Coral Analysis Notes

Jason Chari and Elizabeth Maloney

June 9, 2016

1. Read in Data

```r
setwd("~/Desktop/Coral Project Data/ALL DCM")
#Read in Data
filenames = c("PAL003DCM1.csv","PAL006DCM1.csv","PAL007DCM1.csv","PAL014DCM1.csv","PAL015DCM1.csv",
              "PAL017DCM1.csv","PAL018DCM1.csv","PAL019D3.csv","PAL029DCM1.csv","PAL042D2.csv","PAL043D(
              "PAL044DCM1.csv","PAL045DCM1.csv","PAL048DCM1.csv","PAL050D1.csv","PAL052D1.csv",
              "PAL052D2.csv","PAL130D1.csv","PAL131D1.csv","PAL133D1.csv","PAL134D1.csv","PAL137DCM1.cs
              "PAL227D1.csv","PAL247DCM1.csv","PAL248D1.csv","PAL250D1.csv","PAL250D2.csv","PAL252DCM1.
              "PAL270D1.csv","PAL299DCM1.csv","PAL300D1.csv","PAL301D1.csv","PAL302D1.csv","PAL306D1.cs
              "PAL332D1.csv","PAL333D1.csv","PAL334D1.csv","PAL340D1.csv","PAL355D1.csv","PAL356D1.csv"
              "PAL362D1.csv","PAL374DCM1.csv","PAL375D1.csv","PAL378DCM1.csv","PAL379DCM1.csv","PAL382D(


PAL003DCM1 = read.csv(filenames[1],header=FALSE)
PAL006DCM1 = read.csv(filenames[2],header=FALSE)
PAL007DCM1 = read.csv(filenames[3],header=FALSE)
PAL014DCM1 = read.csv(filenames[4],header=FALSE)
PAL015DCM1 = read.csv(filenames[5],header=FALSE)
PAL017DCM1 = read.csv(filenames[6],header=FALSE)
PAL018DCM1 = read.csv(filenames[7],header=FALSE)
PAL019DCM1 = read.csv(filenames[8],header=FALSE)
PAL029DCM1 = read.csv(filenames[9],header=FALSE)
PAL042DCM1 = read.csv(filenames[10],header=FALSE)
PAL043DCM1 = read.csv(filenames[11],header=FALSE)
PAL044DCM1 = read.csv(filenames[12],header=FALSE)
PAL045DCM1 = read.csv(filenames[13],header=FALSE)
PAL048DCM1 = read.csv(filenames[14],header=FALSE)
PAL050DCM1 = read.csv(filenames[15],header=FALSE)
PAL052DCM1 = read.csv(filenames[16],header=FALSE)
#PAL052DCM2 = read.csv(filenames[17],header=FALSE) Broad peak @ 15.8 looked less like rest of data. Use
PAL130DCM1 = read.csv(filenames[18],header=FALSE)
PAL131DCM1 = read.csv(filenames[19],header=FALSE)
PAL133DCM1 = read.csv(filenames[20],header=FALSE)
PAL134DCM1 = read.csv(filenames[21],header=FALSE)
PAL137DCM1 = read.csv(filenames[22],header=FALSE)
PAL227DCM1 = read.csv(filenames[23],header=FALSE)
PAL247DCM1 = read.csv(filenames[24],header=FALSE)
PAL248DCM1 = read.csv(filenames[25],header=FALSE)
PAL250DCM1 = read.csv(filenames[26],header=FALSE)
#PAL250D2 = read.csv(filenames[27],header=FALSE)  Graph 250D2 looked the same as 052D2. Use D1.
PAL252DCM1 = read.csv(filenames[28],header=FALSE)
PAL270DCM1 = read.csv(filenames[29],header=FALSE)
PAL299DCM1 = read.csv(filenames[30],header=FALSE)
PAL300DCM1 = read.csv(filenames[31],header=FALSE)
PAL301DCM1 = read.csv(filenames[32],header=FALSE)
PAL302DCM1 = read.csv(filenames[33],header=FALSE)
```

```
PAL306DCM1 = read.csv(filenames[34],header=FALSE)
PAL307DCM1 = read.csv(filenames[35],header=FALSE)
PAL332DCM1 = read.csv(filenames[36],header=FALSE)
PAL333DCM1 = read.csv(filenames[37],header=FALSE)
PAL334DCM1 = read.csv(filenames[38],header=FALSE)
PAL340DCM1 = read.csv(filenames[39],header=FALSE)
PAL355DCM1 = read.csv(filenames[40],header=FALSE)
PAL356DCM1 = read.csv(filenames[41],header=FALSE)
PAL359DCM1 = read.csv(filenames[42],header=FALSE)
PAL362DCM1 = read.csv(filenames[43],header=FALSE)
PAL374DCM1 = read.csv(filenames[44],header=FALSE)
PAL375DCM1 = read.csv(filenames[45],header=FALSE)
PAL378DCM1 = read.csv(filenames[46],header=FALSE)
PAL379DCM1 = read.csv(filenames[47],header=FALSE)
PAL382DCM1 = read.csv(filenames[48],header=FALSE)
```

2. Use trap function to generate trapezoidal area

```
DELTA = 0.00033333333333 #Time increment by which intensity is measured (in minutes)

trap <- function(intens, delta) {
  intens1 <- rep(intens)
  intens1 = intens1[-1]
  intens = intens[1:length(intens)-1]
  trapArea = (intens + intens1)*delta/2
  return(c(as.numeric(trapArea), 0))
}


norm <- function(trapArea){
  tot = sum(trapArea)
  return(trapArea/tot)
}
```

3. Generate dataframe with normalized trapezoid areas for all time points as well as the sample number, separating solvent ("greasiness"), and clade.

```
#Process as dataframe - CHANGE TO AS.NUMERIC
dataframe = rbind(
  #S. glaucum clade F
  c(003,"DCM","F", norm(trap(as.numeric(PAL003DCM1[,2]), DELTA))),
  c(006,"DCM","F", norm(trap(as.numeric(PAL006DCM1[-77002,2]), DELTA))),
  c(007,"DCM","F", norm(trap(as.numeric(PAL007DCM1[,2]), DELTA))),
  c(014,"DCM","F", norm(trap(as.numeric(PAL014DCM1[-77002,2]), DELTA))),
  c(015,"DCM","F", norm(trap(as.numeric(PAL015DCM1[-77002,2]), DELTA))),
  c(017,"DCM","F", norm(trap(as.numeric(PAL017DCM1[,2]), DELTA))),
  c(018,"DCM","F", norm(trap(as.numeric(PAL018DCM1[,2]), DELTA))),
  c(043,"DCM","F", norm(trap(as.numeric(PAL043DCM1[,2]), DELTA))),
  c(044,"DCM","F", norm(trap(as.numeric(PAL044DCM1[-77002,2]), DELTA))),
  c(048,"DCM","F", norm(trap(as.numeric(PAL048DCM1[,2]), DELTA))),
  c(050,"DCM","F", norm(trap(as.numeric(PAL050DCM1[,2]), DELTA))),
  c(130,"DCM","F", norm(trap(as.numeric(PAL130DCM1[,2]), DELTA))),
  c(131,"DCM","F", norm(trap(as.numeric(PAL131DCM1[,2]), DELTA))),
  c(133,"DCM","F", norm(trap(as.numeric(PAL133DCM1[-77002,2]), DELTA))),
```

```r
  c(134,"DCM","F", norm(trap(as.numeric(PAL134DCM1[,2]), DELTA))),
  c(248,"DCM","F", norm(trap(as.numeric(PAL248DCM1[,2]), DELTA))),
  c(270,"DCM","F", norm(trap(as.numeric(PAL270DCM1[,2]), DELTA))),
  c(302,"DCM","F", norm(trap(as.numeric(PAL302DCM1[-77002,2]), DELTA))),
  c(333,"DCM","F", norm(trap(as.numeric(PAL333DCM1[-77002,2]), DELTA))),
  c(334,"DCM","F", norm(trap(as.numeric(PAL334DCM1[,2]), DELTA))),
  c(374,"DCM","F", norm(trap(as.numeric(PAL374DCM1[,2]), DELTA))),
  c(378,"DCM","F", norm(trap(as.numeric(PAL378DCM1[,2]), DELTA))),
  c(382,"DCM","F", norm(trap(as.numeric(PAL382DCM1[,2]), DELTA))),
  c(227,"DCM","F", norm(trap(as.numeric(PAL227DCM1[,2]), DELTA))),

  #S. glaucum clade D
  c(029,"DCM","D", norm(trap(as.numeric(PAL029DCM1[,2]), DELTA))),
  c(247,"DCM","D", norm(trap(as.numeric(PAL247DCM1[-77002,2]), DELTA))),
  c(252,"DCM","D", norm(trap(as.numeric(PAL252DCM1[,2]), DELTA))),
  c(299,"DCM","D", norm(trap(as.numeric(PAL299DCM1[-77002,2]), DELTA))),
  c(300,"DCM","D", norm(trap(as.numeric(PAL300DCM1[-77002,2]), DELTA))),
  c(306,"DCM","D", norm(trap(as.numeric(PAL306DCM1[,2]), DELTA))),
  c(332,"DCM","D", norm(trap(as.numeric(PAL332DCM1[-77002,2]), DELTA))),
  c(340,"DCM","D", norm(trap(as.numeric(PAL340DCM1[-77002,2]), DELTA))),
  c(355,"DCM","D", norm(trap(as.numeric(PAL355DCM1[-77002,2]), DELTA))),
  c(359,"DCM","D", norm(trap(as.numeric(PAL359DCM1[,2]), DELTA))),

  #Trocheliophorum
  c(019,"DCM","T", norm(trap(as.numeric(PAL019DCM1[,2]), DELTA))),
  c(042,"DCM","T", norm(trap(as.numeric(PAL042DCM1[-77002,2]), DELTA))),
  c(045,"DCM","T", norm(trap(as.numeric(PAL045DCM1[,2]), DELTA))),
  c(052,"DCM","T", norm(trap(as.numeric(PAL052DCM1[,2]), DELTA))),
  c(137,"DCM","T", norm(trap(as.numeric(PAL137DCM1[,2]), DELTA))),
  c(250,"DCM","T", norm(trap(as.numeric(PAL250DCM1[,2]), DELTA))),
  c(301,"DCM","T", norm(trap(as.numeric(PAL301DCM1[-77002,2]), DELTA))),
  c(307,"DCM","T", norm(trap(as.numeric(PAL307DCM1[,2]), DELTA))),
  c(356,"DCM","T", norm(trap(as.numeric(PAL356DCM1[,2]), DELTA))),
  c(362,"DCM","T", norm(trap(as.numeric(PAL362DCM1[,2]), DELTA))),
  c(375,"DCM","T", norm(trap(as.numeric(PAL375DCM1[,2]), DELTA))),
  c(379,"DCM","T", norm(trap(as.numeric(PAL379DCM1[,2]), DELTA)))
)

#generate column name vector for the dataframe
names <- function(){
  nameVec = c(1:77001)
  nameVec = as.character(nameVec)

  for(i in 1:77001){
    nameVec[i] = paste("t", nameVec[i], sep="")
  }

  nameVec[1:5]
  nameVec = c("Sample", "Greasiness", "Clade", nameVec)
  return(nameVec)
}
colnames(dataframe) <- names()
```

