method call
```

abouheifs cmean

```
## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##     obs = res$obs, alter = alter, names = test.names)
## 
## Number of tests:   1
## 
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs    Std.Obs    Alter Pvalue
## 1    x -0.09396389 -1.398489 greater  0.938
## 
## other elements: adj.method call


## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##     obs = res$obs, alter = alter, names = test.names)
## 
## Number of tests:   1
## 
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs    Std.Obs    Alter Pvalue
## 1    x -0.01730539 0.6000796 greater  0.221
## 
## other elements: adj.method call
```

```
## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##      obs = res$obs, alter = alter, names = test.names)
##
## Number of tests:   1
##
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test        Obs     Std.Obs    Alter Pvalue
## 1    x -0.1482137 -0.9261158 greater  0.835
##
## other elements: adj.method call


## class: krandtest
## Monte-Carlo tests
## Call: as.krandtest(sim = matrix(res$result, ncol = nvar, byrow = TRUE),
##      obs = res$obs, alter = alter, names = test.names)
##
## Number of tests:   1
##
## Adjustment method for multiple comparisons:   none
## Permutation number:   999
##   Test         Obs    Std.Obs    Alter Pvalue
## 1    x -0.03082416 0.2456358 greater  0.353
##
## other elements: adj.method call
```
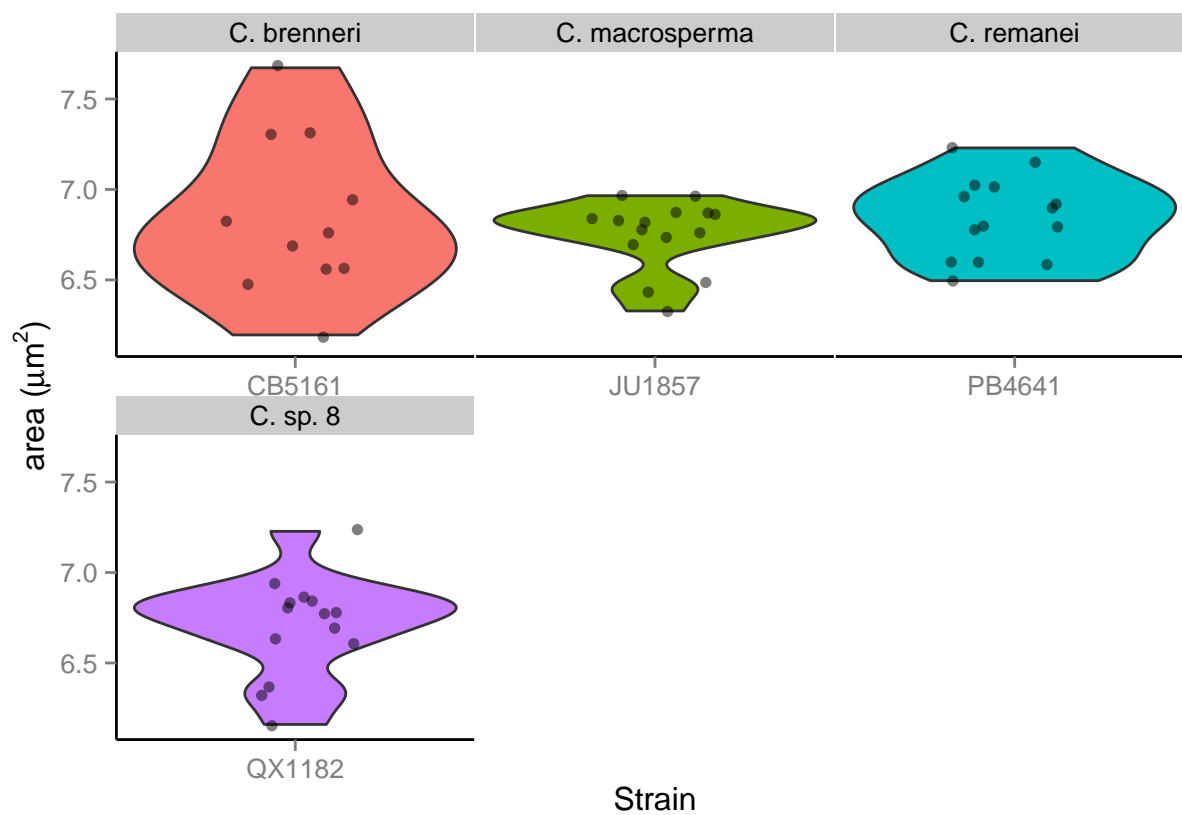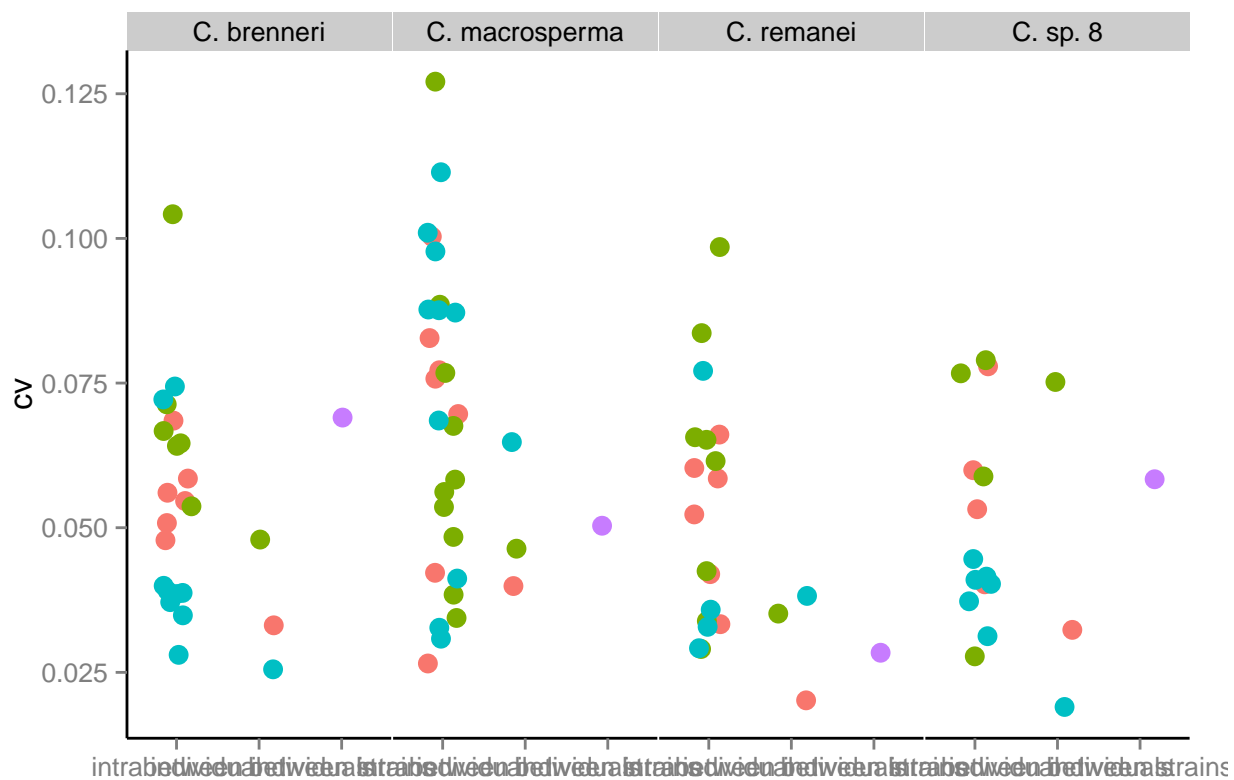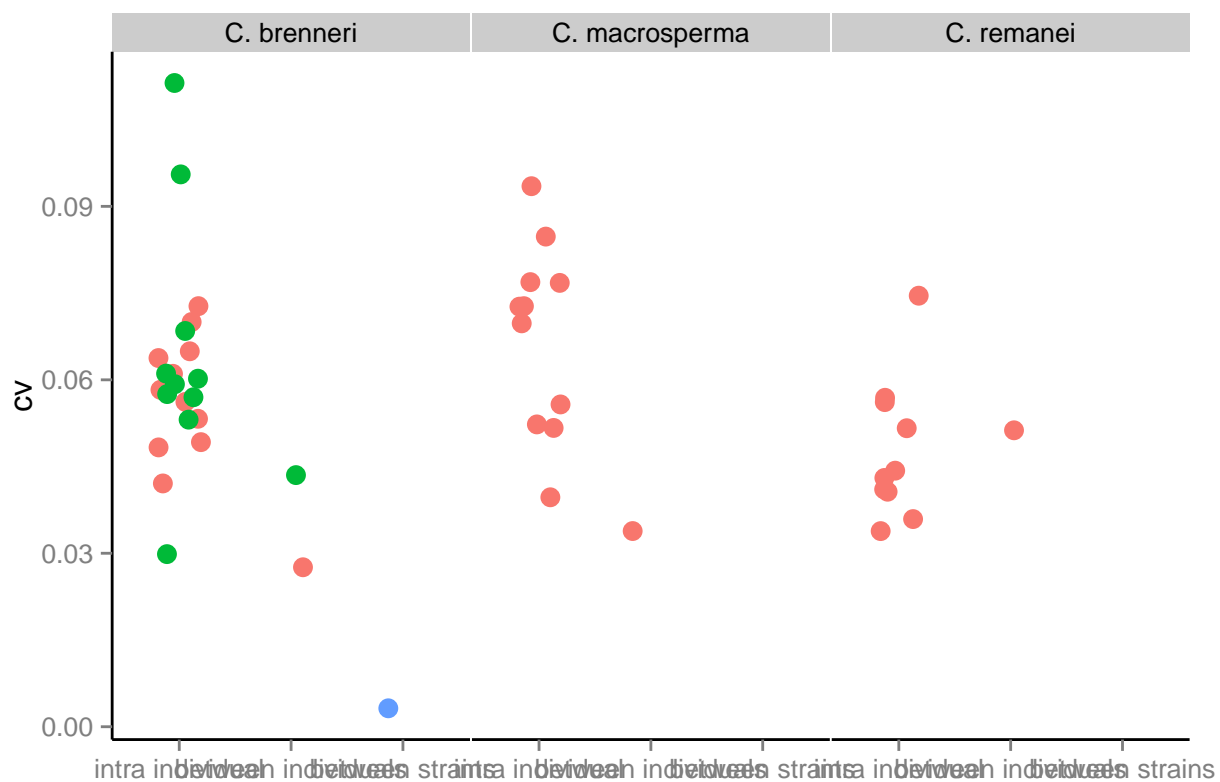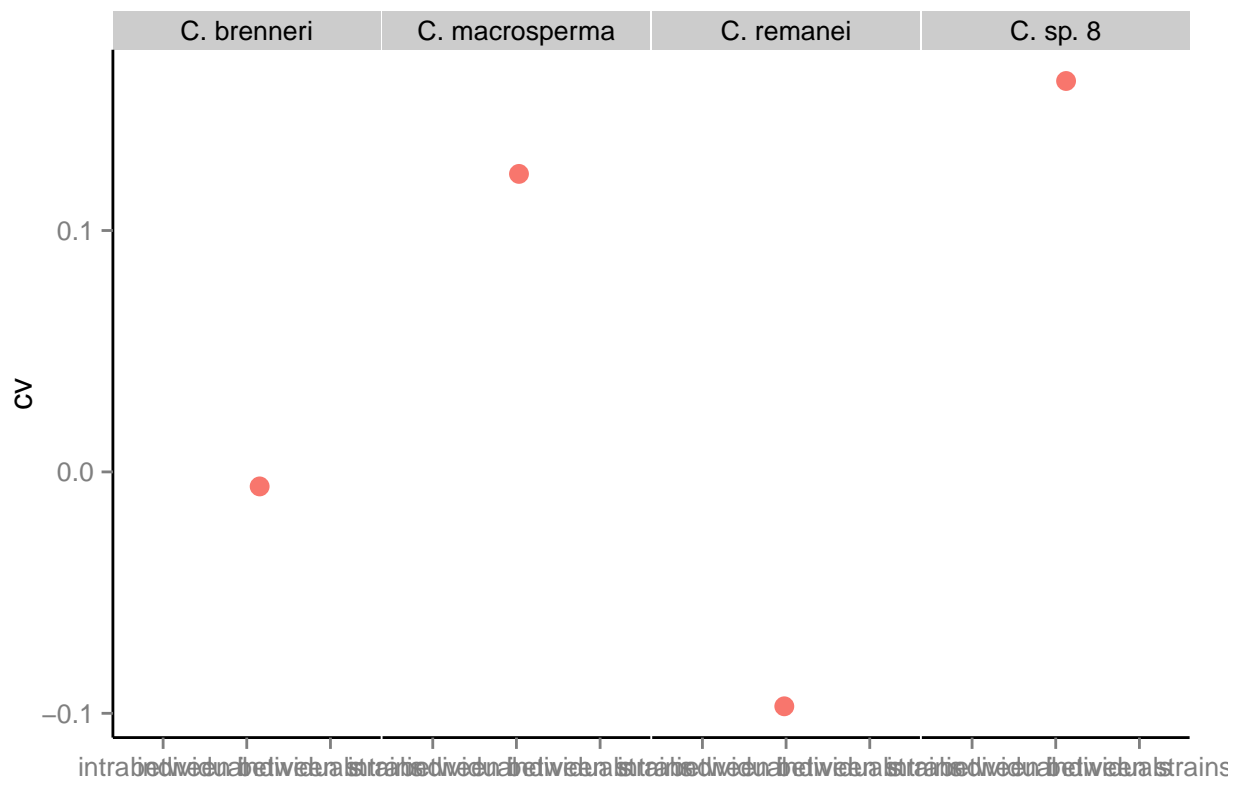
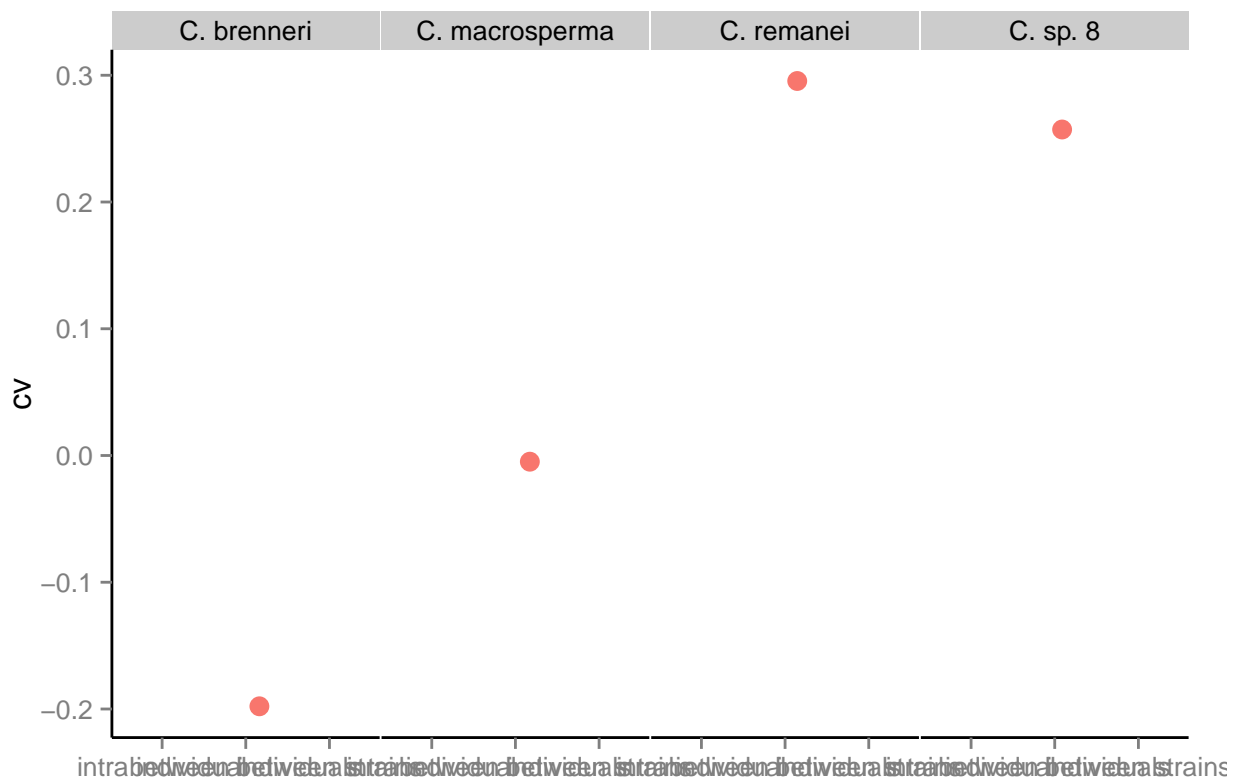figure 3A with violin plots
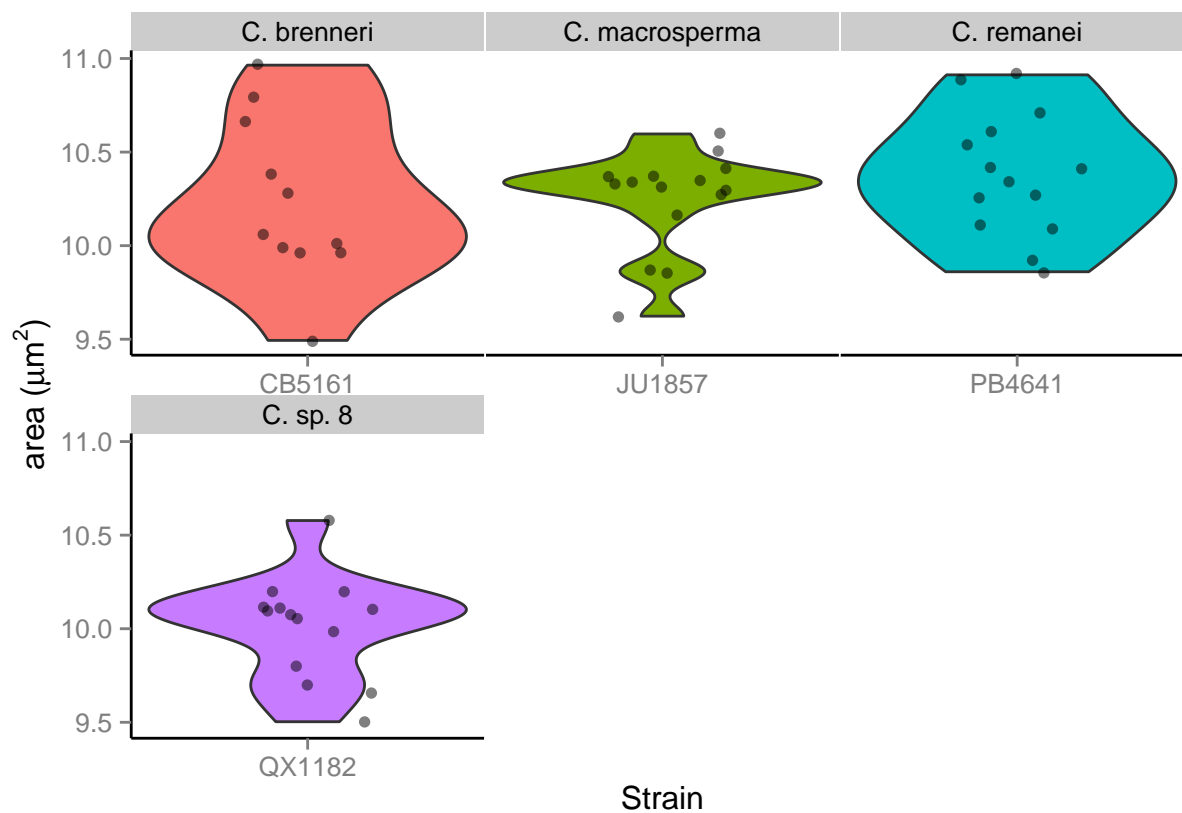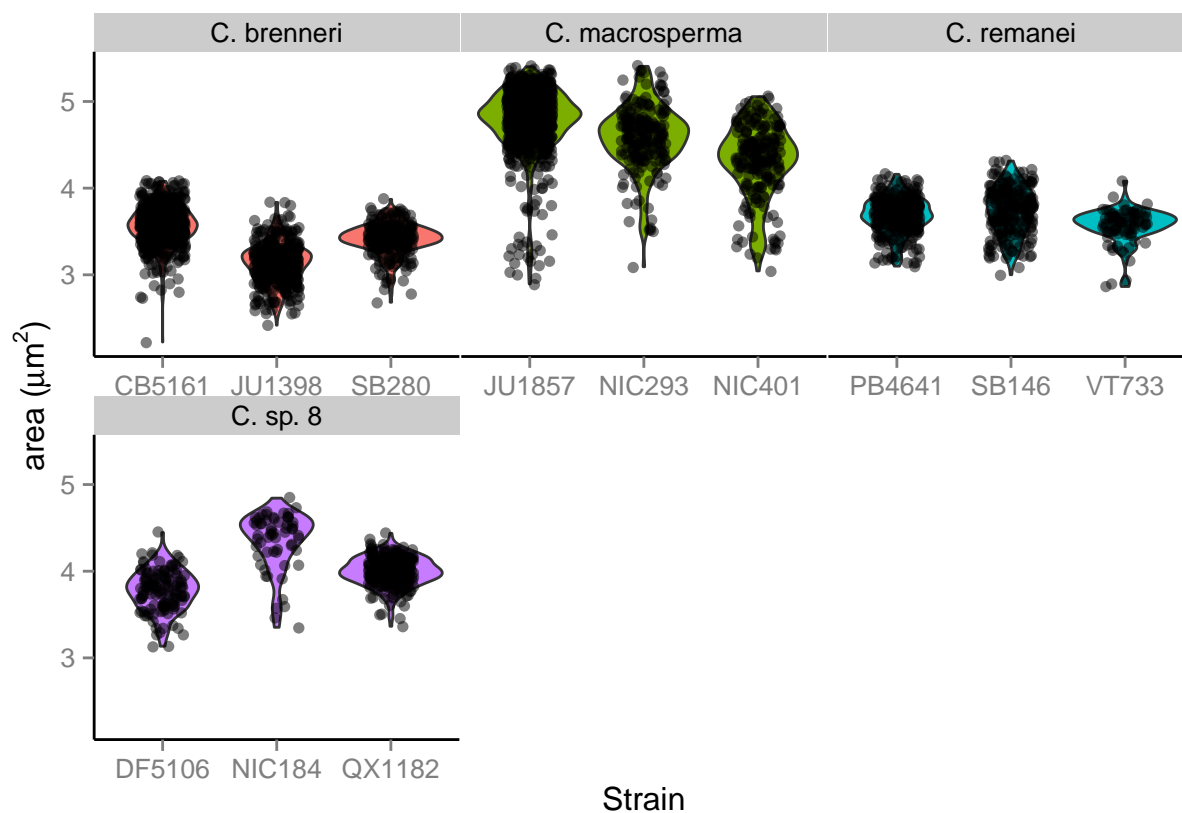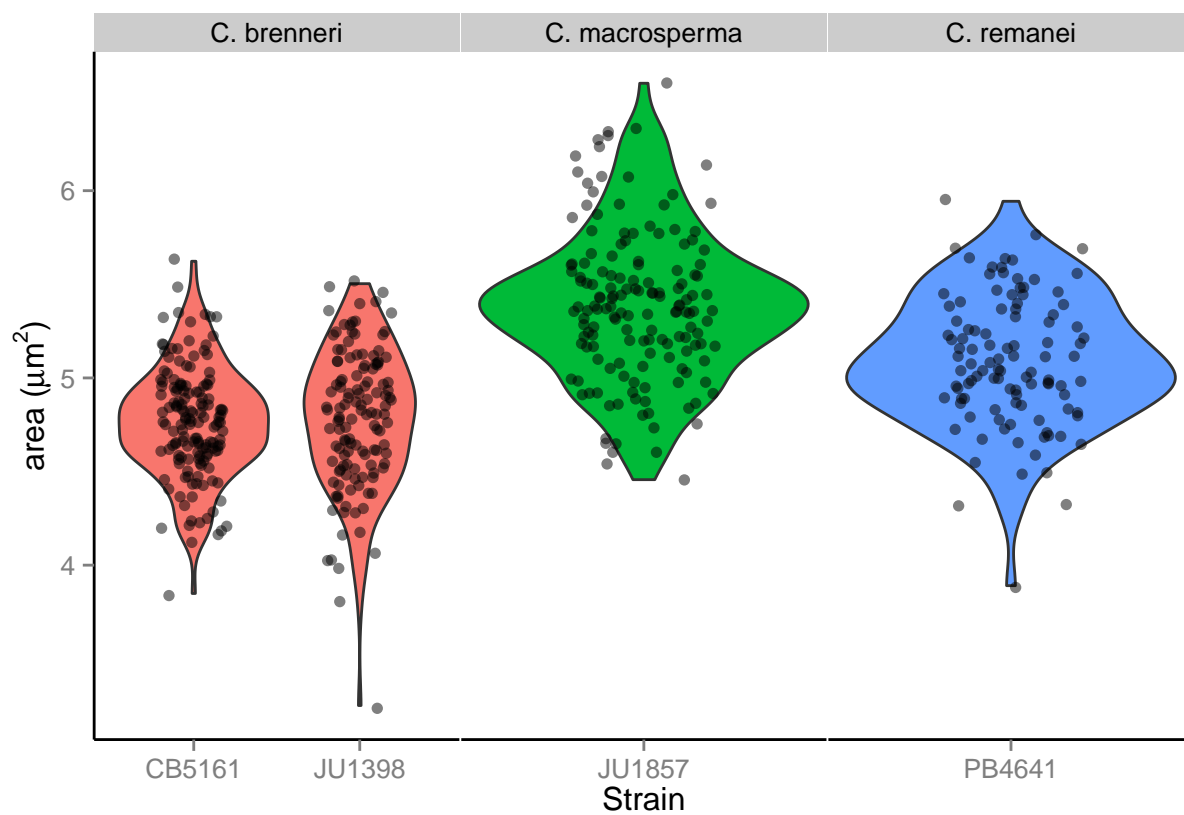
figure 3B

figure mockup a we are starting agin here, as I've modified some of the variables etc above Easier to start with a clean slate Clean and read in data:

mean and sd

