Differential Expression of *Phaeodactylum tricornutum* Grown in Produced Water

By:

Claudia Vo

Codey Phoun

Sudhin Domala

For:

SJSU CS286 Project

Dr. Andreopoulos

Spring 2021

Introduction

Sample abbreviations

HQE = normal media exponential growth (control)

HQ10E = normal media + added 10% produced water (intermediate treatment)

PWE = 100% produced water exponential (High treatment)

HQ_ST = normal media stationary growth (control)

PW_ST = 100% produced water stationary (treatment)

Setup

Import libraries

```
library(readr)
library(edgeR)
```

```
## Loading required package: limma
```

```
library(limma)
library(pheatmap)
```

Gene count data was previously created with STAR

Read in the count data as a data frame

```
URL = "https://raw.githubusercontent.com/codey-phoun/pw_oilfield/main/STAR_results/STAR_counts.txt"
star_data = read_tsv(URL)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   gene_name = col_character(),
##   HQ10E1 = col_double(),
##   HQ10E2 = col_double(),
##   HQE1 = col_double(),
##   HQE2 = col_double(),
##   HQ_ST1 = col_double(),
##   HQ_ST2 = col_double(),
##   PWE1 = col_double(),
##   PWE2 = col_double(),
##   PW_ST1 = col_double(),
##   PW_ST2 = col_double()
## )
```

Create count matrix

```
star_data = as.data.frame(star_data)
# Set row names to be the gene name and remove gene_name column
rownames(star_data) <- star_data$gene_name
star_data[1] <- NULL
head(star_data)
```

```
##               HQ10E1 HQ10E2 HQE1 HQE2 HQ_ST1 HQ_ST2 PWE1 PWE2 PW_ST1 PW_ST2
## Phatr3_J31400      1      5    3    7      0      4    2    1     10     20
## Phatr3_J42422    144    172  214  153    402    451  117  101    203    253
## Phatr3_J31402      0      0    0    0      0      0    0    0      0      0
## Phatr3_J42423     27     44   77   42    557    508   46   39    110    304
## Phatr3_J42424   1819   2172 1876 1254   1524   1841 1518 1508   1035   1562
## Phatr3_J7430     455    510  620  443    467    513  333  456    337    364
```

Create a DGE list object

```
group <- rep(c("HQ10E","HQE","HQ_ST","PWE","PW_ST"),
             each = 2)

dge <- DGEList(counts = star_data,
               group = group) #creates a DGE list object
dim(dge)
```

```
## [1] 12392    10
```

```
full_dge <- dge #store original data just in case
```

Sample library sizes

```
apply(dge$counts, 2, sum) # sum across columns/samples for each gene
```

```
##   HQ10E1   HQ10E2     HQE1     HQE2   HQ_ST1   HQ_ST2     PWE1     PWE2
## 19654912 22355112 22447786 16151357 15135335 18504345 18680923 18905705
##   PW_ST1   PW_ST2
## 17778297 21967279
```

Filtering & Normalizing data

Counts per million (cpm)

```
head(cpm(dge))
```

```
##                    HQ10E1     HQ10E2       HQE1       HQE2    HQ_ST1     HQ_ST2
## Phatr3_J31400  0.05087787  0.2236625  0.1336435  0.4334001   0.00000  0.2161654
## Phatr3_J42422  7.32641286  7.6939896  9.5332341  9.4728883  26.56036 24.3726541
## Phatr3_J31402  0.00000000  0.0000000  0.0000000  0.0000000   0.00000  0.0000000
## Phatr3_J42423  1.37370241  1.9682299  3.4301824  2.6004007  36.80130 27.4530117
## Phatr3_J42424 92.54684020 97.1589854 83.5717162 77.6405351 100.69153 99.4901468
## Phatr3_J7430  23.14942952 22.8135739 27.6196503 27.4280359  30.85495 27.7232185
##                     PWE1        PWE2     PW_ST1    PW_ST2
## Phatr3_J31400  0.1070611  0.05289409  0.5624836  0.910445
## Phatr3_J42422  6.2630738  5.34230276 11.4184165 11.517130
## Phatr3_J31402  0.0000000  0.00000000  0.0000000  0.000000
## Phatr3_J42423  2.4624051  2.06286938  6.1873193 13.838764
## Phatr3_J42424 81.2593682 79.76428279 58.2170497 71.105757
## Phatr3_J7430  17.8256717 24.11970355 18.9556964 16.570100
```

Keep only 100 counts per million in at least 2 samples

```
keep <- rowSums(cpm(dge)>100) >= 2
table(keep)
```

```
## keep
## FALSE  TRUE
##  9141  3251
```

```
dge <- dge[keep,]
dim(dge)  #check number of genes left after filtering
```

```
## [1] 3251   10
```

12392 genes are filtered down to 3251 genes

Reset the library size

```
dge$samples$lib.size <- colSums(dge$counts)
dge$samples
```

```
##        group lib.size norm.factors
## HQ10E1 HQ10E 15289361            1
## HQ10E2 HQ10E 17319367            1
## HQE1     HQE 17111487            1
## HQE2     HQE 12568283            1
## HQ_ST1 HQ_ST 11248709            1
## HQ_ST2 HQ_ST 13869364            1
## PWE1     PWE 14606604            1
## PWE2     PWE 14925823            1
## PW_ST1 PW_ST 14619994            1
## PW_ST2 PW_ST 17756624            1
```