Snapshot of dataframe:

```r
install.packages("stargazer", repos='http://cran.us.r-project.org')
```

```
##
## The downloaded binary packages are in
##     /var/folders/7k/8ngqf1rd4ggc6b8q6cflfr6m0000gn/T//Rtmpgq3ev0/downloaded_packages
```

```r
library(stargazer)
```

```
##
## Please cite as:
##
##   Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##   R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```r
stargazer(dataframe[1:5,1:5])
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard
## % Date and time: Thu, Jun 09, 2016 - 16:26:25
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Sample & Greasiness & Clade & t1 & t2 \\
## \hline \\[-1.8ex]
## 3 & DCM & F & 1.14692140379615e-05 & 1.14776729740806e-05 \\
## 6 & DCM & F & 1.04060759079854e-05 & 1.04002200252189e-05 \\
## 7 & DCM & F & 1.13192549346752e-05 & 1.13065820709645e-05 \\
## 14 & DCM & F & 1.08956168100273e-05 & 1.08845296211337e-05 \\
## 15 & DCM & F & 1.1349807025902e-05 & 1.13571810821435e-05 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

3.Create a bin function to partition the data by summing across equal time intervals.

```r
binner <- function(df, binWidth, delta) {
  rowz = c()
  for(j in 1:nrow(df)) {
    col = c()
    row_normed = as.numeric(df[j,-(1:3)])
    for(i in seq(4, ncol(df)-binWidth, binWidth)){
      end = i + binWidth - 1
      col = c(col, sum(as.numeric(df[j,i:end])))
    }
    rowz = rbind(rowz, col)
  }
```

```
    return(rowz)
}


#We omit the first and last 15000 observations (5 minutes) because there is little observed activity du
binned_matrix_1100 = binner(dataframe[,15000:62004],1100, DELTA)
binned_matrix_2000 = binner(dataframe[,15000:62004],2000, DELTA)
binned_matrix_3000 = binner(dataframe[,15000:62004],3000, DELTA)
```

4. PCA — 22 Second Bin Width (1100 time points)

```
trapAreas.centered = scale(binned_matrix_1100, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:42] 0.01071 0.0083 0.00596 0.00487 0.00413 ...
##  $ rotation: num [1:42, 1:42] -0.0229 -0.0247 -0.0144 -0.0218 -0.0221 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:42] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:42] 8.67e-19 1.13e-19 -3.02e-19 1.51e-19 3.02e-19 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:42] 0.000809 0.027598 -0.004235 -0.007181 -0.001474 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:42] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```
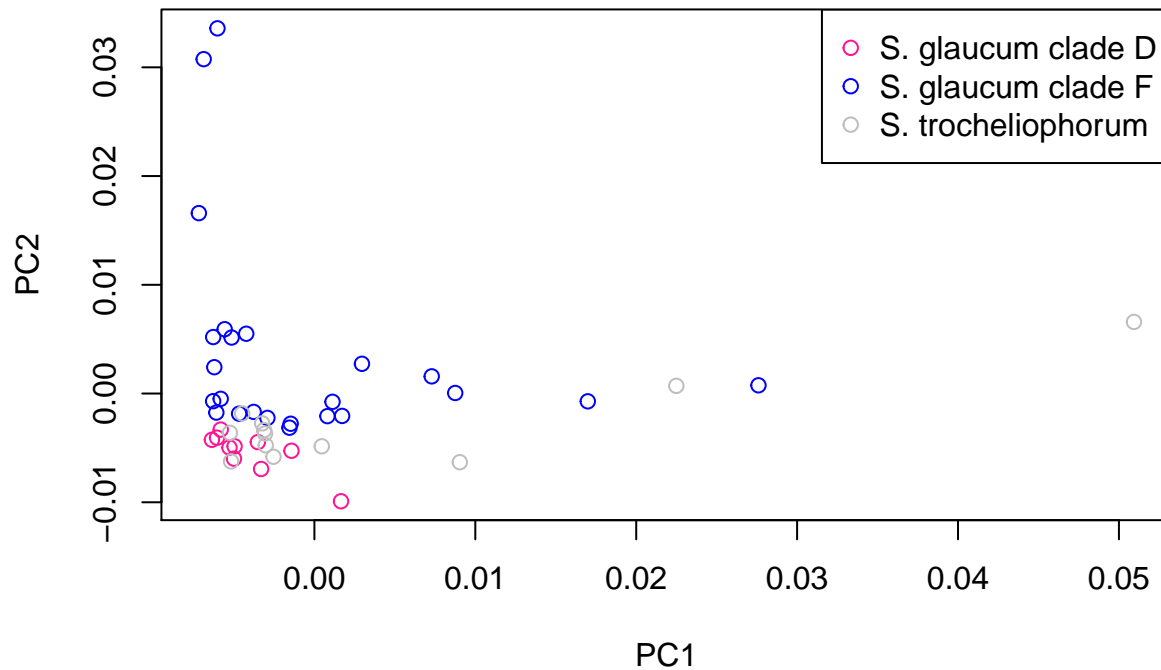
```
PC = pca.result$x


F = 1:24
D = 25:34
T = 35:46


plot(PC[,1],PC[,2],col="white",main = "PCA - 22 Second Bin Width",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topright",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deepp
```
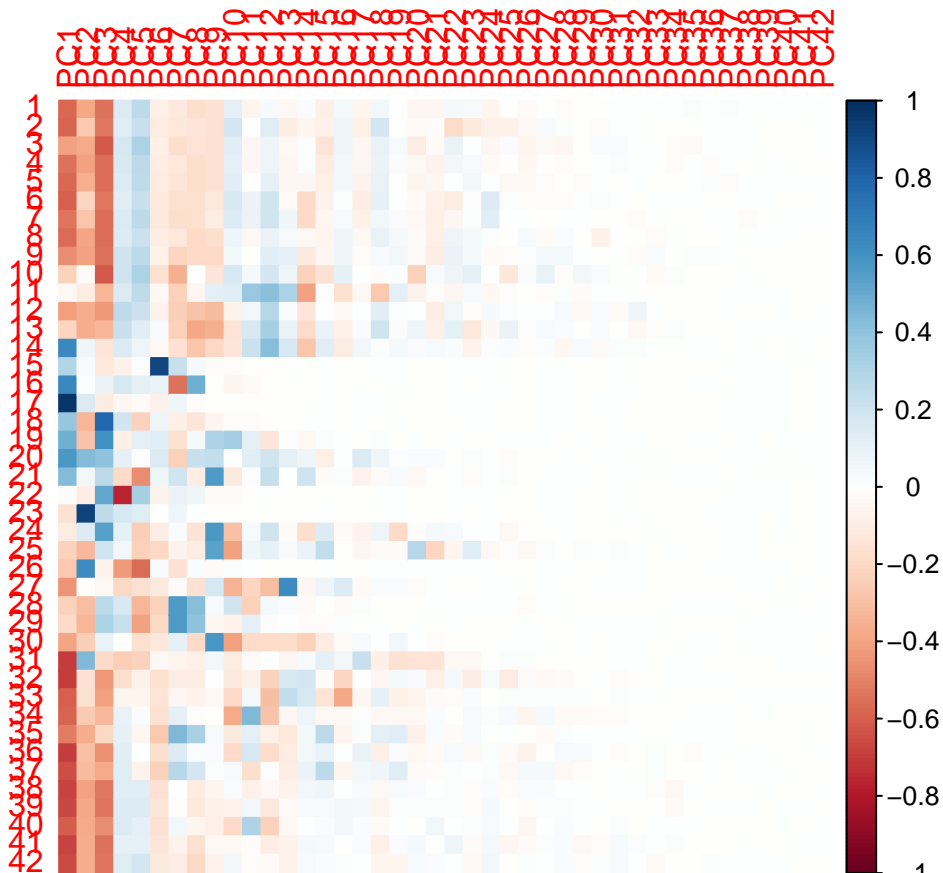
PCA – 22 Second Bin Width


```
correlation_1100 = cor(PC,binned_matrix_1100)
install.packages("corrplot", repos='http://cran.us.r-project.org')
```

```
##
## The downloaded binary packages are in
##   /var/folders/7k/8ngqf1rd4ggc6b8q6cflfr6m0000gn/T//Rtmpgq3ev0/downloaded_packages
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.2.5
```

```
corrplot(t(correlation_1100), method="color")
```

```
par(mfrow=c(1,1))
```

