PCA in diploids

Zebin Zhang

7/2/2021

```
library(ggplot2)
library(grid)
library(ggpubr)
```

Loading libraries

```
setwd("/Users/zebinzhang/Desktop/My_Computer/Uppsala_University/Capsella_Project")
# This file contains reads counts infromation of each gene in all samples
rawTMM <- read.table("InputData/Diploids_individual_TMM_FLR.txt", header=TRUE, row.names=1)
head(rawTMM)[1:5]
```

Set work directly and load files.

```
##                       CR1_F     CR2_F     CR3_F     CR4_F    CG1_F
## Carubv10000018m.g 4275.497 5362.8262 4307.0971 3454.9034 7684.806
## Carubv10000019m.g  357.008  391.8685  348.1086  369.9587 1019.473
## Carubv10000020m.g 3025.966 2636.1311 2675.1350 2537.0918 3325.447
## Carubv10000021m.g 5897.512 6217.7578 6132.0495 4990.0279 6097.419
## Carubv10000022m.g 5522.177 4889.0774 5237.4588 4539.4333 5387.324
## Carubv10000023m.g 5517.350 4844.7843 5467.3307 4674.4514 5601.544
```

```
dim(rawTMM)
```

```
## [1] 17307    36
```

```
# This file contails informations of population and tissue in each accession
Accession <- read.table("InputData/DiploidsPhenotypicFile.txt", header = T)
head(Accession)
```

```
##   accession tissue species   MateType NewName
## 1   75.13_F flower      CR     Selfer   CR1_F
## 2   79.17_F flower      CR     Selfer   CR2_F
## 3    81.1_F flower      CR     Selfer   CR3_F
## 4    83.4_F flower      CR     Selfer   CR4_F
## 5    85.3_F flower      CG Outcrosser   CG1_F
## 6   86.12_F flower      CG Outcrosser   CG2_F
```

```
CR_F <- c(Accession$species=="CR" & Accession$tissue=="flower")
CG_F <- c(Accession$species=="CG" & Accession$tissue=="flower")
CO_F <- c(Accession$species=="CO" & Accession$tissue=="flower")

CR_L <- c(Accession$species=="CR" & Accession$tissue=="leaf")
CG_L <- c(Accession$species=="CG" & Accession$tissue=="leaf")
CO_L <- c(Accession$species=="CO" & Accession$tissue=="leaf")

CR_R <- c(Accession$species=="CR" & Accession$tissue=="root")
CG_R <- c(Accession$species=="CG" & Accession$tissue=="root")
CO_R <- c(Accession$species=="CO" & Accession$tissue=="root")
```

Define species

```
dim(rawTMM)
```

remove genes with 0 expression value in any individual

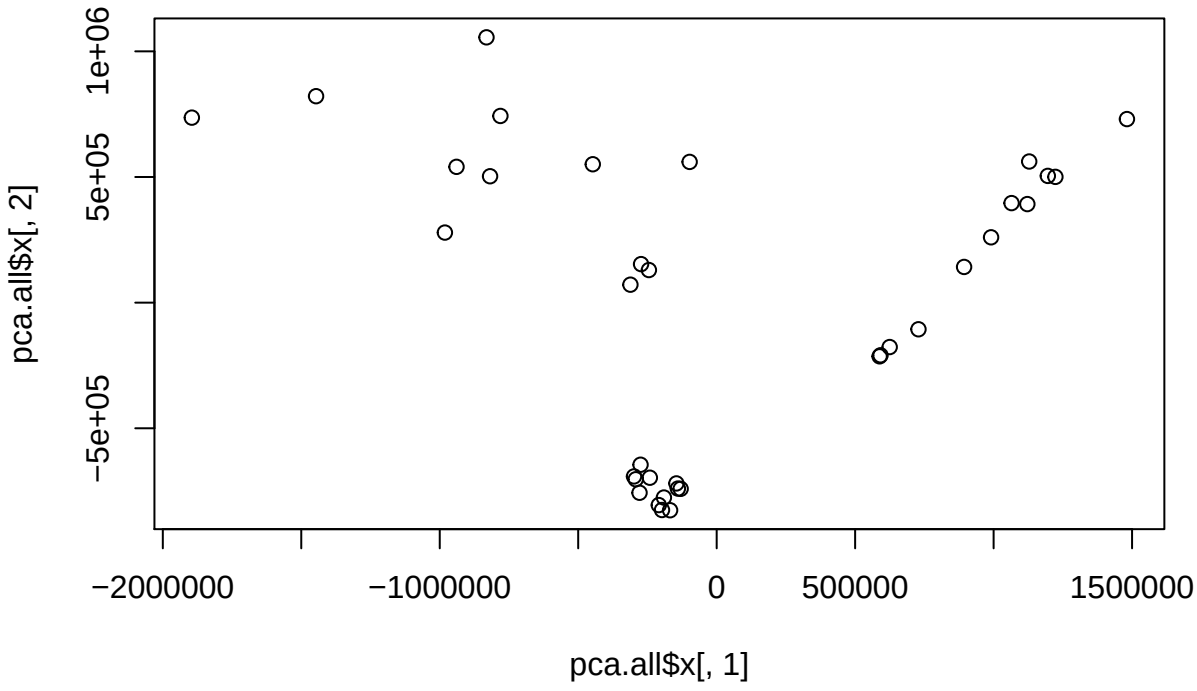
```
## [1] 17307    36
```

```
TMM <- na.omit(rawTMM)
dim(TMM)
```

```
## [1] 17307    36
```

PCA in all tissues: Flower, Leaf, Root Perform PCA

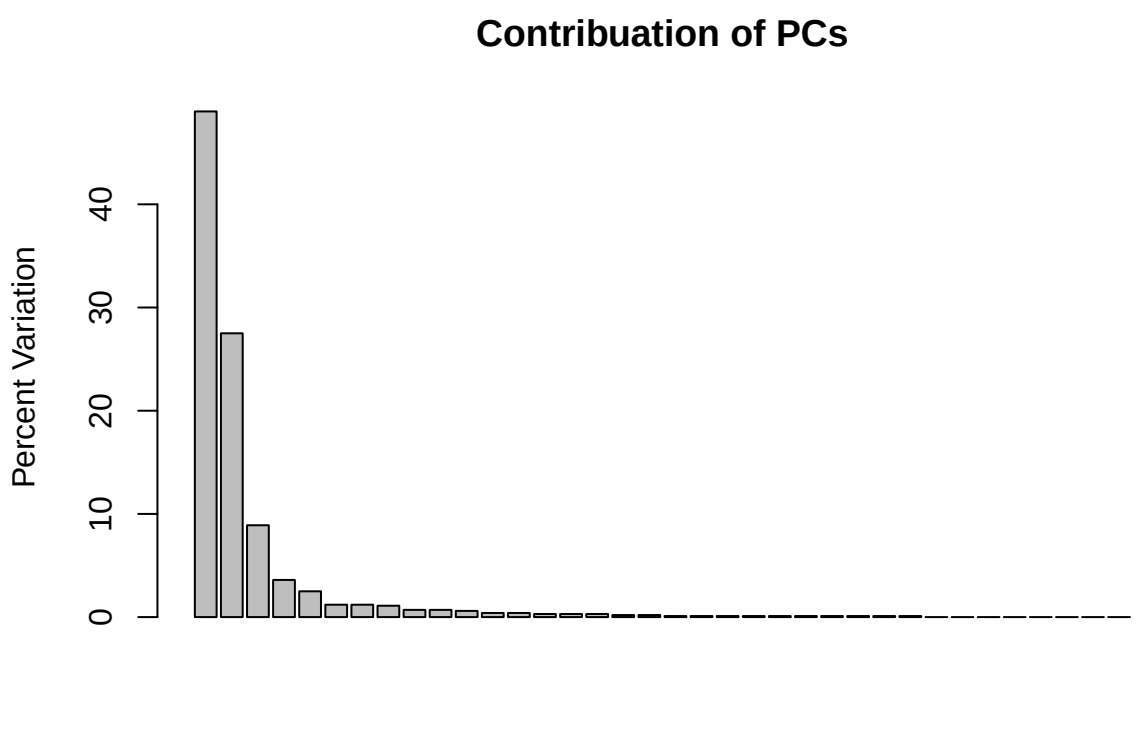
```
pca.all <- prcomp(t(TMM))
# NOTE: By default, prcomp() expects the samples to be rows and the genes to be columns
# So the data have to be transpose the matrix using the t() function
# If don't transpose the matrix, you will ultimately get a graph that shows how the genes are related t
# prcomp() returns thress things:
# 1) x -- x contains the principal components (PCs) for drawing a graph.
## plot PC1 and PC2
plot(pca.all$x[,1],pca.all$x[,2])
```




```
# 2) sdev -- "standard deviation", to calculate how much variation in the original data each PC account
## how much variation in the original data PC1 accounts for.
pca.all.var <- pca.all$sdev^2
head(pca.all.var)
```

```
## [1] 649042747103 363638260151 118449733159  47121684231  32665486503
## [6]  16427751269
```

```
## Since the percentage of variation that each PC accounts for is way more interesting than the acutal
## we calculate the percentage..
pca.all.var.per <- round(pca.all.var/sum(pca.all.var)*100,1)
## Plotting the percentage is easy with barplot()
barplot(pca.all.var.per, main = "Contribuation of PCs", xlab = "Pricipal Component", ylab = "Percent Va
```




```
# 3) rotation -- the loading scores rotation
```

```
dim(pca.all$x)[1] # check how many PCs in total
```

Format the data the way of ggplot2 needs

```
## [1] 36
```

```
pca.all.data <- data.frame(ID = 1:dim(pca.all$x)[1]) # create a new dataframe

for (i in 1:dim(pca.all$x)[1]){ # name each PCs
  pcs <- paste0("PC", i)
  pca.all.data[[pcs]] <- pca.all$x[,i]
}

# Set ID, Tissue, species, and matingSystem for ggplot
pca.all.data$ID <- row.names(pca.all.data)
pca.all.data$Tissue <- c(rep("Flower",12), rep("Leaf",12), rep("Root",12))
pca.all.data$Species <- rep(c(rep("CR",4), rep("CG",4), rep("CO", 4)), 3)
pca.all.data$Species <- factor(pca.all.data$Species, levels = c("CG","CR","CO"))
pca.all.data$MatingSystem <- rep(c(rep("Selfer",4), rep("Outcrosser",4), rep("Selfer", 4)), 3)
```

```
ZscoreAll <- pca.all.data
head(ZscoreAll)[1:7]
```