```
groupedmeans=ff %>% group_by(Species,sex) %>% summarise(mean=mean(area,na.rm=T),n=n(),sd=sd(area,na.rm=T
limits <- aes(ymax = mean + 1.96*sd/sqrt(n), ymin=mean - 1.96*sd/sqrt(n))
dodge <- position_dodge(width=0.9)
ggplot(data=groupedmeans, aes(x=Species, y=mean, fill=sex)) + geom_bar(stat="identity", position=positio
    scale_fill_manual(values=c("gray","white"))+ylab( expression(paste("area (", mu, m^{2},")")))+xlab(
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
     panel.background = element_blank(),axis.line = element_line(colour = "black"),legend.position = "n
```
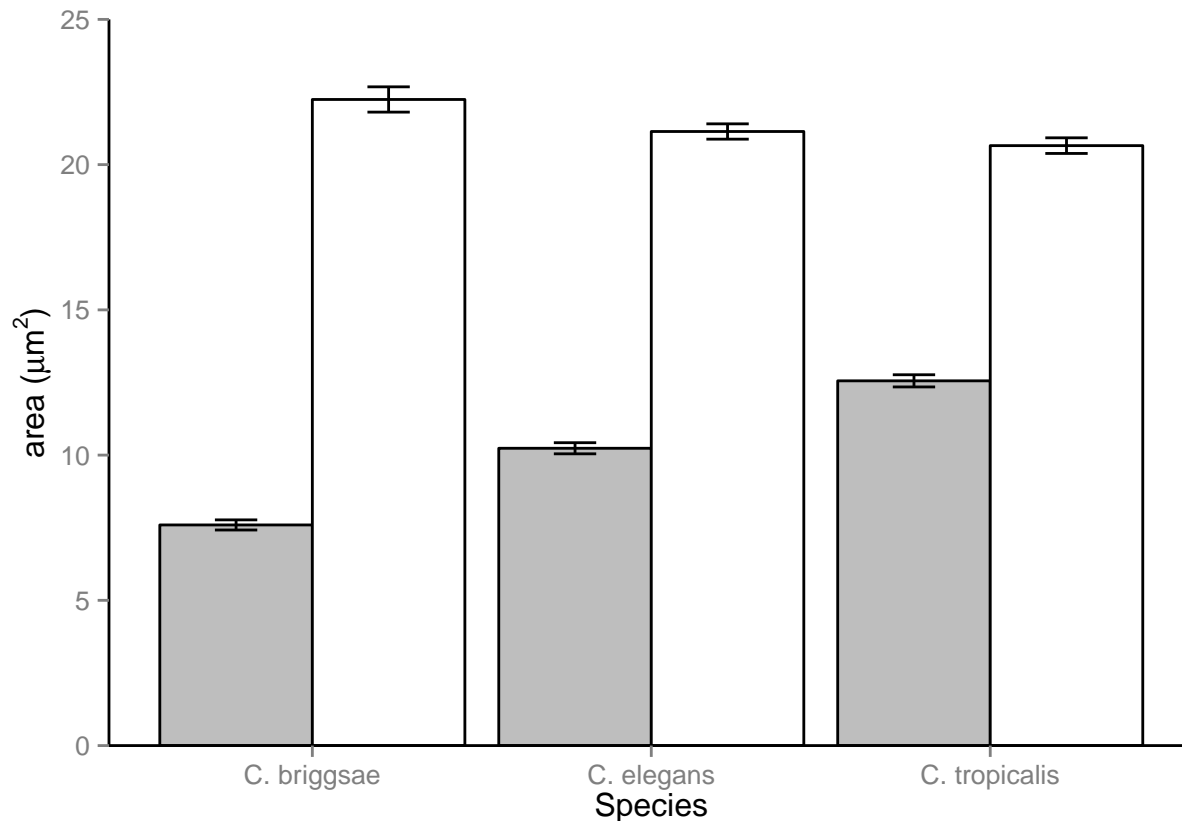
Figure mockup b group to get the data I want it's ugly, but it works