5. PCA — 40 Second Bin Width (2000 time points)

```
trapAreas.centered = scale(binned_matrix_2000, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:23] 0.01264 0.00829 0.00793 0.00503 0.00459 ...
##  $ rotation: num [1:23, 1:23] 0.0433 0.0352 0.0407 0.0407 0.0362 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:23] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:23] -1.21e-18 -6.79e-19 -1.13e-18 -4.53e-19 9.80e-19 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:23] 0.00152 -0.02339 0.00685 0.01005 0.00481 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:23] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```
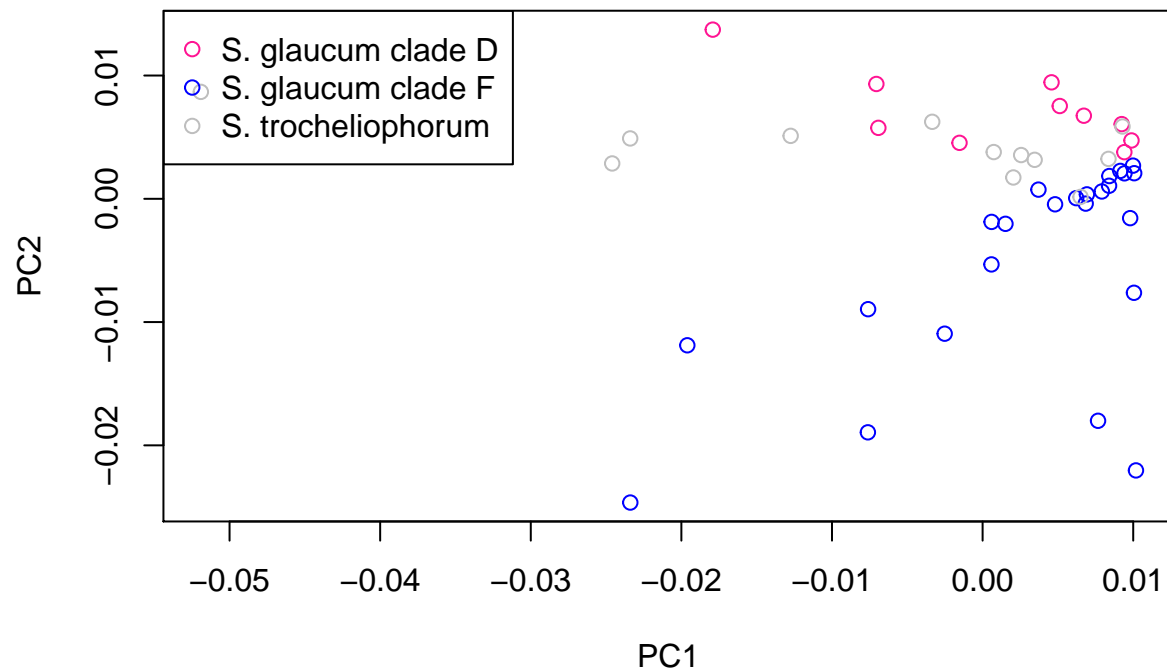
```
PC = pca.result$x
```

```
plot(PC[,1],PC[,2],col="white",main = "PCA - 40 Second Bin Width",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```
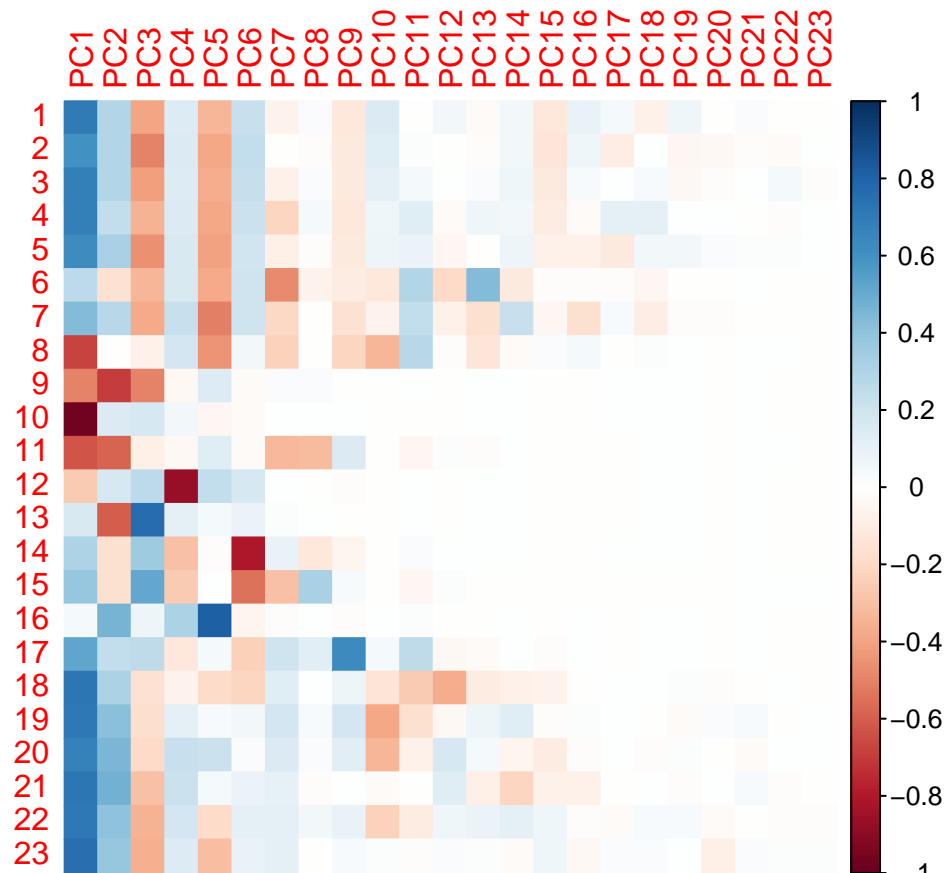

PCA – 40 Second Bin Width

```
correlation_2000 = cor(PC,binned_matrix_2000)
library(corrplot)
corrplot(t(correlation_2000), method="color")
```

```
par(mfrow=c(1,1))
```