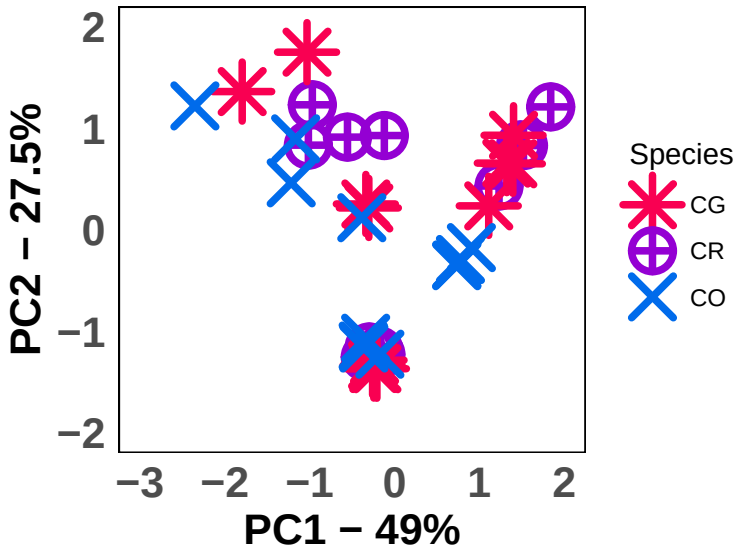
```
##   ID       PC1       PC2       PC3       PC4         PC5        PC6
## 1  1 -278544.1 -756548.9  94069.71 117903.16  -81764.805   2940.939
## 2  2 -145390.7 -719431.4 -37627.00  39356.61   20915.336   2026.887
## 3  3 -130108.3 -740925.1 -19649.46  51476.50   -3602.299 -31941.098
## 4  4 -241297.2 -696687.0  44972.37  42775.38    6835.263 -40415.846
## 5  5 -190536.9 -775558.3  30242.91 162820.18 -161845.001 220102.247
## 6  6 -197479.0 -825155.7  33887.92 184174.11 -221366.092 203952.704
```

```
for(i in 2:13){
  ZscoreAll[,i] <- scale(ZscoreAll[,i])
}
head(ZscoreAll)[1:7]
```

```
##   ID        PC1       PC2         PC3       PC4         PC5         PC6
## 1  1 -0.3457461 -1.254591  0.27332715 0.5431440 -0.45239921  0.02294547
## 2  2 -0.1804679 -1.193039 -0.10932830 0.1813039  0.11572316  0.01581395
## 3  3 -0.1614984 -1.228682 -0.05709310 0.2371366 -0.01993128 -0.24920732
## 4  4 -0.2995130 -1.155322  0.13067086 0.1970532  0.03781906 -0.31532805
## 5  5 -0.2365061 -1.286115  0.08787323 0.7500631 -0.89547760  1.71725748
## 6  6 -0.2451231 -1.368362  0.09846410 0.8484342 -1.22480382  1.59125730
```

```
ggplot(ZscoreAll, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(plot.title = element_text(size=16,face="bold"))
```

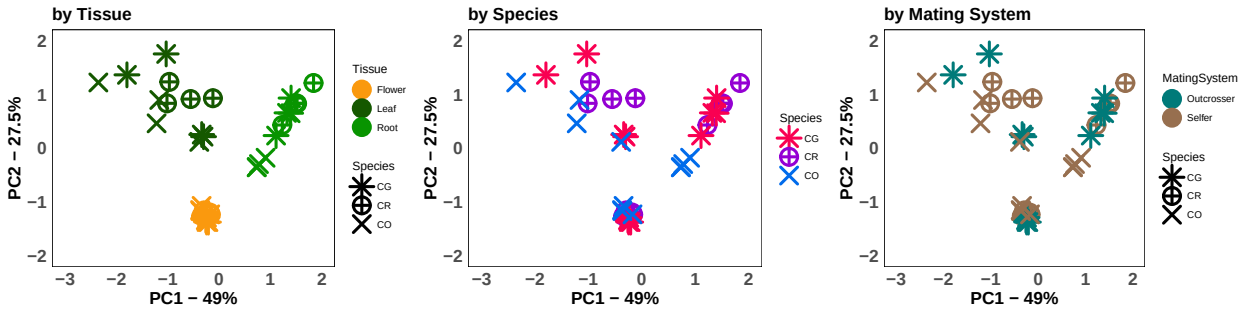
Output raw PCA plot




```
# Tissue
p1 <- ggplot(ZscoreAll, aes(PC1, PC2, col = Tissue, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#fa990e","#155800","#089400"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  # stat_ellipse(geom="polygon", aes(fill = tissue), # add frame
  #              alpha = 0.05,
  #              show.legend = FALSE,
  #              level = 0.90) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  #theme(legend.position = "none") +
  ggtitle("by Tissue") +
  theme(plot.title = element_text(size=16,face="bold"))

# Species
p2 <- ggplot(ZscoreAll, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
```

```
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  # stat_ellipse(geom="polygon", aes(fill = tissue), # add frame
  #              alpha = 0.05,
  #              show.legend = FALSE,
  #              level = 0.90) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  #theme(legend.position = "none") +
  ggtitle("by Species") +
  theme(plot.title = element_text(size=16,face="bold"))
# Mating System
p3 <- ggplot(ZscoreAll, aes(PC1, PC2, col = MatingSystem, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#007a79","#976f4f"))  +
  #scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  #theme(legend.position = "none") +
  ggtitle("by Mating System") +
  theme(plot.title = element_text(size=16,face="bold"))

grid.newpage()
ggarrange(p1, p2, p3, ncol=3, nrow=1) # set the common legend
```

We also want to know how this plot looks like when color by tissue, by species, and by mating system



Causing all dots of flowers crowded together, I have to add zoom in plot for flower The PCA analysis only for flower is need to run, where all steps are typically same as above mentioned but in different input data

```
## PCA in Flowers
### Define flower
setwd("/Users/zebinzhang/Desktop/My_Computer/Uppsala_University/Capsella_Project")
TMM.Flower <- read.table("InputData/TMM_Diploids_F.txt", header = T, sep = "\t")
head(TMM.Flower)
```

```
##                      CG1_F    CG2_F     CG3_F     CG4_F    CR1_F     CR2_F
## Carubv10000018m.g 7684.806 6298.346 5793.3873 6236.9802 4275.497 5362.8262
## Carubv10000019m.g 1019.473 1690.816  401.1901  991.6462  357.008  391.8685
## Carubv10000020m.g 3325.447 2731.951 2855.3056 3079.5167 3025.966 2636.1311
## Carubv10000021m.g 6097.419 5552.219 5613.8344 6340.5385 5897.512 6217.7578
## Carubv10000022m.g 5387.324 5080.437 5374.0826 5583.7381 5522.177 4889.0774
## Carubv10000023m.g 5601.544 5348.137 5554.6822 5333.5295 5517.350 4844.7843
##                       CR3_F     CR4_F      CO1_F     CO2_F     CO3_F     CO4_F
## Carubv10000018m.g 4307.0971 3454.9034 2968.84893 3585.6948 3102.7761 3045.1453
## Carubv10000019m.g  348.1086  369.9587   89.85919  289.2978  256.7454  268.3176
## Carubv10000020m.g 2675.1350 2537.0918 2212.86192 2552.3922 2331.6557 2050.6528
## Carubv10000021m.g 6132.0495 4990.0279 7223.50950 7483.9381 6995.9217 6537.8976
## Carubv10000022m.g 5237.4588 4539.4333 4020.72119 4467.3627 3717.1696 3824.5072
## Carubv10000023m.g 5467.3307 4674.4514 4440.56247 4828.6397 4427.7404 3767.8463
```

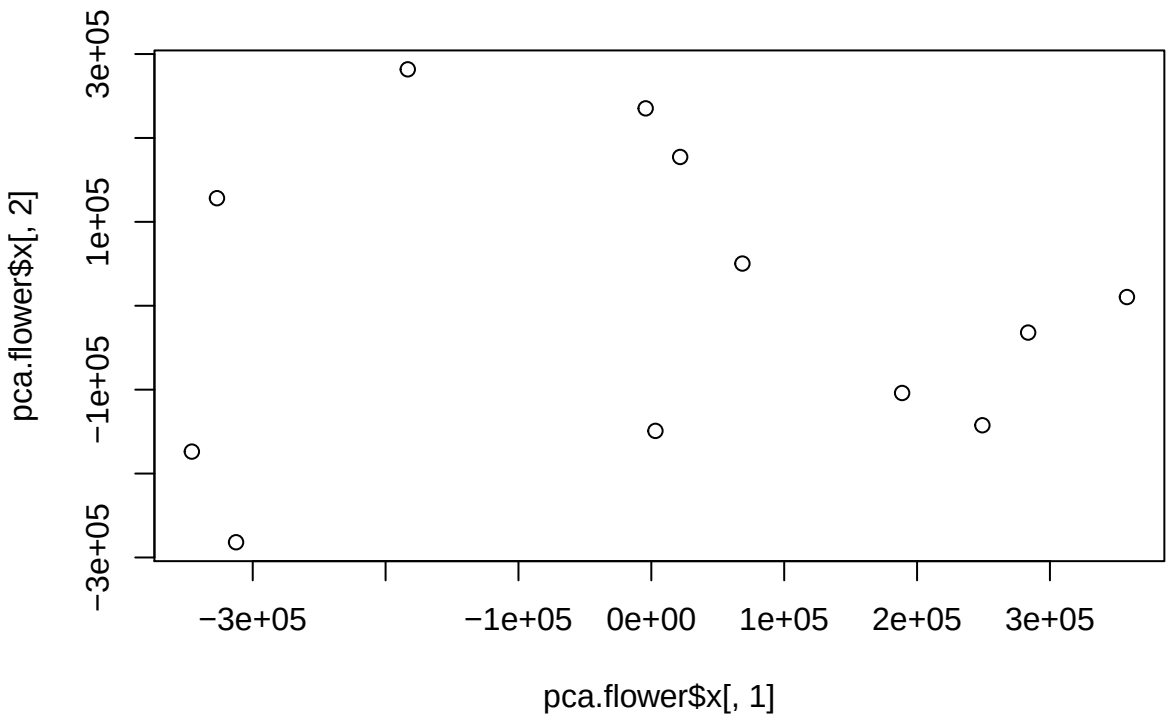
```
dim(TMM.Flower)
```

```
## [1] 17307    12
```

```
dim(na.omit(TMM.Flower))
```

```
## [1] 17307    12
```