```
groupedstrains=ff %>% group_by(Species,Strain) %>%
    do(data.frame(meanM=mean(.[which(.$sex=="M"),]$area,na.rm=T),meanH=mean(.[which(.$sex=="H"),]$area,
                  sdM=sd(.[which(.$sex=="M"),]$area,na.rm=T),sdH=sd(.[which(.$sex=="H"),]$area,na.rm=T)
groupedstrains$sem=1.96*groupedstrains$sdM/sqrt(groupedstrains$nM)
groupedstrains$seh=1.96*groupedstrains$sdH/sqrt(groupedstrains$nH)
p=ggplot(data=groupedstrains, aes(x=meanM, y=meanH, colour=Species, ymin = meanH - seh,ymax=meanH + seh
p+geom_point(size=5)+coord_cartesian(ylim=c(0,15),xlim=c(0,33))+theme(panel.grid.major = element_blank(
    panel.background = element_blank(),axis.line = element_line(colour = "black"),legend.position = "n
    ylab(expression(paste("Hermaphrodite Sperm area (", mu, m^{2},")")))+xlab(expression(paste("Male Sp
    geom_smooth(aes(group=Species), method="lm",size=1.5,colour="black")
```
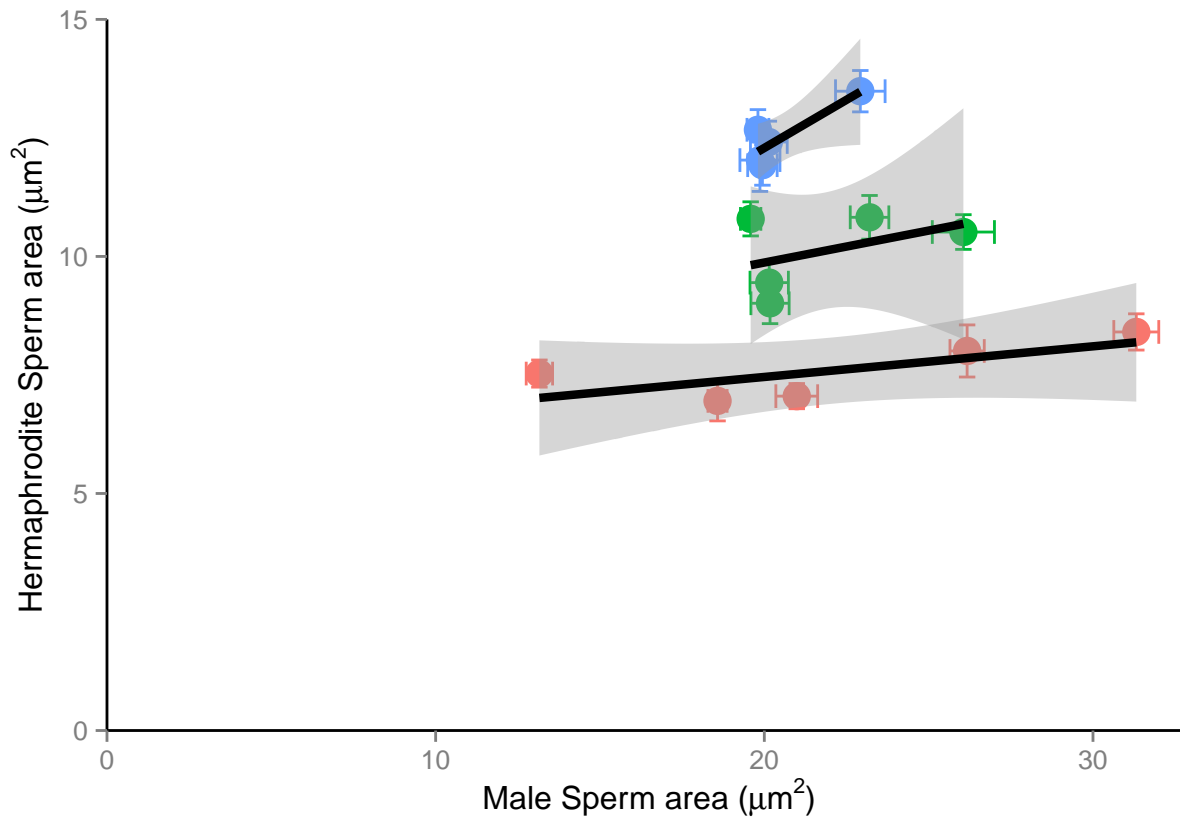
figure1a - bars need to rerun the above