6. PCA — 60 Second Bin Width (3000 time points)

```
trapAreas.centered = scale(binned_matrix_3000, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:15] 0.01384 0.00844 0.00789 0.00518 0.00467 ...
##  $ rotation: num [1:15, 1:15] 0.05585 0.05207 0.05521 0.02565 0.00469 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:15] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:15] 2.11e-18 -1.66e-18 2.41e-18 1.96e-18 2.11e-18 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:15] -0.000768 -0.029461 0.008086 0.011389 0.004603 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:15] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```
PC = pca.result$x
```

```r
plot(PC[,1],PC[,2],col="white",main = "PCA - 60 Second Bin Width",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```
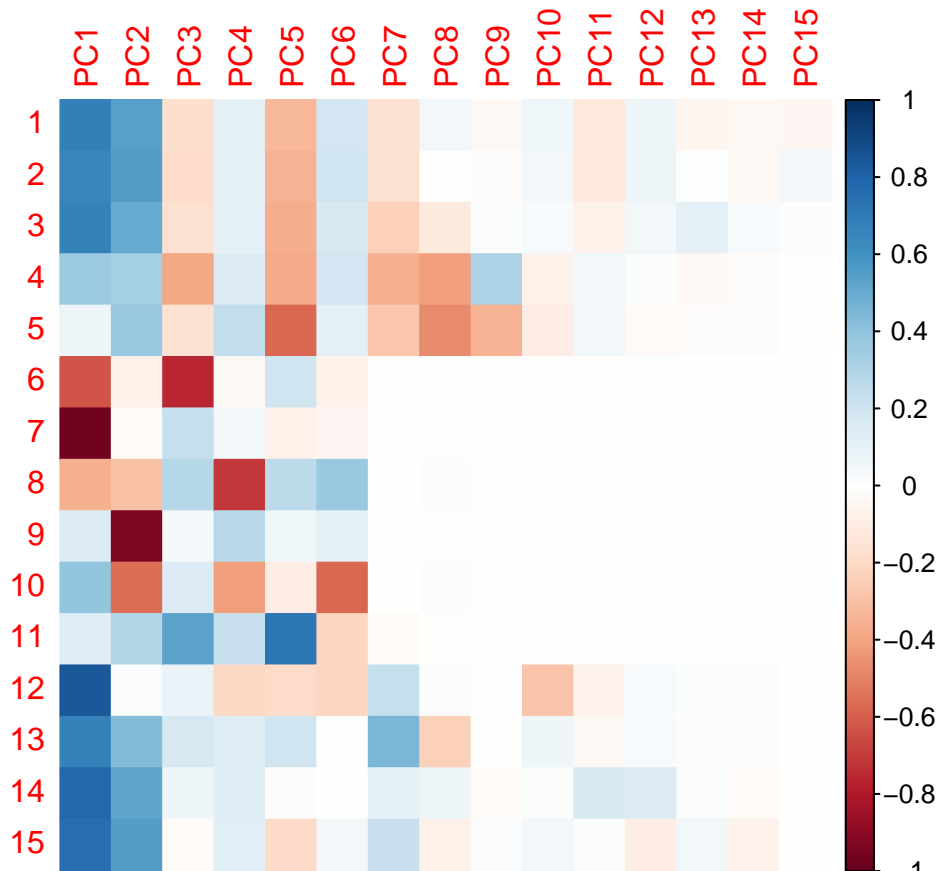
PCA – 60 Second Bin Width
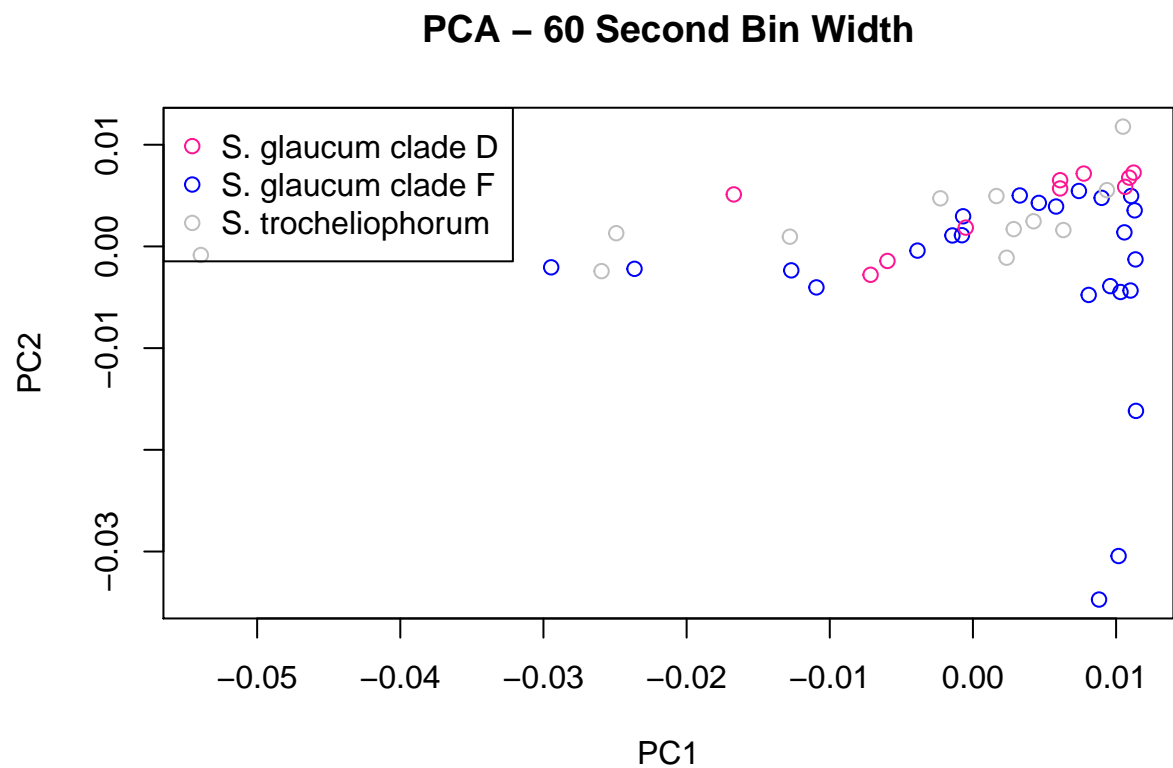

```r
correlation_3000 = cor(PC,binned_matrix_3000)
require(corrplot)
corrplot(t(correlation_3000), method="color")
```

```
par(mfrow=c(1,1))
```

7. LDA — 22 Second Bin Width (1100 time points)

```
library(MASS)
lda_info = lda(as.factor(dataframe[,3])~binned_matrix_1100)
lda_info$scaling #Linear discriminant coefficient matrix
```

```
##                             LD1           LD2
## binned_matrix_11001   -104626.0663    10228.28146
## binned_matrix_11002      -855.8969   -14860.51149
## binned_matrix_11003   -144901.0236    35402.80852
## binned_matrix_11004    355613.5442  -145622.37549
## binned_matrix_11005    106546.7389   140585.78999
## binned_matrix_11006   -176209.2856    62750.74561
## binned_matrix_11007    143856.7211   -32617.94652
## binned_matrix_11008    -34205.4316   -25502.08948
## binned_matrix_11009   -223364.8249     3097.94805
## binned_matrix_110010    15922.0793   -26619.60456
## binned_matrix_110011    29712.8314    -5738.30435
## binned_matrix_110012   -97583.1300   -26677.16401
## binned_matrix_110013    28772.8417    22603.32418
## binned_matrix_110014    22246.6172    12033.51032
## binned_matrix_110015    -1805.8863      156.54772
## binned_matrix_110016     1934.0323     2493.76383
```

```
## binned_matrix_110017    -1996.5652       587.91484
## binned_matrix_110018    -1307.0681        43.90009
## binned_matrix_110019      332.9433      5053.36970
## binned_matrix_110020    -8482.5663     -4631.48285
## binned_matrix_110021      207.6220      9370.50302
## binned_matrix_110022      609.4370      1984.38849
## binned_matrix_110023      565.8616      1791.67880
## binned_matrix_110024   -14281.8951     -3935.47212
## binned_matrix_110025     2625.7796     -4885.36575
## binned_matrix_110026     -786.8961      -920.71196
## binned_matrix_110027    -6632.6838      4705.91993
## binned_matrix_110028   -45931.1360     -4643.71608
## binned_matrix_110029    29642.2433      5981.37515
## binned_matrix_110030     9381.4340     -2700.87486
## binned_matrix_110031     9167.8694      9691.93675
## binned_matrix_110032   -85420.5136    -63585.90400
## binned_matrix_110033    78414.3501     38386.20431
## binned_matrix_110034  -115118.9355    -33229.79348
## binned_matrix_110035   -17510.0956    -39640.35319
## binned_matrix_110036   -55233.2862     36399.53826
## binned_matrix_110037   105351.5463     13398.16476
## binned_matrix_110038  -472786.4083    170248.74985
## binned_matrix_110039   601819.1063    -61359.28125
## binned_matrix_110040  -160550.6088     36207.85879
## binned_matrix_110041   131813.5400    109218.74026
## binned_matrix_110042   -25569.3245   -154738.39059
```