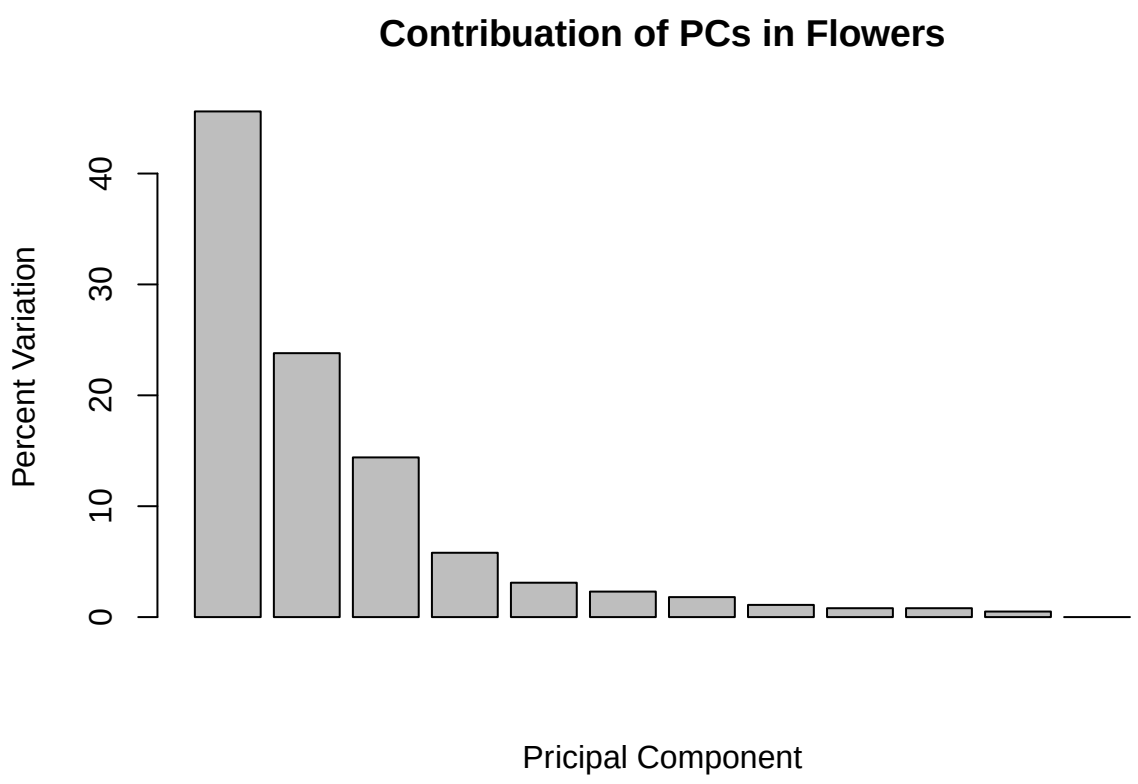
```
### perform PCA in Flower
pca.flower <- prcomp(t(na.omit(TMM.Flower)))
# 1) X
## plot PC1 and PC2
plot(pca.flower$x[,1],pca.flower$x[,2])
```




```
# 2) sdev
pca.flower.var <- pca.flower$sdev^2
head(pca.flower.var)
```

```
## [1] 60837790941 31752477312 19214354990  7725694762  4081414720  3116208228
```

```
## calculate the percentage..
pca.flower.var.per <- round(pca.flower.var/sum(pca.flower.var)*100,1)
## Plotting the percentage is easy with barplot()
barplot(pca.flower.var.per, main = "Contribuation of PCs in Flowers", xlab = "Pricipal Component", ylab
```




```
# 3) rotation -- the loading scores rotation
# Format the data the way ggplot2 likes it
dim(pca.flower$x)[1]
```

```
## [1] 12
```

```
pca.flower.data <- data.frame(ID = 1:dim(pca.flower$x)[1])
for (i in 1:dim(pca.flower$x)[1]){
  pcs <- paste0("PC", i)
  pca.flower.data[[pcs]] <- pca.flower$x[,i]
}

pca.flower.data$ID <- row.names(pca.flower$x)

pca.flower.data$ID <- row.names(pca.flower.data)
pca.flower.data$tissue <- "Flower"
pca.flower.data$Species <- c(rep("CG",4), rep("CR",4), rep("CO", 4))
pca.flower.data$Species <- factor(pca.flower.data$Species, levels = c("CG","CR","CO"))
#pca.flower.data$species <- c(c("CR","CG","CO"))

ZscoreFlower <- pca.flower.data
head(ZscoreFlower)
```

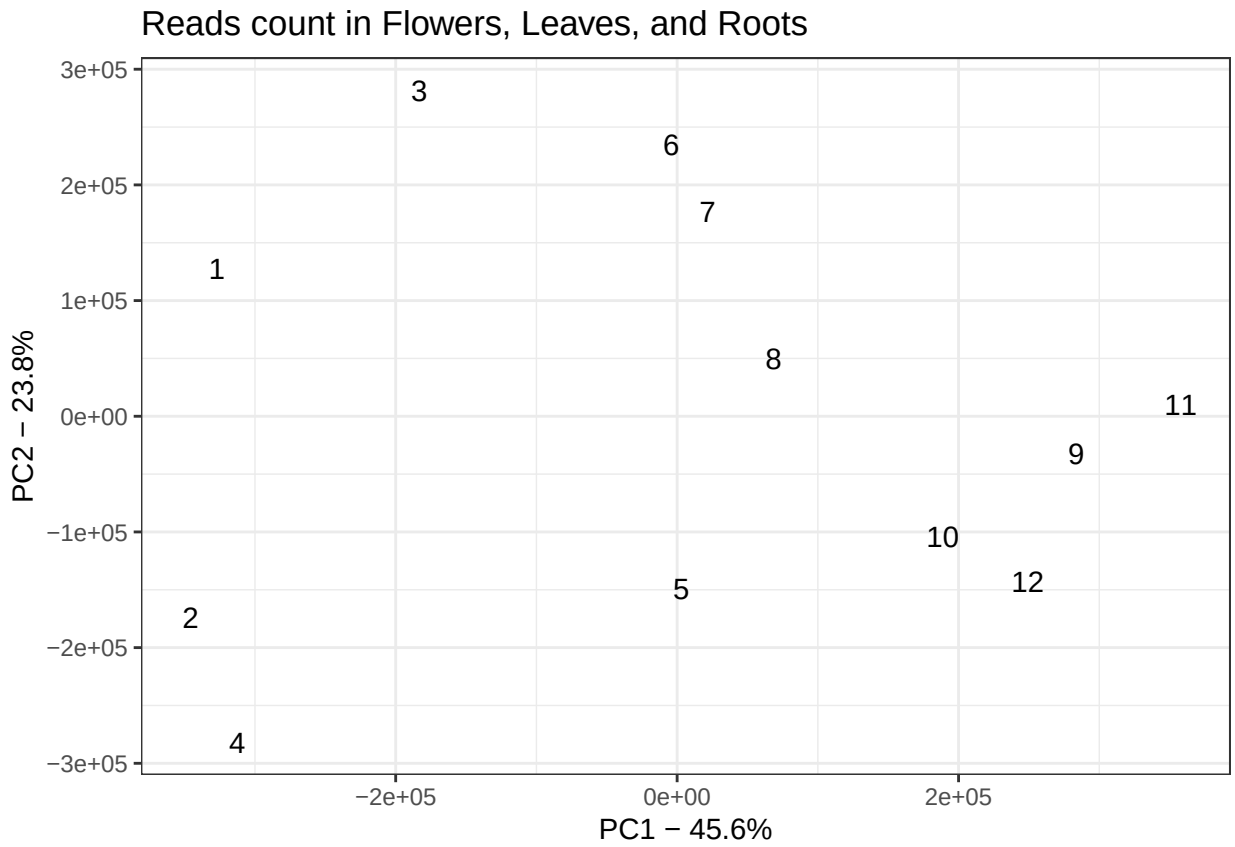
```
##   ID         PC1       PC2          PC3         PC4       PC5       PC6
## 1  1 -326902.058  128200.6  155351.0038    8347.959 -33883.23 -65092.43
```

```
## 2  2 -345819.621 -173820.0   62148.3607    6351.159 137084.95  49138.03
## 3  3 -183308.300  281754.4  108521.1338 -113651.996 -69034.07  37027.22
## 4  4 -312534.713 -281851.5    -878.2356   70210.075 -64291.80 -28275.65
## 5  5    3081.972 -149151.9 -224566.3383 -174966.191 -16428.40  31686.95
## 6  6   -4257.591  235252.3 -100864.0258   97625.278  61320.68  11853.72
##           PC7        PC8          PC9       PC10      PC11          PC12 tissue
## 1  -8207.9039  82616.243    -43.30106  13936.050 -5044.585 -1.625375e-09 Flower
## 2 -45283.8653 -14082.197 -17223.44109  17818.393  2786.605 -4.885407e-10 Flower
## 3 -23164.8764 -61672.001  -9213.33212  -9187.486 10439.164 -9.453557e-10 Flower
## 4  77513.4889 -36549.057  11864.09877 -22810.445  2737.705 -8.174623e-10 Flower
## 5   -935.6943  46954.285   4967.81207 -20801.146 17560.981  2.672199e-10 Flower
## 6 -10080.8005   4860.015  48589.98730 -58728.041 -7841.190  5.280437e-10 Flower
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

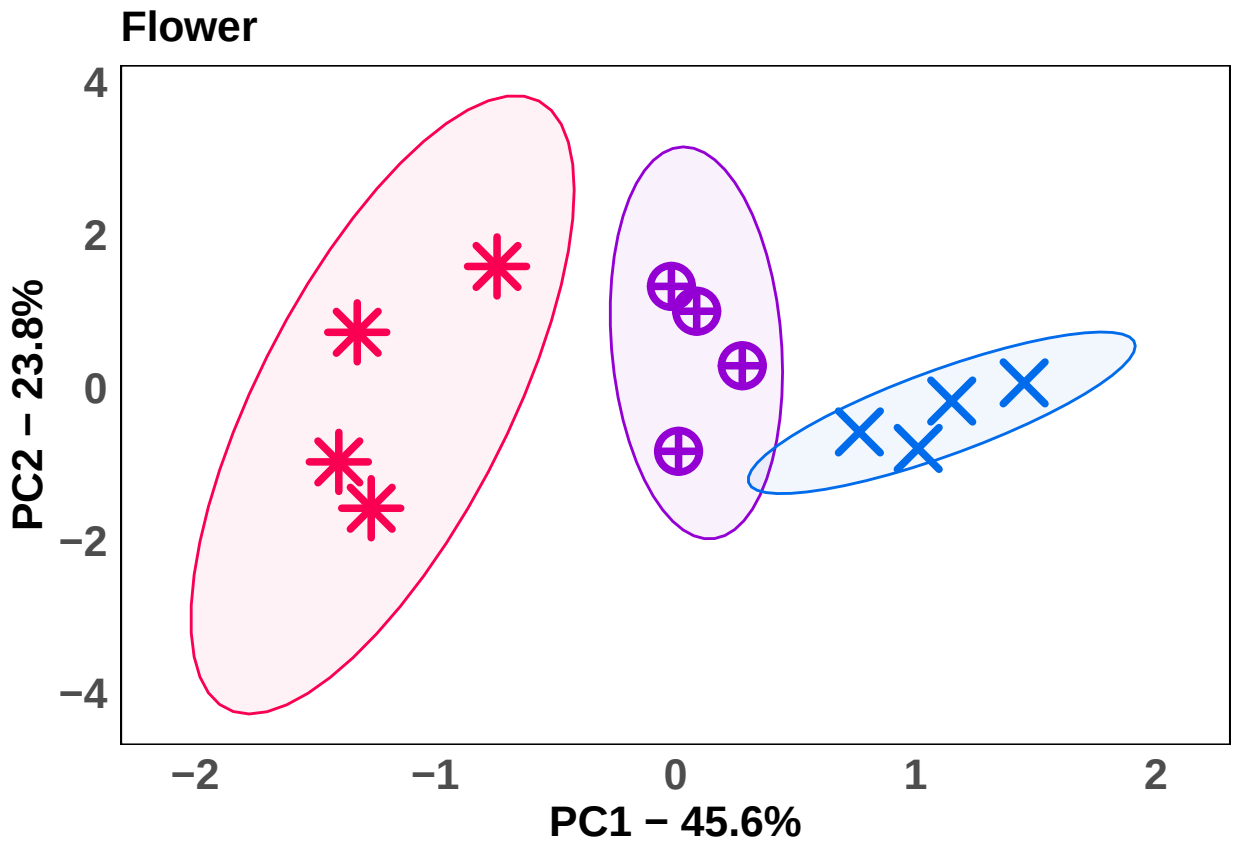
```
for(i in 2:13){
  ZscoreFlower[,i] <- scale(ZscoreFlower[,i])
}
head(ZscoreFlower)
```