```
ff<-spermsizes<-read.table("C:/Users/jeremy/Desktop/geiger/compileddata.csv",header=T,sep=",")
ff<-renameworms(ff)
spermsize2<-ff %>% group_by(Species,Strain) %>% summarise(means=mean(area,na.rm=T),cv=sd(area,na.rm=T)/m
countspec=function(dataframe){
    holding=c()
    for(i in 1:nrow(dataframe)){
        holding=c(holding,sum(dataframe$Species==dataframe[i,]$Species))
    }
    return(holding)
}
spermsize2$Species<-ordered(spermsize2$Species,levels=c(tr$tip.label))
spermsize2<-spermsize2[order(spermsize2$Species),]
spermsize2$Strain<-factor(spermsize2$Strain, levels = spermsize2$Strain)
spermsize2$specnum=countspec(spermsize2)
spermsize2$cols=cut(spermsize2$means,3)
findchange<-function(df){
    holding=c()
    for(i in 1:length(df$Species)){
        if(i==1){
            holding=c(holding,0)
        } else if(df$Species[i]==df$Species[i-1]){
            holding=c(holding,0)
        } else {
            holding=c(holding,1)
```

```
        }
    }
    return(holding)
}

spermsize2$change<-findchange(spermsize2)
w<-(1-(spermsize2$specnum-1)*0.05)/spermsize2$specnum
pos <- 0.5 * (cumsum(w) + cumsum(c(0, w[-length(w)])))