```r
lda_info_p = predict(lda_info,as.data.frame(binned_matrix_1100))
lda_info_pclass = predict(lda_info)$class
par(mar=c(5,4,4,4))
v = as.numeric(lda_info_pclass)
v[which(v==1)] =  "deeppink"    #D
v[which(v==2)] =  12            #F
v[which(v==3)] =  "grey"        #Trochi

plot(lda_info_p$x[,1],pch=16, col = v,main = "LDA - 22 Second Bin Width",xlab="df1",ylab="df2") #Plot t
legend("topright",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deepp
```

LDA – 22 Second Bin Width
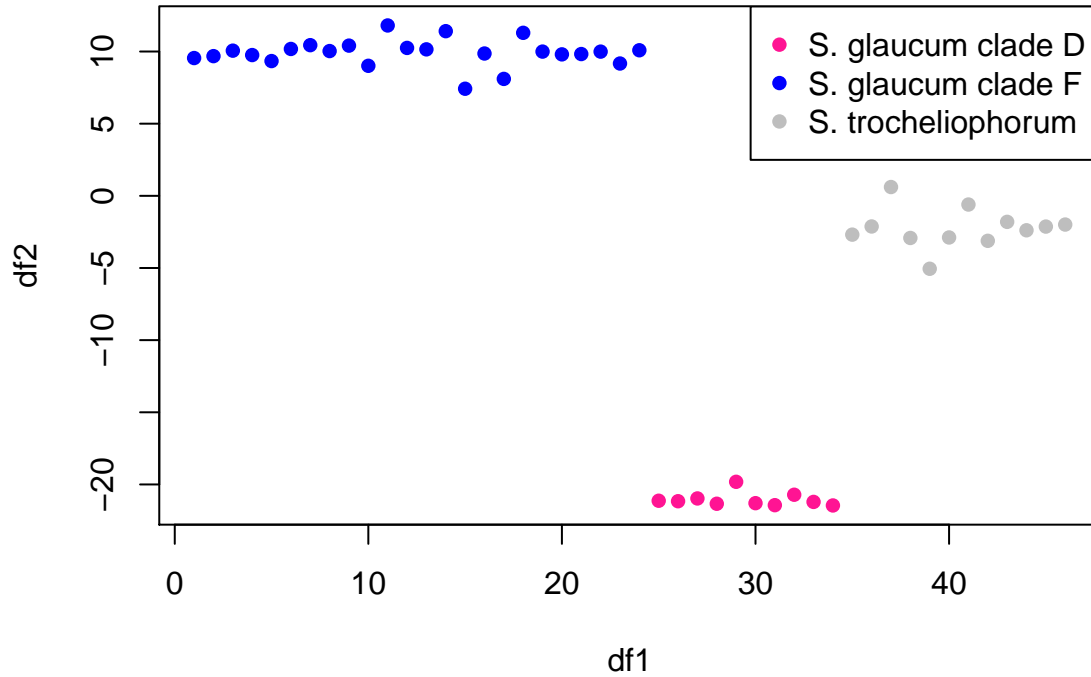

```r
clade.manova = manova(binned_matrix_1100~as.factor(dataframe[,3]))
clade.wilks = summary(clade.manova,test="Wilks");clade.wilks
```

```
##                          Df     Wilks approx F num Df den Df  Pr(>F)
## as.factor(dataframe[, 3])  2 0.00013874   3.9951     84      4 0.09128 .
## Residuals                 43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Wilk's Lambda statistic corresponds to a p-value of ~.09, suggesting that the clades account for a large (but only marginally significant) proportion of the variance in binned area.

8. LDA — 40 Second Bin Width (2000 time points)

```r
lda_info = lda(as.factor(dataframe[,3])~binned_matrix_2000)
lda_info$scaling  #Linear discriminant coefficient matrix
```

```
##                          LD1         LD2
## binned_matrix_20001   4161.89020 -2569.85753
## binned_matrix_20002     30.39786 -1210.66777
## binned_matrix_20003   -562.85208  6238.28242
## binned_matrix_20004  -2063.06313  1245.36566
## binned_matrix_20005   -814.89961 -4187.61634
## binned_matrix_20006   1002.24976  1006.37155
## binned_matrix_20007   -383.68851  3158.21210
## binned_matrix_20008    752.62717  -351.43071
## binned_matrix_20009    211.96921   255.17764
## binned_matrix_200010    81.06543   420.10313
```

```
## binned_matrix_200011    455.07376     60.13171
## binned_matrix_200012    177.50249    259.18391
## binned_matrix_200013    290.56170    323.63210
## binned_matrix_200014    634.44067    471.09024
## binned_matrix_200015   -272.25482   -218.28673
## binned_matrix_200016    157.97134     81.33965
## binned_matrix_200017    292.45486    438.12204
## binned_matrix_200018   1492.99163   1189.57987
## binned_matrix_200019  -2706.26921  -1990.82575
## binned_matrix_200020  -3413.65047  -3019.88153
## binned_matrix_200021   3407.10516   5316.86918
## binned_matrix_200022   8402.10360   7454.26853
## binned_matrix_200023  -3375.75442  -3276.59642
```

```r
lda_info_p = predict(lda_info,as.data.frame(binned_matrix_2000))
lda_info_pclass = predict(lda_info)$class
par(mar=c(5,4,4,4))
v = as.numeric(lda_info_pclass)
v[which(v==1)] =  "deeppink"    #D
v[which(v==2)] =  12            #F
v[which(v==3)] =  "grey"        #Trochi

plot(lda_info_p$x[,1],pch=16, col = v,main = "LDA - 40 Second Bin Width",xlab="df1",ylab="df2") #Plot t
legend("topright",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deepp
```


LDA – 40 Second Bin Width
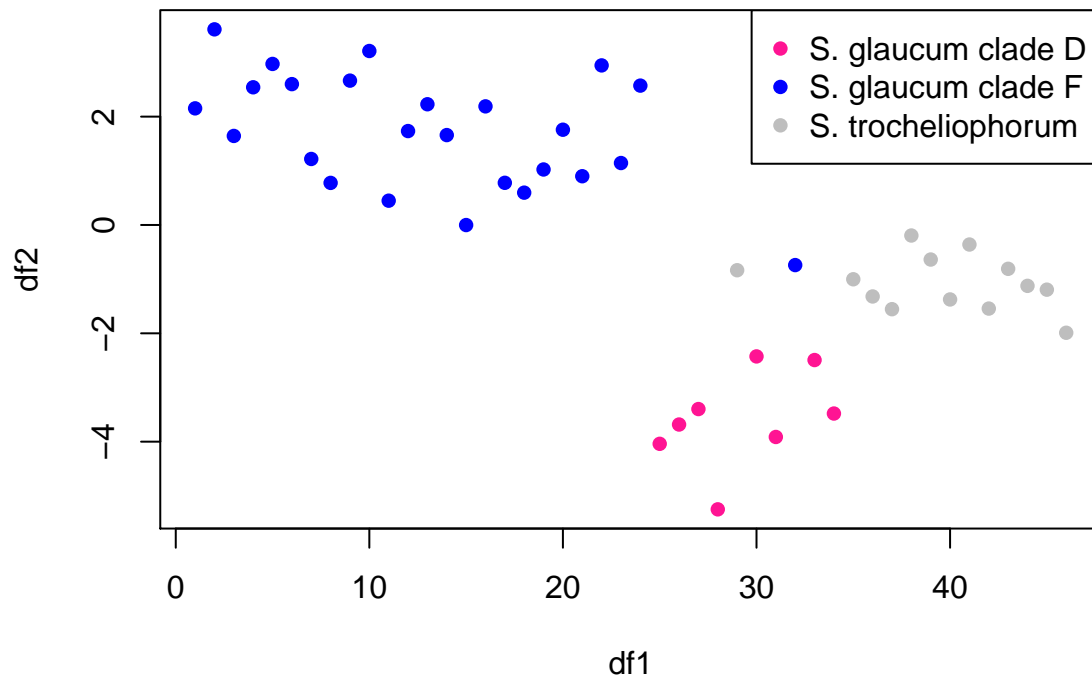
```r
clade.manova = manova(binned_matrix_2000~as.factor(dataframe[,3]))
clade.wilks = summary(clade.manova,test="Wilks");clade.wilks
```

```
##                         Df   Wilks approx F num Df den Df   Pr(>F)
```

```
## as.factor(dataframe[, 3])  2 0.065312   2.6597     46     42 0.000858 ***
## Residuals                  43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Wilk's Lambda statistic corresponds to a p-value of ~0.0008, suggesting that the clades account for a large, significant proportion of the variance in binned area.

9. LDA — 60 Second Bin Width (3000 time points)