```
##   ID         PC1        PC2          PC3         PC4        PC5        PC6
## 1  1 -1.32535108  0.7194510  1.120730417  0.09497549 -0.5303706 -1.1660503
## 2  2 -1.40204810 -0.9754631  0.448349586  0.07225771  2.1457762  0.8802470
## 3  3 -0.74318239  1.5811823  0.782891211 -1.29302909 -1.0805829  0.6632968
## 4  4 -1.26710190 -1.5817271 -0.006335752  0.79878641 -1.0063528 -0.5065232
## 5  5  0.01249516 -0.8370280 -1.620062437 -1.99060626 -0.2571520  0.5676324
## 6  6 -0.01726145  1.3202164 -0.727651440  1.11069165  0.9598461  0.2123446
##           PC7        PC8          PC9       PC10       PC11       PC12 tissue
## 1 -0.16804663  2.1562284 -0.001302233  0.4326155 -0.1960269 -1.8059882 Flower
## 2 -0.92713085 -0.3675359 -0.517976651  0.5531347  0.1082844 -0.5477301 Flower
## 3 -0.47427205 -1.6095977 -0.277081153 -0.2852062  0.4056543 -1.0533369 Flower
## 4  1.58699232 -0.9539058  0.356800138 -0.7081024  0.1063842 -0.9117834 Flower
## 5 -0.01915718  1.2254753  0.149401658 -0.6457280  0.6824002  0.2887523 Flower
## 6 -0.20639186  0.1268431  1.461292130 -1.8230889 -0.3046999  0.5774343 Flower
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```
# graph parttern 1
ggplot(data = pca.flower.data, aes(x=PC1, y=PC2, label=ID)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.flower.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.flower.var.per[2], "%", sep = "")) +
  theme_bw() +
  ggtitle("Reads count in Flowers, Leaves, and Roots")
```




```r
# graph parttern 2
# PC1 vs PC2
ggplot(ZscoreFlower, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.flower.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.flower.var.per[2], "%", sep = "")) +
  stat_ellipse(geom="polygon", aes(fill = Species), # add frame
               alpha = 0.05,
               show.legend = FALSE,
               level = 0.90) +
  scale_x_continuous(breaks=seq(-2, 2, 1), limits=c(-2.1, 2.1)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(legend.position = "none") +
  ggtitle("Flower") +
  theme(plot.title = element_text(size=16,face="bold"))
```



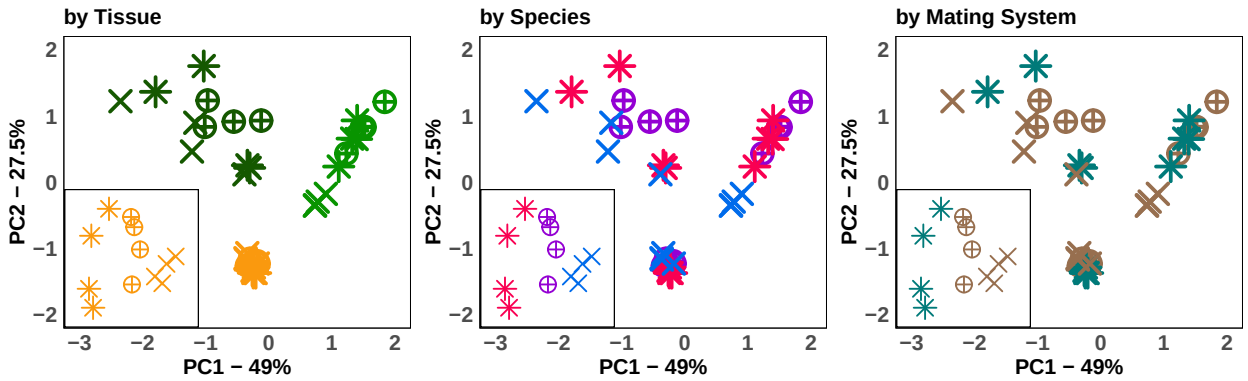

```
# by Tissue
p1 <- ggplot(ZscoreAll, aes(PC1, PC2, col = Tissue, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#fa990e","#155800","#089400"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(legend.position = "none") +
  ggtitle("by Tissue") +
  theme(plot.title = element_text(size=16,face="bold"))

p1_i <- ggplot(ZscoreFlower, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 5, stroke = 1, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#fa990e","#fa990e","#fa990e"))  +
```

```
  scale_fill_manual(values = c("#fa990e","#fa990e","#fa990e")) +
  xlim(-2,2) + ylim(-2,2)+
  #xlab("PC1") +
  #ylab("PC2") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(legend.position = "none") +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

p_tissue <- p1 + annotation_custom(ggplotGrob(p1_i), xmin = -3.41, xmax = -1,
                   ymin = -2.35, ymax = 0)

# Species
p2 <- ggplot(ZscoreAll, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  # stat_ellipse(geom="polygon", aes(fill = tissue), # add frame
  #              alpha = 0.05,
  #              show.legend = FALSE,
  #              level = 0.90) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(legend.position = "none") +
  ggtitle("by Species") +
  theme(plot.title = element_text(size=16,face="bold"))

p2_i <- ggplot(ZscoreFlower, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 5, stroke = 1, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  xlim(-2,2) + ylim(-2,2)+
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  #xlab("PC1") +
  #ylab("PC2") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
```

```
        panel.border = element_rect(fill= "transparent")) +
  theme(legend.position = "none") +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

p_species <- p2 + annotation_custom(ggplotGrob(p2_i), xmin = -3.41, xmax = -1,
                                    ymin = -2.35, ymax = 0)

# Mating System
p3 <- ggplot(ZscoreAll, aes(PC1, PC2, col = MatingSystem, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#007a79","#976f4f"))  +
  #scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.all.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.all.var.per[2], "%", sep = "")) +
  scale_x_continuous(breaks=seq(-3, 2, 1), limits=c(-3, 2)) +
  scale_y_continuous(breaks=seq(-2, 2, 1), limits=c(-2, 2)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(legend.position = "none") +
  ggtitle("by Mating System") +
  theme(plot.title = element_text(size=16,face="bold"))

p3_i <- ggplot(ZscoreFlower, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 5, stroke = 1, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#007a79","#976f4f","#976f4f"))  +
  scale_fill_manual(values = c("#007a79","#976f4f","#976f4f")) +
  xlim(-2,2) + ylim(-2,2)+
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  #xlab("PC1") +
  #ylab("PC2") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(legend.position = "none") +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

p_MST <- p3 + annotation_custom(ggplotGrob(p3_i), xmin = -3.41, xmax = -1,
```

```
                                    ymin = -2.35, ymax = 0)

grid.newpage()
ggarrange(p_tissue, p_species, p_MST, ncol=3, nrow=1) # set the common legend
```

Insert the flower plot into main figures.



PCA plot separate by tissues

Load data

```
setwd("/Users/zebinzhang/Desktop/My_Computer/Uppsala_University/Capsella_Project")
TMM.Flower <- read.table("InputData/TMM_Diploids_F.txt", header = T, sep = "\t")
TMM.Leaf <- read.table("InputData/TMM_Diploids_L.txt", header = T, sep = "\t")
TMM.Root <- read.table("InputData/TMM_Diploids_R.txt", header = T, sep = "\t")
```

```
head(TMM.Flower)[1:5]
```

PCA in Flowers

```
##                       CG1_F    CG2_F     CG3_F     CG4_F    CR1_F
## Carubv10000018m.g 7684.806 6298.346 5793.3873 6236.9802 4275.497
## Carubv10000019m.g 1019.473 1690.816  401.1901  991.6462  357.008
## Carubv10000020m.g 3325.447 2731.951 2855.3056 3079.5167 3025.966
## Carubv10000021m.g 6097.419 5552.219 5613.8344 6340.5385 5897.512
## Carubv10000022m.g 5387.324 5080.437 5374.0826 5583.7381 5522.177
## Carubv10000023m.g 5601.544 5348.137 5554.6822 5333.5295 5517.350
```

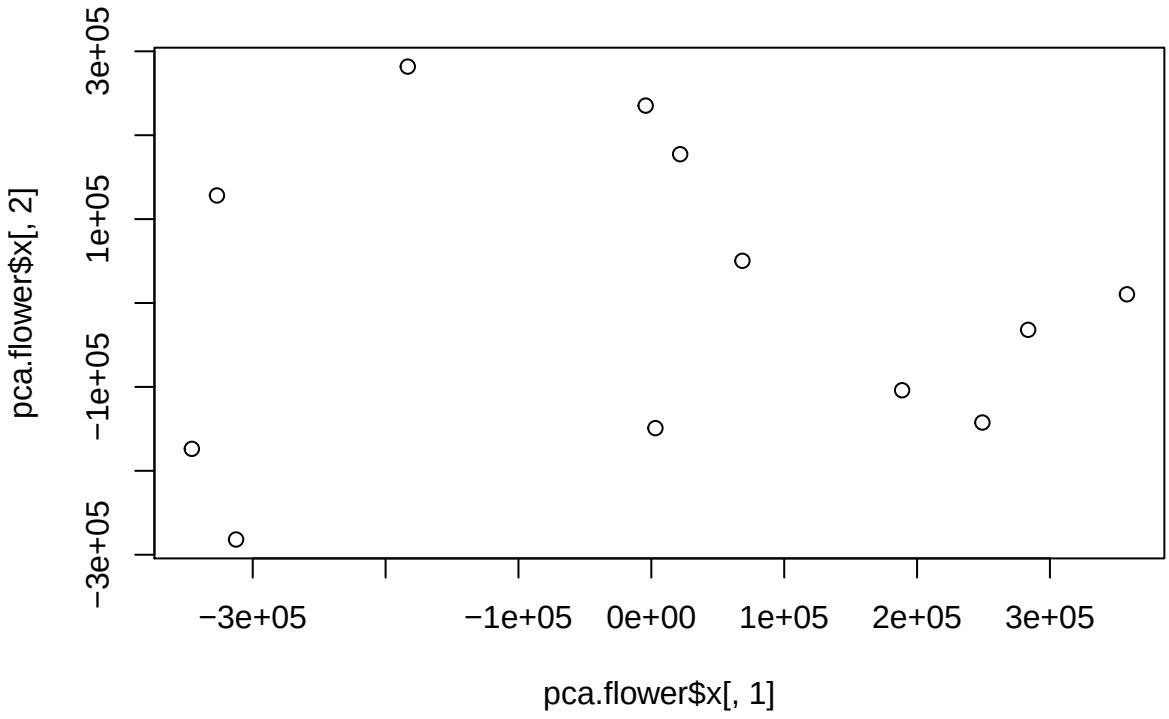
```
dim(TMM.Flower)
```