gapsizes<-function(df,pos){
    pos1<-pos
    for(i in 1:length(df$specnum)){
        if(df$specnum[i]==1){
            pos1[i:length(pos1)]<-pos1[i:length(pos1)]+0.1
        } else if(df$change[i]==1){
            pos1[i:length(pos1)]<-pos1[i:length(pos1)]+0.1
        } else {
            pos1[i:length(pos1)]<-pos1[i:length(pos1)]+0.05
        }
    }
    return(pos1)
}

pos<-gapsizes(spermsize2,pos)
ggplot(data=spermsize2,aes(x = pos, width = w, y = means, fill=cols,ymin=means-sem,ymax=means+sem)) + s
    geom_bar(stat = "identity",colour="black") + scale_x_continuous(labels = spermsize2$Strain, breaks
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
     panel.background = element_blank(),axis.line = element_line(colour = "black"),legend.position = "n
```

Warning: position_stack requires constant width: output may be incorrect
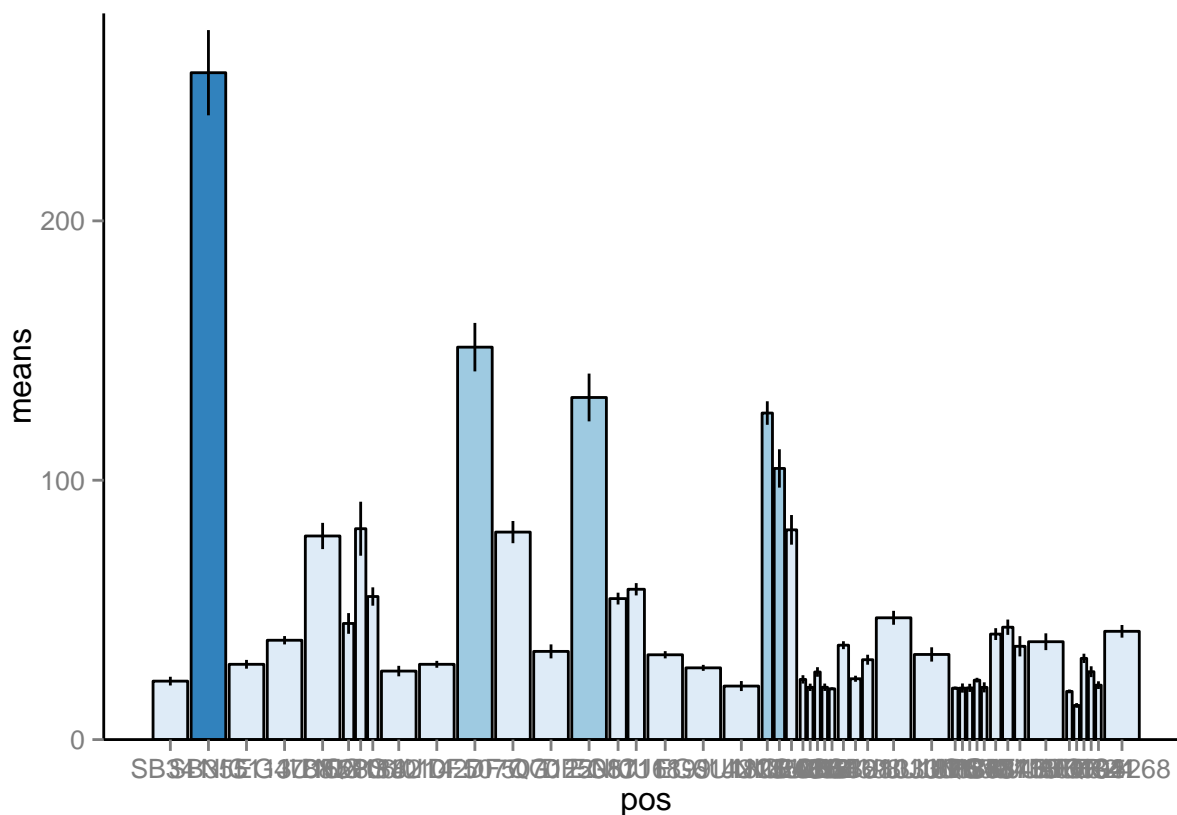
large tree first group all the data

```
spermsize1<-ff %>% group_by(Species) %>% summarise(means=mean(area,na.rm=T))
spermsize2<-ff %>% group_by(Species,Strain) %>% summarise(means=mean(area,na.rm=T))
spermsize3<-ff %>% group_by(Species,Strain,individual) %>% summarise(means=mean(area,na.rm=T))
```

function to fix the format of spermsizes, and fit tree

```
fitace <- function(spermsizes,tree){
  holding<-spermsizes$means
  names(holding)<-spermsizes$Species
  return(fastAnc(tree,holding,CI=T))
}
```

function to choose a subset - works on any level, just chooses one from each species

```
subsetter<-function(spermsizes){
  x=lapply(split(spermsizes, spermsizes$Species),function(subdf) subdf[sample(1:nrow(subdf), 1),])
  do.call("rbind",x)
}
```

example usage fitace(subsetter(spermsizes2),tr) now need to loop and keep the values

```
looper<-function(spermsizes,reps,tree){
  outputdata=data.frame()
  outputCI=data.frame()
  for(i in 1:reps){
    outputdata=rbind(outputdata,fitace(subsetter(spermsizes),tree)$ace)
    outputCI=rbind(outputCI,fitace(subsetter(spermsizes),tree)$CI95)
  }
  names(outputdata)=seq(from=27,to=51)
  return(c(outputdata,outputCI))
}
```

ok we can call eg x=looper(spermsize3,10,tr) then mean and sd, etc. etc.