```
lda_info = lda(as.factor(dataframe[,3])~binned_matrix_3000)
lda_info$scaling  #Linear discriminant coefficient matrix
```

```
##                             LD1          LD2
## binned_matrix_30001    2513.38799 -4167.63794
## binned_matrix_30002   -1524.91947  2327.32334
## binned_matrix_30003   -1182.37343  1693.53248
## binned_matrix_30004     810.16567  -300.67936
## binned_matrix_30005     340.25057   412.54361
## binned_matrix_30006     222.42659   -80.54531
## binned_matrix_30007     118.88929    61.06036
## binned_matrix_30008      77.65326   -48.15544
## binned_matrix_30009     181.99095   -10.51186
## binned_matrix_300010    137.47217   -60.40893
## binned_matrix_300011     26.49760  -123.33882
## binned_matrix_300012    868.15305   -46.57229
## binned_matrix_300013  -1150.10789  -228.17895
## binned_matrix_300014    683.00505   292.85770
## binned_matrix_300015   1852.66675   232.20270
```

```
lda_info_p = predict(lda_info,as.data.frame(binned_matrix_3000))
lda_info_pclass = predict(lda_info)$class
par(mar=c(5,4,4,4))
v = as.numeric(lda_info_pclass)
v[which(v==1)] =  "deeppink"    #D
v[which(v==2)] =  12            #F
v[which(v==3)] =  "grey"        #Trochi

plot(lda_info_p$x[,1],pch=16, col = v,main = "LDA - 60 Second Bin Width",xlab="df1",ylab="df2") #Plot t.
legend("topright",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deepp:
```

LDA – 60 Second Bin Width
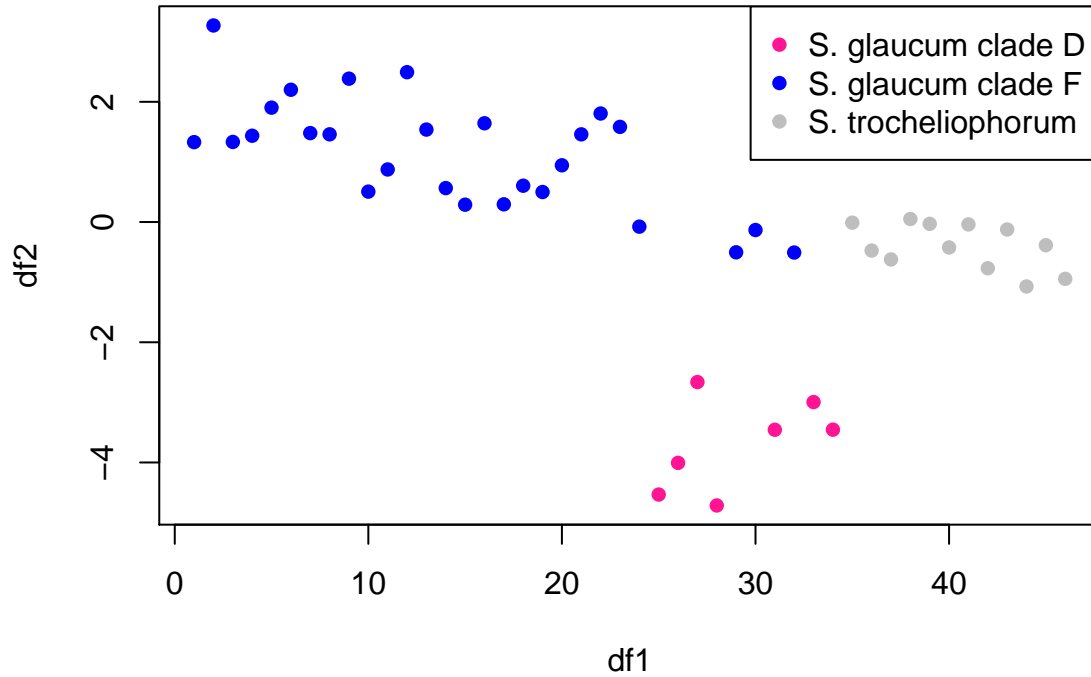

```
clade.manova = manova(binned_matrix_3000~as.factor(dataframe[,3]))
clade.wilks = summary(clade.manova,test="Wilks");clade.wilks
```

```
##                          Df   Wilks approx F num Df den Df    Pr(>F)
## as.factor(dataframe[, 3])  2 0.11567   3.7513     30     58 7.922e-06 ***
## Residuals                 43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Wilk's Lambda statistic corresponds to a p-value of ~0, suggesting that the clades account for a large, significant proportion of the variance in binned area.