```
## [1] 17307    12
```

```r
dim(na.omit(TMM.Flower))
```

```
## [1] 17307    12
```

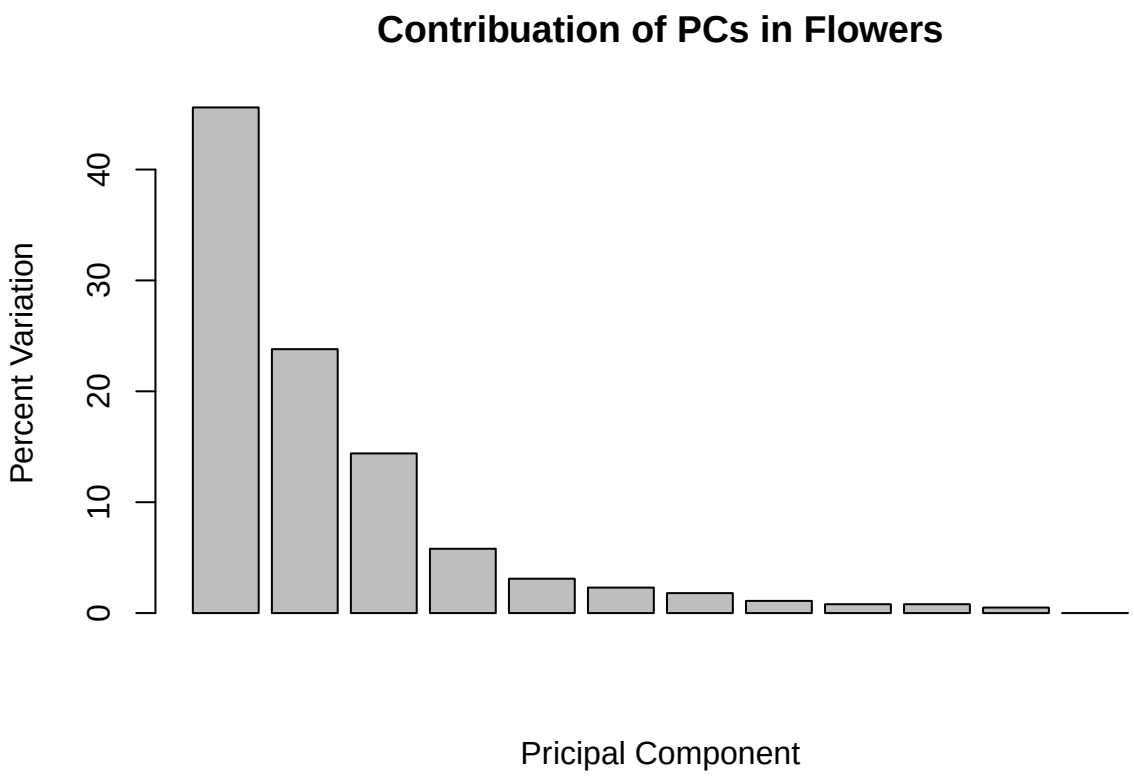
```r
### perform PCA in Flower
pca.flower <- prcomp(t(na.omit(TMM.Flower)))
# 1) X
## plot PC1 and PC2
plot(pca.flower$x[,1],pca.flower$x[,2])
```




```r
# 2) sdev
pca.flower.var <- pca.flower$sdev^2
head(pca.flower.var)
```

```
## [1] 60837790941 31752477312 19214354990  7725694762  4081414720  3116208228
```

```r
## calculate the percentage..
pca.flower.var.per <- round(pca.flower.var/sum(pca.flower.var)*100,1)
## Plotting the percentage is easy with barplot()
barplot(pca.flower.var.per, main = "Contribuation of PCs in Flowers", xlab = "Pricipal Component", ylab
```




```
# 3) rotation -- the loading scores rotation

### Draw graph in Flowers
# Format the data the way ggplot2 likes it
dim(pca.flower$x)[1]
```

```
## [1] 12
```

```
pca.flower.data <- data.frame(ID = 1:dim(pca.flower$x)[1])
for (i in 1:dim(pca.flower$x)[1]){
  pcs <- paste0("PC", i)
  pca.flower.data[[pcs]] <- pca.flower$x[,i]
}

pca.flower.data$ID <- row.names(pca.flower$x)

pca.flower.data$ID <- row.names(pca.flower.data)
pca.flower.data$tissue <- "Flower"
pca.flower.data$Species <- c(rep("CG",4), rep("CR",4), rep("CO", 4))
pca.flower.data$Species <- factor(pca.flower.data$Species, levels = c("CG","CR","CO"))
#pca.flower.data$species <- c(c("CR","CG","CO"))

ZscoreFlower <- pca.flower.data
head(ZscoreFlower)
```

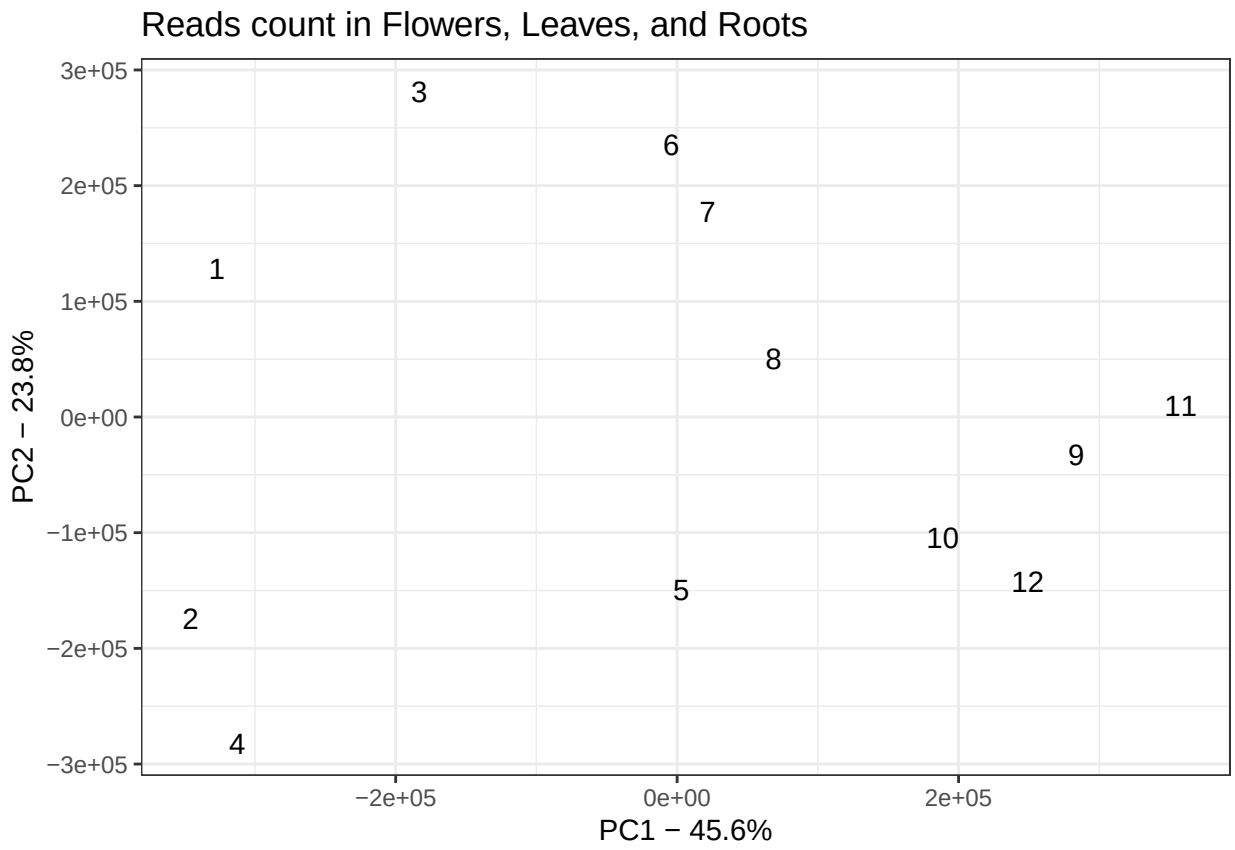
```
##   ID         PC1       PC2          PC3         PC4       PC5       PC6
## 1  1 -326902.058  128200.6  155351.0038    8347.959 -33883.23 -65092.43
## 2  2 -345819.621 -173820.0   62148.3607    6351.159 137084.95  49138.03
## 3  3 -183308.300  281754.4  108521.1338 -113651.996 -69034.07  37027.22
## 4  4 -312534.713 -281851.5    -878.2356   70210.075 -64291.80 -28275.65
## 5  5    3081.972 -149151.9 -224566.3383 -174966.191 -16428.40  31686.95
## 6  6   -4257.591  235252.3 -100864.0258   97625.278  61320.68  11853.72
##           PC7        PC8          PC9       PC10      PC11          PC12 tissue
## 1  -8207.9039  82616.243    -43.30106  13936.050 -5044.585 -1.625375e-09 Flower
## 2 -45283.8653 -14082.197 -17223.44109  17818.393  2786.605 -4.885407e-10 Flower
## 3 -23164.8764 -61672.001  -9213.33212  -9187.486 10439.164 -9.453557e-10 Flower
## 4  77513.4889 -36549.057  11864.09877 -22810.445  2737.705 -8.174623e-10 Flower
## 5   -935.6943  46954.285   4967.81207 -20801.146 17560.981  2.672199e-10 Flower
## 6 -10080.8005   4860.015  48589.98730 -58728.041 -7841.190  5.280437e-10 Flower
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```
for(i in 2:13){
  ZscoreFlower[,i] <- scale(ZscoreFlower[,i])
}
head(ZscoreFlower)
```

```
##   ID         PC1        PC2          PC3         PC4        PC5        PC6
## 1  1 -1.32535108  0.7194510  1.120730417  0.09497549 -0.5303706 -1.1660503
## 2  2 -1.40204810 -0.9754631  0.448349586  0.07225771  2.1457762  0.8802470
## 3  3 -0.74318239  1.5811823  0.782891211 -1.29302909 -1.0805829  0.6632968
## 4  4 -1.26710190 -1.5817271 -0.006335752  0.79878641 -1.0063528 -0.5065232
## 5  5  0.01249516 -0.8370280 -1.620062437 -1.99060626 -0.2571520  0.5676324
## 6  6 -0.01726145  1.3202164 -0.727651440  1.11069165  0.9598461  0.2123446
##           PC7        PC8          PC9       PC10       PC11       PC12 tissue
## 1 -0.16804663  2.1562284 -0.001302233  0.4326155 -0.1960269 -1.8059882 Flower
## 2 -0.92713085 -0.3675359 -0.517976651  0.5531347  0.1082844 -0.5477301 Flower
## 3 -0.47427205 -1.6095977 -0.277081153 -0.2852062  0.4056543 -1.0533369 Flower
## 4  1.58699232 -0.9539058  0.356800138 -0.7081024  0.1063842 -0.9117834 Flower
## 5 -0.01915718  1.2254753  0.149401658 -0.6457280  0.6824002  0.2887523 Flower
## 6 -0.20639186  0.1268431  1.461292130 -1.8230889 -0.3046999  0.5774343 Flower
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```
# graph parttern 1
ggplot(data = pca.flower.data, aes(x=PC1, y=PC2, label=ID)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.flower.var.per[1], "%", sep = "")) +
```