```
x<-fitace(spermsize1,tr)
AncSperm=x$ace
#or AncSperm=apply(x,2,mean) if it's bootstrapped
spermsize1$Species <- factor(spermsize1$Species, levels = c("C. sp. 1","C. plicata","C. guadeloupensis"
spermsize=spermsize1[order(spermsize1$Species),]$means
names(spermsize)<-spermsize1[order(spermsize1$Species),]$Species
plotfancytree<-function(tree,nodesizes,terminalsizes){
  nodesizes<<-nodesizes
  terminalsizes<<-terminalsizes
  print(ggtree(tree)+geom_text(subset=.(isTip),aes(label=label),hjust=-0.19,fontface="italic")+
    geom_point(subset=.(!isTip),color="#1b9e77", size=sqrt(nodesizes/pi))+
    geom_point(subset=.(isTip),color="#7570b3", size=sqrt(terminalsizes/pi))+
    geom_text(subset=.(isTip),aes(label=round(terminalsizes,1)), hjust=1, vjust=-0.4, size=3)+
    geom_text(subset=.(!isTip),aes(label=round(nodesizes,1)), hjust=1.5, vjust=-0.4, size=3)+
    scale_x_continuous(expand = c(0.1, 0.1)))
  rm(list=c("terminalsizes","nodesizes"),pos=".GlobalEnv")
}
plotfancytree(tr,AncSperm,spermsize)
```
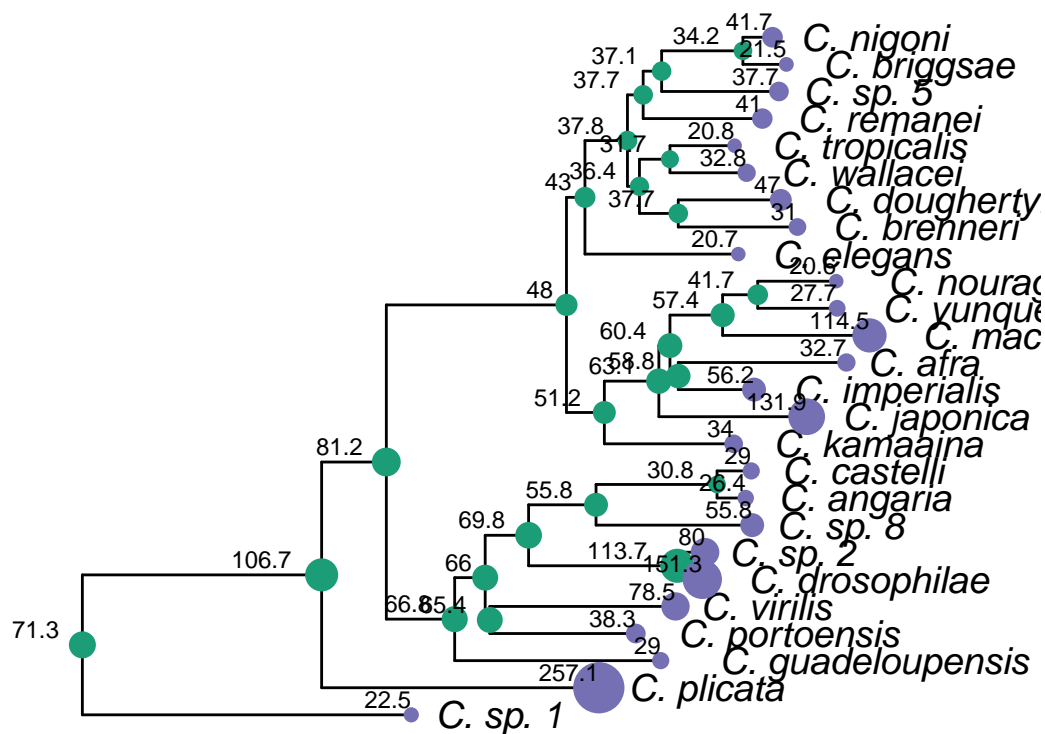
surface analysis

```r
library(surface)
```

```
## Warning: package 'surface' was built under R version 3.1.3
```

```
## Loading required package: ouch
```

```
## Warning: package 'ouch' was built under R version 3.1.3
```

```
## Loading required package: subplex
```

```r
tr<-read.tree(text="(C. sp. 1:0.356,(C. plicata:0.3,((C. guadeloupensis:0.223,((C. portoensis:0.158,C. v

tr$tip.label<-c("C. sp. 1","C. plicata","C. guadeloupensis","C. portoensis","C. virilis","C. sp. 8","C.
tr<-nameNodes(tr)
spermsizes<-read.table("C:/Users/jeremy/Desktop/geiger/compileddata.csv",header=T,sep=",")
spermsizes<-renameworms(spermsizes)
clusterer<-function(df){
  x<-df %>% group_by(Species) %>%
    summarise(means=mean(area,na.rm=T),cv=sd(area,na.rm=T)/mean(area,na.rm=T))
  rownames(x) <- x$Species
  x
}
```

```r
surfacesperm<-clusterer(spermsizes)

rownames(surfacesperm)<-surfacesperm$Species
surfacesperm <- surfacesperm[match(tr$tip.label,rownames(surfacesperm)),]
valuesin<-surfacesperm$means
names(valuesin)<-surfacesperm$Species
z<-runSurface(nameNodes(tr),as.data.frame(valuesin))

runonesim<-function(tr,fit,shifts){
  runSurface(tr,surfaceSimulate(tr, type="hansen-fit", hansenfit=fit,
                   shifts=shifts, sample_optima=TRUE)$dat)
}


getvals<-function(tr,fit,shifts){
  sim<-runonesim(tr,fit,shifts)
  lengthofsim<-length(sim$bwd)
  sim$bwd[[lengthofsim]]$n_regimes
}

distributionofvals<-function(n,tr,fit,shifts){
  rdply(n,getvals(tr,fit,shifts),.progress = "text")
}

#increase/run this to bootstap - for now it takes ~30 secs per run
#distributionofvals(2,tr,z$bwd[[2]]$fit,z$bwd[[2]]$savedshifts)
```