APPENDIX

Experimentation with different bin sizes:

```
binned_matrix_1100 = binner(dataframe[,15000:62004],1100, DELTA)
binned_matrix_1300 = binner(dataframe[,15000:62004],1300, DELTA)
binned_matrix_1500 = binner(dataframe[,15000:62004],1500, DELTA)
binned_matrix_1600 = binner(dataframe[,15000:62004],1600, DELTA)
binned_matrix_1800 = binner(dataframe[,15000:62004],1800, DELTA)
binned_matrix_2000 = binner(dataframe[,15000:62004],2000, DELTA)
```
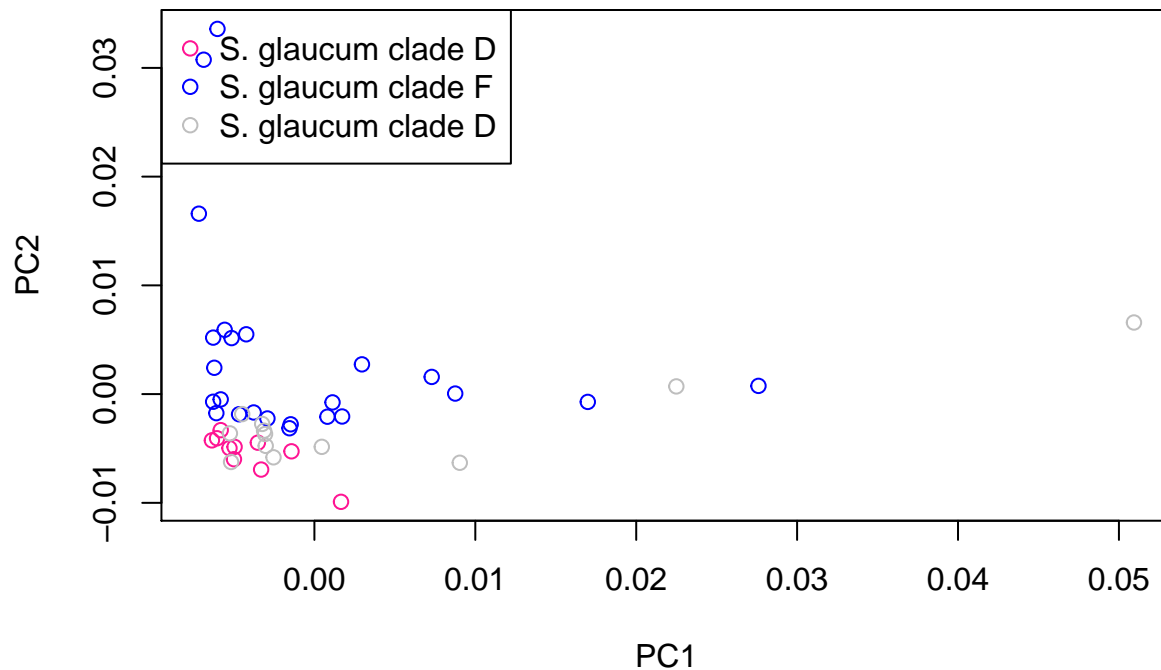
```
trapAreas.centered = scale(binned_matrix_1100, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:42] 0.01071 0.0083 0.00596 0.00487 0.00413 ...
##  $ rotation: num [1:42, 1:42] -0.0229 -0.0247 -0.0144 -0.0218 -0.0221 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:42] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:42] 8.67e-19 1.13e-19 -3.02e-19 1.51e-19 3.02e-19 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:42] 0.000809 0.027598 -0.004235 -0.007181 -0.001474 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:42] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```
PC = pca.result$x

plot(PC[,1],PC[,2],col="white",main = "22 Second Bin Width (1100 time points)",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. glaucum clade D"),col=c("deeppi
```
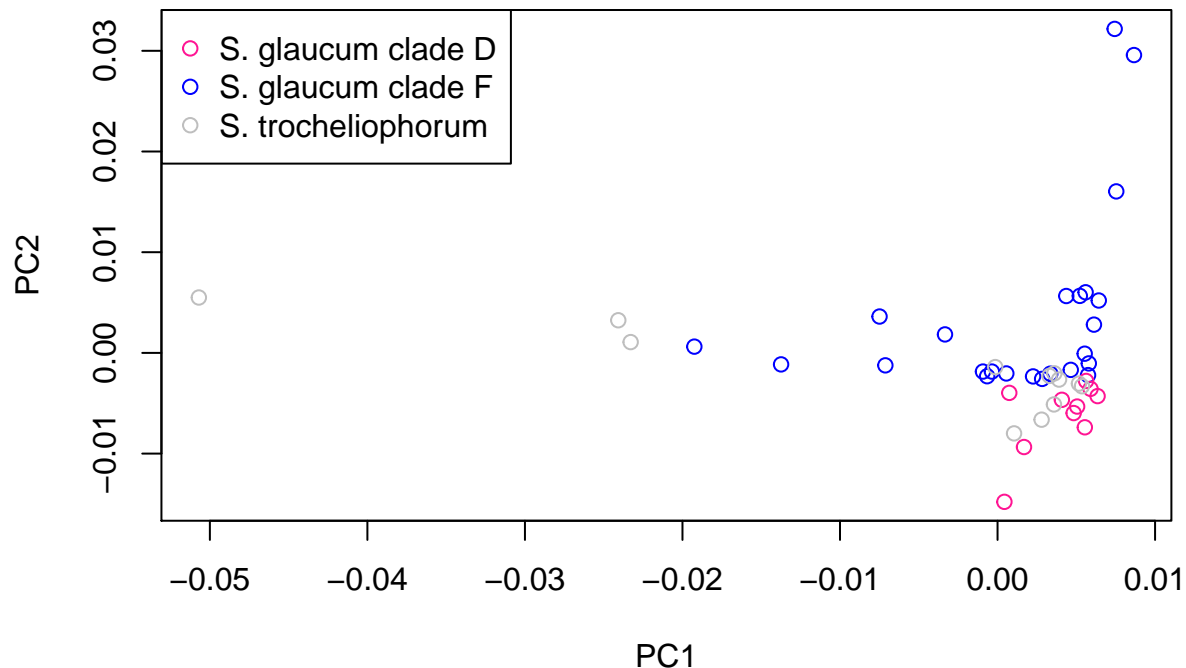
22 Second Bin Width (1100 time points)

```
trapAreas.centered = scale(binned_matrix_1300, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:36] 0.01073 0.00827 0.00659 0.00598 0.00473 ...
##  $ rotation: num [1:36, 1:36] 0.0291 0.0269 0.0191 0.0256 0.0279 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:36] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:36] -7.54e-20 4.90e-19 -6.79e-19 1.89e-19 -3.77e-20 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:36] -0.000359 -0.019246 0.004364 0.007537 0.002256 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:36] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```
PC = pca.result$x
```

```
plot(PC[,1],PC[,2],col="white",main = "26 Second Bin Width (1300 time points)",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```
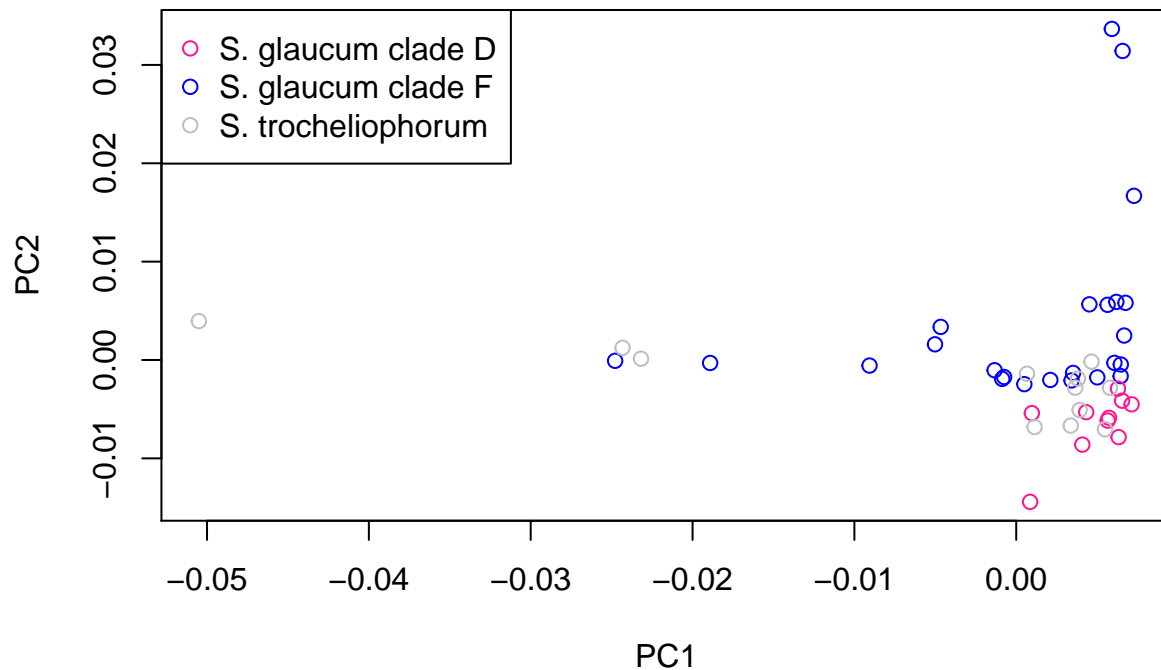
26 Second Bin Width (1300 time points)

```
trapAreas.centered = scale(binned_matrix_1500, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:31] 0.01124 0.00857 0.00662 0.00636 0.005 ...
##  $ rotation: num [1:31, 1:31] 0.0321 0.0279 0.0248 0.0302 0.0285 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:31] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:31] -6.03e-19 -2.26e-19 -3.77e-19 -1.21e-18 -6.79e-19 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:31] -0.000749 -0.024784 0.004518 0.007275 0.002107 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:31] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```
PC = pca.result$x
```

```
plot(PC[,1],PC[,2],col="white",main = "30 Second Bin Width (1500 time points)",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```
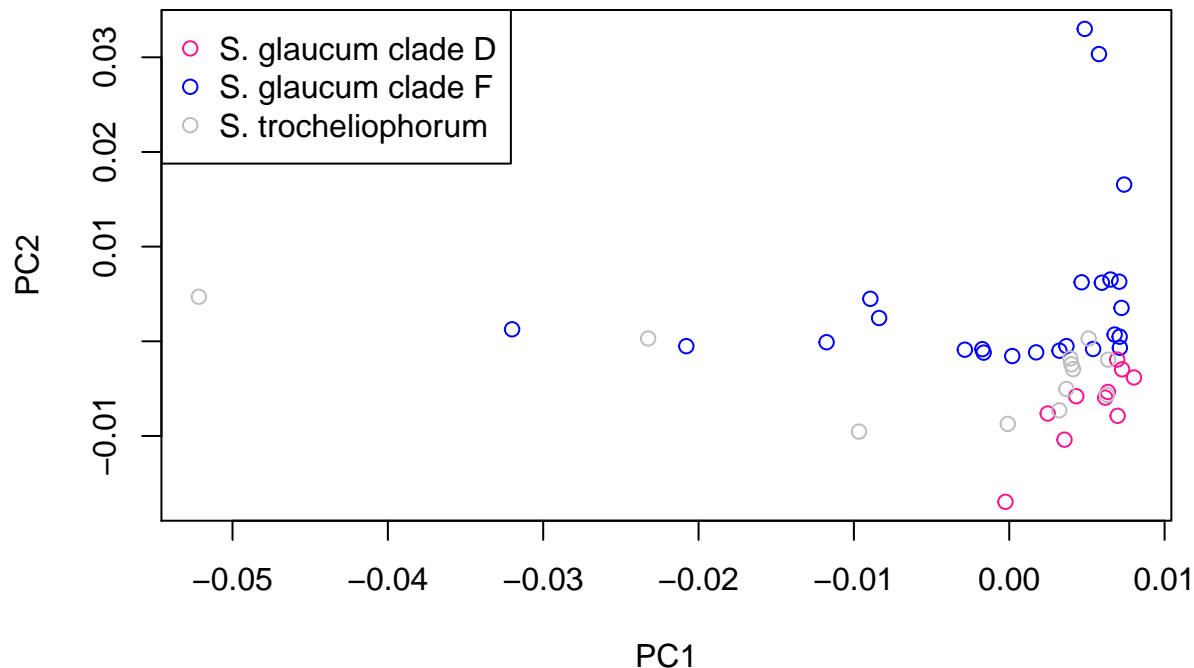