```
  ylab(paste("PC2 - ", pca.flower.var.per[2], "%", sep = "")) +
  theme_bw() +
  ggtitle("Reads count in Flowers, Leaves, and Roots")
```




```
# graph parttern 2
# PC1 vs PC2
p1 <- ggplot(ZscoreFlower, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  # stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.flower.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.flower.var.per[2], "%", sep = "")) +
  stat_ellipse(geom="polygon", aes(fill = Species), # add frame
               alpha = 0.05,
               show.legend = FALSE,
               level = 0.90) +
  scale_x_continuous(breaks=seq(-2, 2, 1), limits=c(-2.1, 2.1)) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(legend.position = "none") +
```

```
  ggtitle("Flower") +
  theme(plot.title = element_text(size=16,face="bold"))
```

```
################################
## PCA in Leaves
### Define leaf
head(TMM.Leaf)[1:5]
```

```
##                        CG1_L     CG2_L     CG3_L     CG4_L     CR1_L
## Carubv10000018m.g 5411.7256 8093.0227 9968.4895 8261.5091 5948.4891
## Carubv10000019m.g  318.1181  747.9534  339.2023  958.7714  182.9841
## Carubv10000020m.g 1943.3791 2520.3253 2816.4044 3657.3709 3404.2495
## Carubv10000021m.g 2959.6804 5365.9697 4522.8324 4894.9462 4799.9591
## Carubv10000022m.g 3701.1556 6682.6636 4724.8480 5168.4485 5436.1021
## Carubv10000023m.g 4053.9332 4449.8566 5459.5006 6262.2343 5360.7120
```

```
dim(TMM.Leaf)
```

```
## [1] 17307    12
```

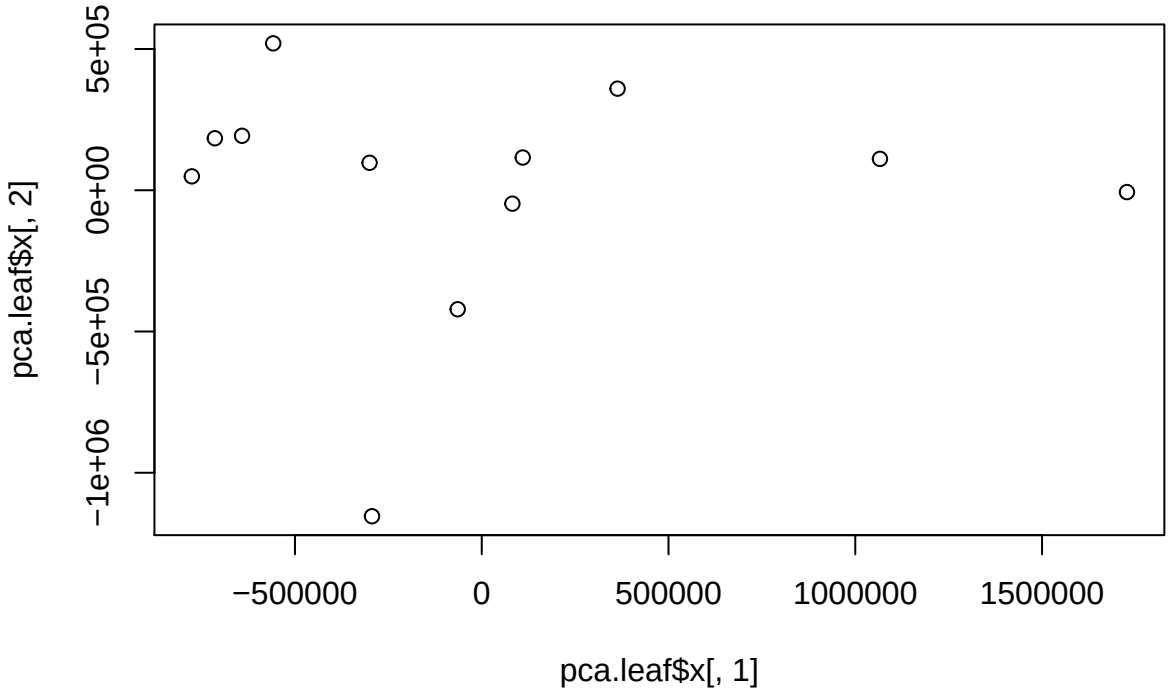
```
dim(na.omit(TMM.Leaf))
```

```
## [1] 17307    12
```

```
### perform PCA in Leaf
pca.leaf <- prcomp(t(na.omit(TMM.Leaf)))
summary(pca.leaf)
```

```
## Importance of components:
##                             PC1       PC2       PC3       PC4       PC5
## Standard deviation     7.561e+05 4.285e+05 3.107e+05 2.216e+05 1.622e+05
## Proportion of Variance 5.838e-01 1.875e-01 9.855e-02 5.015e-02 2.687e-02
## Cumulative Proportion  5.838e-01 7.712e-01 8.698e-01 9.199e-01 9.468e-01
##                             PC6       PC7       PC8       PC9      PC10
## Standard deviation     1.211e+05 1.069e+05 1.021e+05 7.987e+04 7.133e+04
## Proportion of Variance 1.496e-02 1.167e-02 1.065e-02 6.510e-03 5.200e-03
## Cumulative Proportion  9.618e-01 9.734e-01 9.841e-01 9.906e-01 9.958e-01
##                            PC11      PC12
## Standard deviation     6.413e+04 1.163e-09
## Proportion of Variance 4.200e-03 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00
```

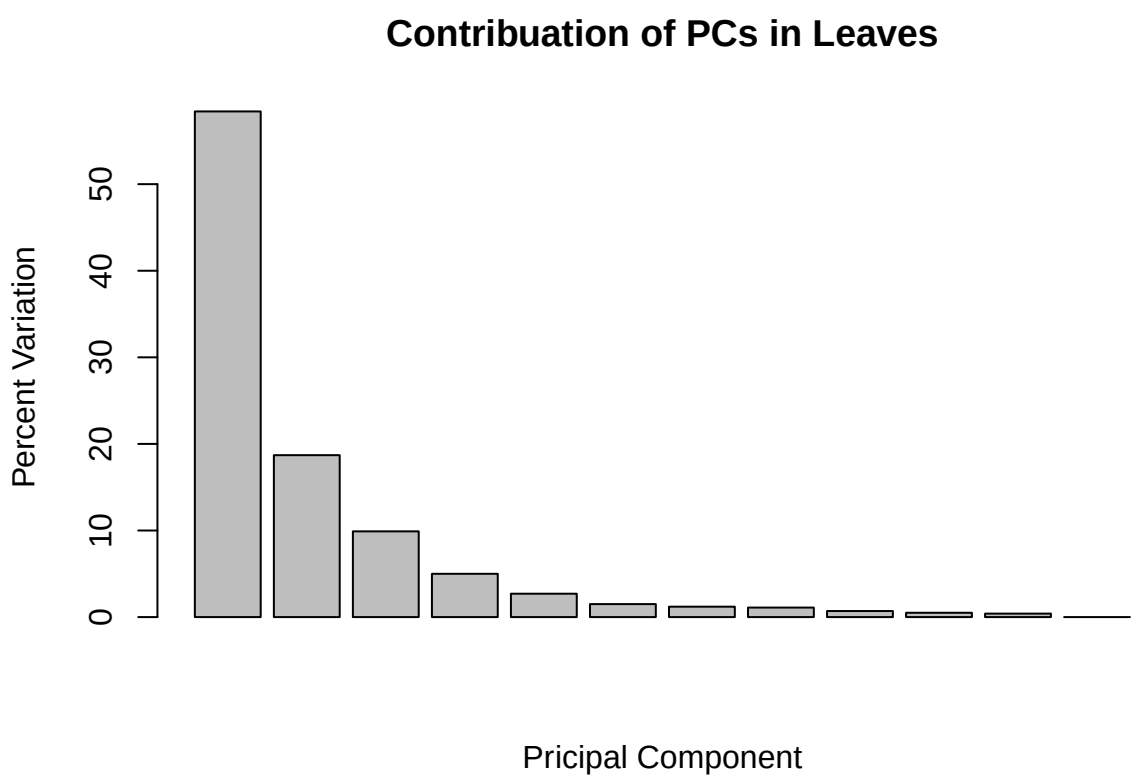
```
# 1) X
## plot PC1 and PC2
plot(pca.leaf$x[,1],pca.leaf$x[,2])
```




```
# 2) sdev
pca.leaf.var <- pca.leaf$sdev^2
head(pca.leaf.var)
```

```
## [1] 571670638623 183599311207  96505957875  49116163347  26310749196
## [6]  14654120331
```

```
## calculate the percentage..
pca.leaf.var.per <- round(pca.leaf.var/sum(pca.leaf.var)*100,1)
## Plotting the percentage is easy with barplot()
barplot(pca.leaf.var.per, main = "Contribuation of PCs in Leaves", xlab = "Pricipal Component", ylab =
```




```
# 3) rotation -- the loading scores rotation

### Draw graph in Leaves
# Format the data the way ggplot2 likes it
dim(pca.leaf$x)[1]
```

```
## [1] 12
```

```
pca.leaf.data <- data.frame(ID = 1:dim(pca.leaf$x)[1])
for (i in 1:dim(pca.leaf$x)[1]){
  pcs <- paste0("PC", i)
  pca.leaf.data[[pcs]] <- pca.leaf$x[,i]
}

pca.leaf.data$ID <- row.names(pca.leaf$x)

pca.leaf.data$tissue <- "Leaf"
pca.leaf.data$Species <- c(rep("CG",4), rep("CR",4), rep("CO", 4))
pca.leaf.data$Species <- factor(pca.leaf.data$Species, levels = c("CG","CR","CO"))

ZscoreLeaf <- pca.leaf.data
head(ZscoreLeaf)
```

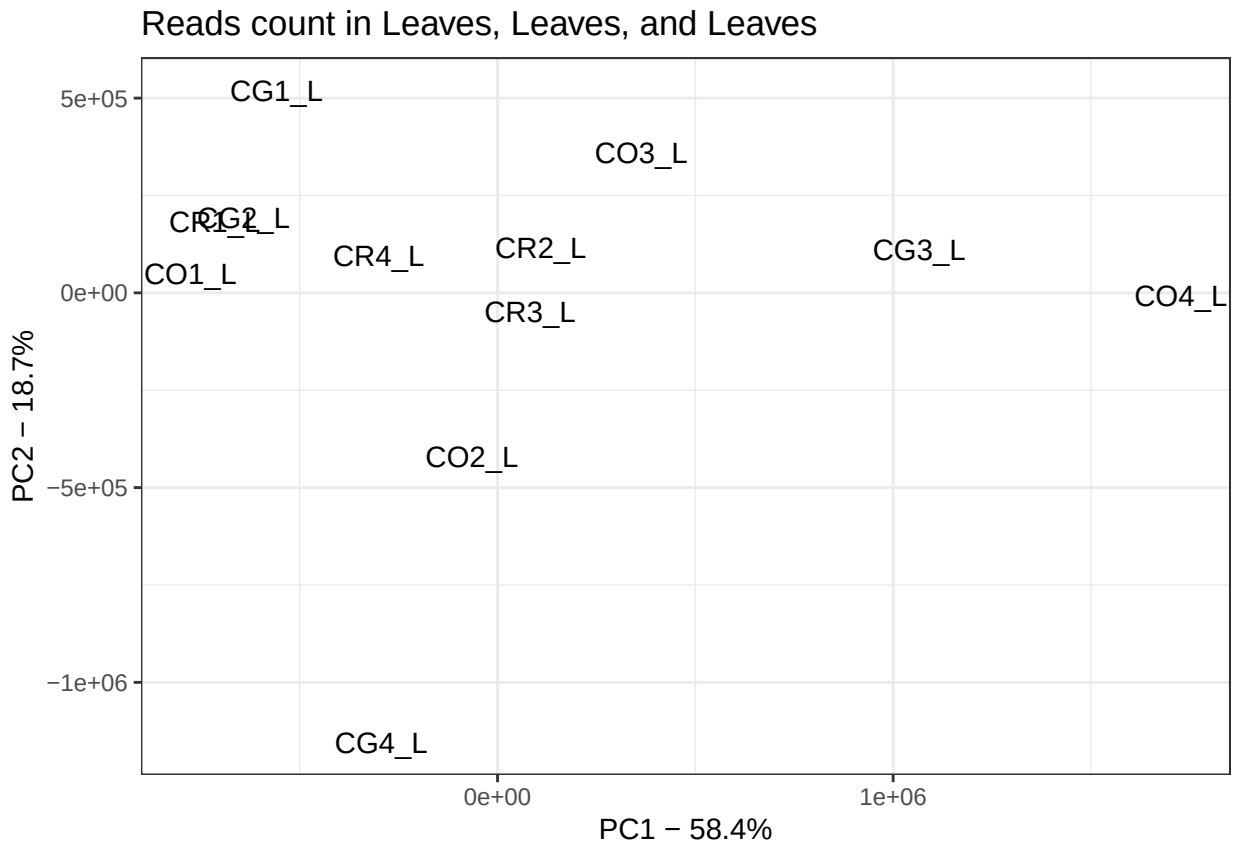
```
##      ID       PC1        PC2        PC3         PC4        PC5        PC6
## 1 CG1_L -558284.7   519933.8  -15655.37   -5317.096  186978.37  -79680.01
```

```
## 2 CG2_L -641695.2   192732.1  238297.58   25806.270  -80235.10 -175014.60
## 3 CG3_L 1066130.9   111055.4 -372594.89  383516.138   49494.19 -138165.97
## 4 CG4_L -293763.0 -1153923.5 -127065.44   10758.389   77744.59 -112676.68
## 5 CR1_L -714465.7   184021.7 -502423.39 -397331.844   32316.65   21645.42
## 6 CR2_L  109772.2   115744.7  -27783.09   32344.135 -457428.60  -13100.84
##         PC7        PC8       PC9        PC10        PC11          PC12 tissue
## 1  17894.53 -137590.77  11961.75   28958.102 -132516.543  2.164350e-09   Leaf
## 2 -79804.84  -21891.00 -41636.17 -155047.414   50658.583  1.570166e-10   Leaf
## 3  76916.97  123072.10  32904.87  -15638.491   -6782.283 -3.445514e-09   Leaf
## 4 -15757.88  -93994.23 -12251.03   47439.694   18083.597 -2.094976e-10   Leaf
## 5 148881.52   67233.20 -25503.16   -9553.748   50148.379  1.549079e-09   Leaf
## 6  34998.94  -53352.01 -47415.13   73062.446  -29658.038  1.306159e-09   Leaf
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```
for(i in 2:13){
  ZscoreLeaf[,i] <- scale(ZscoreLeaf[,i])
}
head(ZscoreLeaf)
```

```
##      ID        PC1        PC2         PC3         PC4        PC5        PC6
## 1 CG1_L -0.7383849  1.2134238 -0.05039485 -0.02399177  1.1527225 -0.6582177
## 2 CG2_L -0.8487032  0.4497990  0.76708337  0.11644290 -0.4946497 -1.4457540
## 3 CG3_L  1.4100600  0.2591817 -1.19938836  1.73049929  0.3051319 -1.1413562
## 4 CG4_L -0.3885296 -2.6930318 -0.40902550  0.04854394  0.4792957 -0.9307953
## 5 CR1_L -0.9449491  0.4294706 -1.61730820 -1.79283844  0.1992323  0.1788076
## 6 CR2_L  0.1451842  0.2701256 -0.08943417  0.14594302 -2.8200494 -0.1082229
##          PC7        PC8        PC9       PC10       PC11        PC12 tissue
## 1  0.1673878 -1.3472679  0.1497683  0.4059747 -2.0664447  0.89431569   Leaf
## 2 -0.7465050 -0.2143533 -0.5213098 -2.1736687  0.7899629  0.06359728   Leaf
## 3  0.7194915  1.2051033  0.4119887 -0.2192420 -0.1057620 -1.42728020   Leaf
## 4 -0.1474013 -0.9203772 -0.1533903  0.6650751  0.2819931 -0.08808157   Leaf
## 5  1.3926573  0.6583372 -0.3193148 -0.1339376  0.7820069  0.63969068   Leaf
## 6  0.3273847 -0.5224148 -0.5936659  1.0242902 -0.4624834  0.53916018   Leaf
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```
# graph parttern 1
ggplot(data = pca.leaf.data, aes(x=PC1, y=PC2, label=ID)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.leaf.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.leaf.var.per[2], "%", sep = "")) +
  theme_bw() +
  ggtitle("Reads count in Leaves, Leaves, and Leaves")
```




```
# graph parttern 2
# PC1 vs PC2
p2 <- ggplot(ZscoreLeaf, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  #stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.leaf.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.leaf.var.per[2], "%", sep = "")) +
  stat_ellipse(geom="polygon", aes(fill = Species), # add frame
               alpha = 0.05,
               show.legend = FALSE,
               level = 0.85) +
  xlim(-4,4) +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold")) +
  theme(legend.position = "none") +
  ggtitle("Leaf")+
  theme(plot.title = element_text(size=16,face="bold"))
```

```
################################
## PCA in Roots
### Define root
head(TMM.Root)[1:5]
```

```
##                       CG1_R     CG2_R     CG3_R    CG4_R     CR1_R
## Carubv10000018m.g 13582.814 18591.777 10782.824 9705.911 8412.2444
## Carubv10000019m.g  1750.108  2336.253  1175.333 1692.486  540.4639
## Carubv10000020m.g  5463.298  6814.073  4037.237 4050.582 5296.0009
## Carubv10000021m.g  7312.001 10293.684  6695.445 5861.066 7109.8276
## Carubv10000022m.g  7785.271 10254.409  6654.702 5701.691 7748.3278
## Carubv10000023m.g  7828.652 10194.108  7969.401 6606.896 8037.3918
```

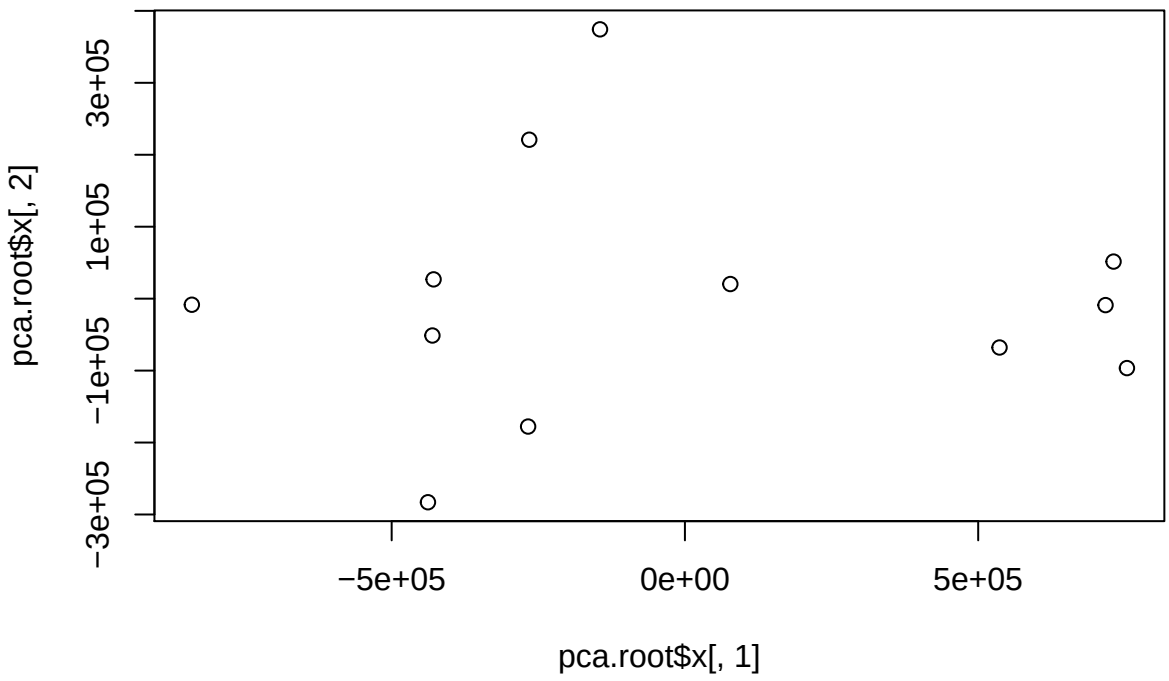
```
dim(TMM.Root)
```

```
## [1] 17307    12
```

```
dim(na.omit(TMM.Root))
```

```
## [1] 17307    12
```

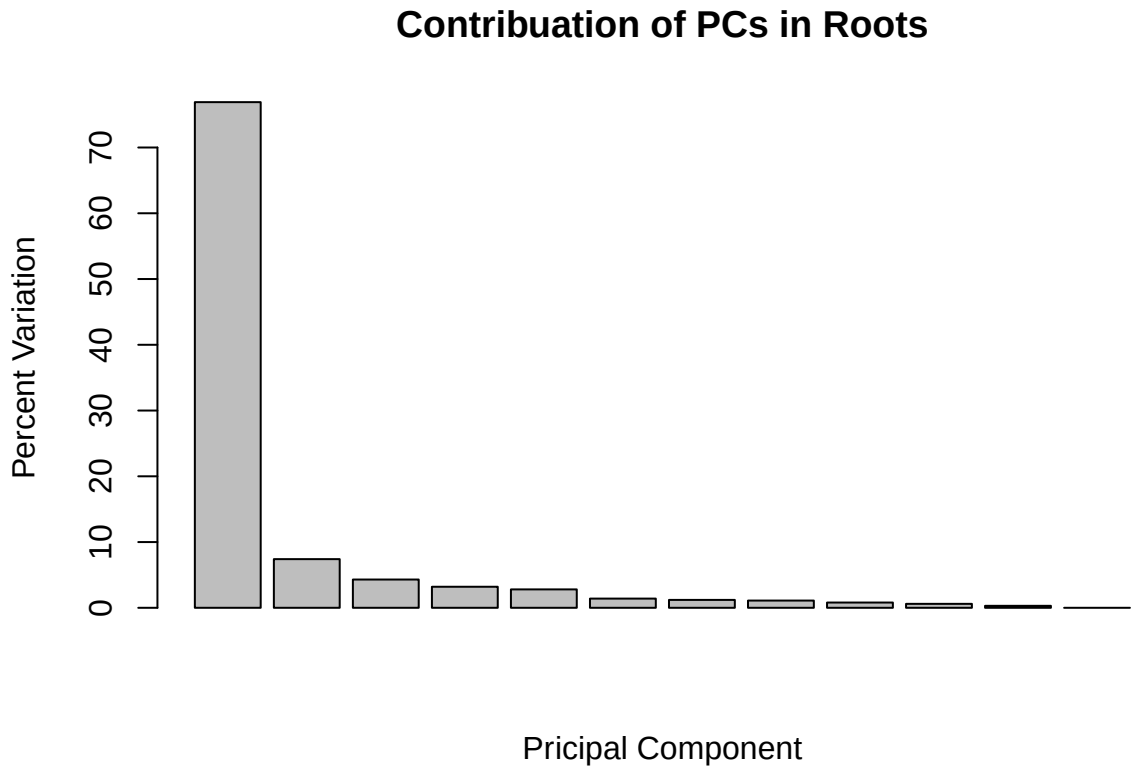
```
### perform PCA in Root
pca.root <- prcomp(t(na.omit(TMM.Root)))
# 1) X
## plot PC1 and PC2
plot(pca.root$x[,1],pca.root$x[,2])
```