30 Second Bin Width (1500 time points)

```
trapAreas.centered = scale(binned_matrix_1600, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
## $ sdev    : num [1:29] 0.01168 0.00876 0.0071 0.00505 0.00491 ...
## $ rotation: num [1:29, 1:29] 0.0324 0.0275 0.0266 0.0317 0.0287 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:29] "PC1" "PC2" "PC3" "PC4" ...
## $ center  : num [1:29] -1.51e-18 1.58e-18 1.28e-18 1.51e-19 -5.28e-19 ...
## $ scale   : logi FALSE
## $ x       : num [1:46, 1:29] -0.00165 -0.032 0.00466 0.00739 0.00173 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:29] "PC1" "PC2" "PC3" "PC4" ...
## - attr(*, "class")= chr "prcomp"
```

```
PC = pca.result$x
```

```
plot(PC[,1],PC[,2],col="white",main = "32 Second Bin Width (1600 time points)",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```


32 Second Bin Width (1600 time points)

```
trapAreas.centered = scale(binned_matrix_1800, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:26] 0.01213 0.00748 0.00695 0.00482 0.00464 ...
##  $ rotation: num [1:26, 1:26] 0.0398 0.0315 0.0369 0.0373 0.0372 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:26] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:26] -1.13e-18 1.66e-18 -3.77e-19 4.53e-19 -1.36e-18 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:26] 0.000902 -0.023019 0.006769 0.009209 0.004085 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:26] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```
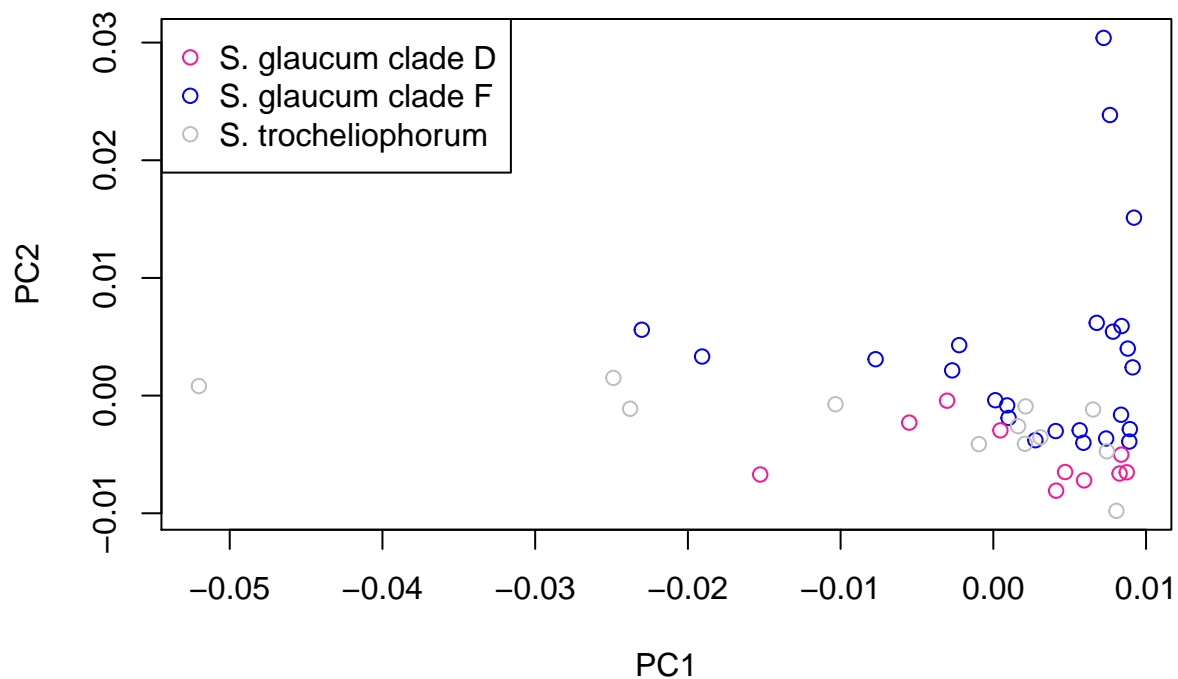
```
PC = pca.result$x

plot(PC[,1],PC[,2],col="white",main = "36 Second Bin Width (1800 time points)",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```


36 Second Bin Width (1800 time points)

```
trapAreas.centered = scale(binned_matrix_2000, center = TRUE, scale = FALSE)
pca.result <- prcomp(trapAreas.centered)
str(pca.result)
```

```
## List of 5
##  $ sdev    : num [1:23] 0.01264 0.00829 0.00793 0.00503 0.00459 ...
##  $ rotation: num [1:23, 1:23] 0.0433 0.0352 0.0407 0.0407 0.0362 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:23] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : num [1:23] -1.21e-18 -6.79e-19 -1.13e-18 -4.53e-19 9.80e-19 ...
##  $ scale   : logi FALSE
##  $ x       : num [1:46, 1:23] 0.00152 -0.02339 0.00685 0.01005 0.00481 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:46] "col" "col" "col" "col" ...
##   .. ..$ : chr [1:23] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```
PC = pca.result$x
```

```
plot(PC[,1],PC[,2],col="white",main = "40 Second Bin Width (2000 time points)",xlab="PC1",ylab="PC2")
points(PC[F,1],PC[F,2],col="blue")
points(PC[D,1],PC[D,2],col="deeppink")
points(PC[T,1],PC[T,2],col="grey")
legend("topleft",legend=c("S. glaucum clade D","S. glaucum clade F","S. trocheliophorum"),col=c("deeppi
```

40 Second Bin Width (2000 time points)
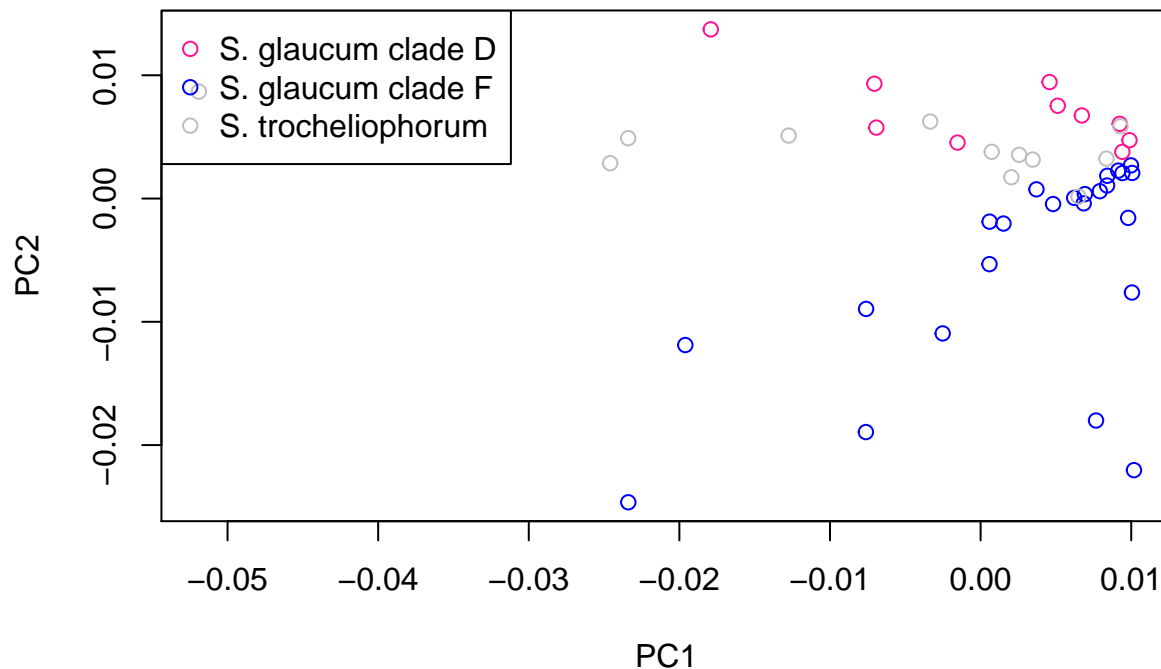

```
barplot(binned_matrix_2000[26,],col="skyblue",main="Sample 29",ylab = "Area",xlab="Bin")
```

```
sum(binned_matrix_2000[25,])
```