```r
# 2) sdev
pca.root.var <- pca.root$sdev^2
head(pca.root.var)
```

```
## [1] 303683761259  29189074778  17038370821  12707001775  11002447048
## [6]   5572785778
```

```r
## calculate the percentage..
pca.root.var.per <- round(pca.root.var/sum(pca.root.var)*100,1)
## Plotting the percentage is easy with barplot()
barplot(pca.root.var.per, main = "Contribuation of PCs in Roots", xlab = "Pricipal Component", ylab = "
```




```r
# 3) rotation -- the loading scores rotation

### Draw graph in Roots
library(ggplot2)
# Format the data the way ggplot2 likes it
dim(pca.root$x)[1]
```

```
## [1] 12
```

```r
pca.root.data <- data.frame(ID = 1:dim(pca.root$x)[1])
for (i in 1:dim(pca.root$x)[1]){
```

```
  pcs <- paste0("PC", i)
  pca.root.data[[pcs]] <- pca.root$x[,i]
}

pca.root.data$ID <- row.names(pca.root$x)
pca.root.data$tissue <- "Root"
pca.root.data$Species <- c(rep("CG",4), rep("CR",4), rep("CO", 4))
pca.root.data$Species <- factor(pca.root.data$Species, levels = c("CG","CR","CO"))

ZscoreRoot <- pca.root.data
head(ZscoreRoot)
```

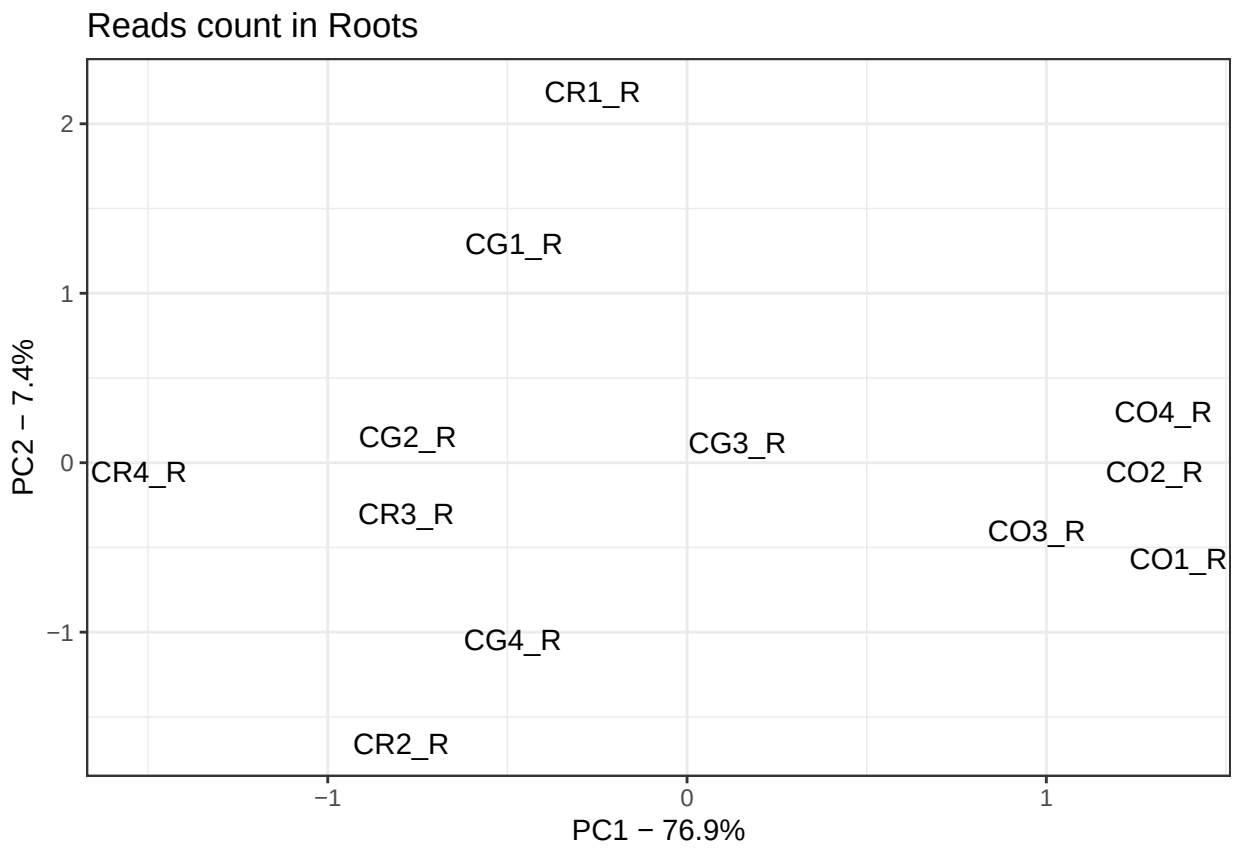
```
##      ID        PC1        PC2        PC3        PC4        PC5        PC6
## 1 CG1_R -265388.93  220897.27   30546.10  235041.65  -86107.91 -44112.952
## 2 CG2_R -428516.10   26781.77 -228283.11  -88752.86 -223122.16  -6918.289
## 3 CG3_R   77415.35   20300.63 -161616.25  137791.18  148007.42  50075.931
## 4 CG4_R -267337.64 -177765.01 -159511.42   35659.99   84211.33  85299.912
## 5 CR1_R -144773.23  374316.54  -14460.95 -164759.91  153728.95 -41870.567
## 6 CR2_R -438100.44 -282991.31   23207.28  -99606.69   67029.67 -85681.098
##          PC7       PC8        PC9       PC10        PC11          PC12 tissue
## 1  71586.426 -10617.60  58972.962 -36359.847   1773.9689  1.268819e-11   Root
## 2  -7389.171  37574.30 -52839.089   2823.173  -2634.8250 -2.225742e-09   Root
## 3 -14159.606 -74279.26 -99888.805 -12227.399    682.1711  5.467950e-10   Root
## 4 -43586.374  64996.32 109328.413   7006.303   3546.4037 -6.135187e-10   Root
## 5   6648.299  48202.61   9318.052   5584.389   3063.7982 -1.417067e-09   Root
## 6 110975.248 -28525.29  -3008.762 -55700.341 -11724.1136 -1.953455e-09   Root
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```
for(i in 2:13){
  ZscoreRoot[,i] <- scale(ZscoreRoot[,i])
}
head(ZscoreRoot)
```

```
##      ID        PC1        PC2        PC3        PC4        PC5         PC6
## 1 CG1_R -0.4815840  1.2929454  0.2340140  2.0850827 -0.8209154 -0.59092171
## 2 CG2_R -0.7776002  0.1567578 -1.7488789 -0.7873373 -2.1271497 -0.09267498
## 3 CG3_R  0.1404806  0.1188227 -1.2381436  1.2223621  1.4110384  0.67079970
## 4 CG4_R -0.4851202 -1.0404857 -1.2220184  0.3163440  0.8028343  1.14264786
## 5 CR1_R -0.2627105  2.1909318 -0.1107855 -1.4616049  1.4655850 -0.56088350
## 6 CR2_R -0.7949923 -1.6563913  0.1777912 -0.8836229  0.6390318 -1.14775409
##          PC7        PC8         PC9        PC10        PC11         PC12 tissue
## 1  1.0211018 -0.1624989  1.05329609 -0.75690766  0.05369203  0.006187718   Root
## 2 -0.1053984  0.5750623 -0.94374106  0.05877036 -0.07974724 -1.246446404   Root
## 3 -0.2019712 -1.1368196 -1.78408010 -0.25453935  0.02064701  0.305075890   Root
## 4 -0.6217118  0.9947472  1.95267773  0.14585113  0.10733764 -0.344240029   Root
## 5  0.0948307  0.7377251  0.16642657  0.11625094  0.09273080 -0.793908781   Root
## 6  1.5829402 -0.4365702 -0.05373848 -1.15952124 -0.35484924 -1.094073557   Root
```

```
##   Species
## 1      CG
## 2      CG
## 3      CG
## 4      CG
## 5      CR
## 6      CR
```

```r
# graph parttern 1
ggplot(data = ZscoreRoot, aes(x=PC1, y=PC2, label=ID)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.root.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.root.var.per[2], "%", sep = "")) +
  theme_bw() +
  ggtitle("Reads count in Roots")
```




```r
# graph parttern 2
p3 <- ggplot(ZscoreRoot, aes(PC1, PC2, col = Species, fill = Species))  +
  geom_point(size = 6, stroke = 2, aes(shape = Species)) +
  scale_shape_manual(values=c(8,10,4))+
  scale_color_manual(values = c("#f90052","darkviolet","#006beb"))  +
  scale_fill_manual(values = c("#f90052","darkviolet","#006beb")) +
  #stat_ellipse(aes(x= PC1, y=PC2, group = species), geom = "polygon", alpha = 0.1) +
  xlab(paste("PC1 - ", pca.root.var.per[1], "%", sep = "")) +
  ylab(paste("PC2 - ", pca.root.var.per[2], "%", sep = "")) +
```

```
  stat_ellipse(geom="polygon", aes(fill = Species), # add frame
               alpha = 0.05,
               show.legend = FALSE,
               level = 0.85) +
  xlim(-2,2) +
  #labs(fill = "Species") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
        panel.border = element_rect(fill= "transparent")) +
  theme(axis.text=element_text(size=16,face="bold"),
        axis.title=element_text(size=16,face="bold"),
        legend.title = element_text(size=16, face = "bold"),
        legend.text = element_text(size=16)) +
  theme(legend.position = "none") +
  ggtitle("Root") +
  theme(plot.title = element_text(size=16,face="bold"))
```

Combine all tissues into one figure

```
grid.newpage()
ggarrange(p3, p2, p1, ncol=3, nrow=1, common.legend = TRUE, legend="right") # set the common legend
```