Library sizes before filtering:

```
apply(full_dge$counts, 2, sum)
```

```
##   HQ10E1   HQ10E2     HQE1     HQE2   HQ_ST1   HQ_ST2     PWE1     PWE2
## 19654912 22355112 22447786 16151357 15135335 18504345 18680923 18905705
##   PW_ST1   PW_ST2
## 17778297 21967279
```

Library sizes after filtering:

```
apply(dge$counts, 2, sum)
```

```
##   HQ10E1   HQ10E2     HQE1     HQE2   HQ_ST1   HQ_ST2     PWE1     PWE2
## 15289361 17319367 17111487 12568283 11248709 13869364 14606604 14925823
##   PW_ST1   PW_ST2
## 14619994 17756624
```

Normalize data by the trimmed mean of M-values (TMM) method proposed by Robinson and Oshlack (2010)

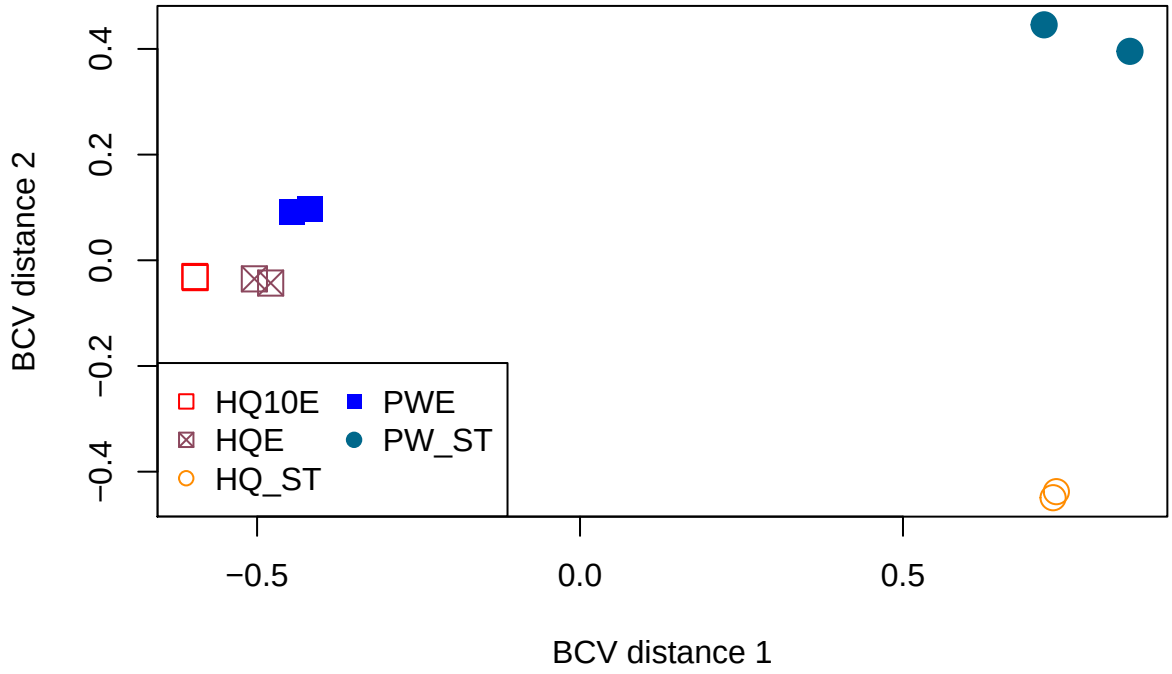
```
dge_norm=calcNormFactors(dge, method="TMM")
dge_norm
```

```
## An object of class "DGEList"
## $counts
##                HQ10E1 HQ10E2 HQE1 HQE2 HQ_ST1 HQ_ST2 PWE1 PWE2 PW_ST1 PW_ST2
## Phatr3_J42426    8520   9594 8689 7289   5396   7052 6769 7244   5813   8494
## Phatr3_EG02408   3806   4483 3948 2637    224    238 3580 3084    216    368
## Phatr3_J31409    6448   7590 9257 5980   1008   1394 4863 4622   1381   1864
## Phatr3_J42429    1747   1981 2134 1377   2150   2386 1650 1699   1427   1398
## Phatr3_J4937     6608   7884 7049 4583   3737   5446 6837 6271   2997   3700
## 3246 more rows ...
##
## $samples
##        group lib.size norm.factors
## HQ10E1 HQ10E 15289361    1.0347646
## HQ10E2 HQ10E 17319367    1.0442405
## HQE1     HQE 17111487    1.0884649
## HQE2     HQE 12568283    1.0460172
```

```
## HQ_ST1 HQ_ST 11248709    1.0977999
## HQ_ST2 HQ_ST 13869364    1.0696829
## PWE1     PWE 14606604    1.0251548
## PWE2     PWE 14925823    1.0196206
## PW_ST1 PW_ST 14619994    0.8029116
## PW_ST2 PW_ST 17756624    0.8247657
```

Multidimensional scaling plot to look at the inter-sample relationship by biological coefficient of variation (BCV) distance

```
colors_mds <- c("red", "palevioletred4", "darkorange", "blue", "deepskyblue4")
plotMDS(dge_norm, method="bcv", col=rep(colors_mds,each=2), pch = rep(c(0,7,1,15,19),each=2), cex = 1.75
legend("bottomleft",as.character(unique(dge_norm$samples$group)),col=colors_mds,pch=c(0,7,1,15,19), nco
```



Exponential growth samples cluster together stronger than stationary growth

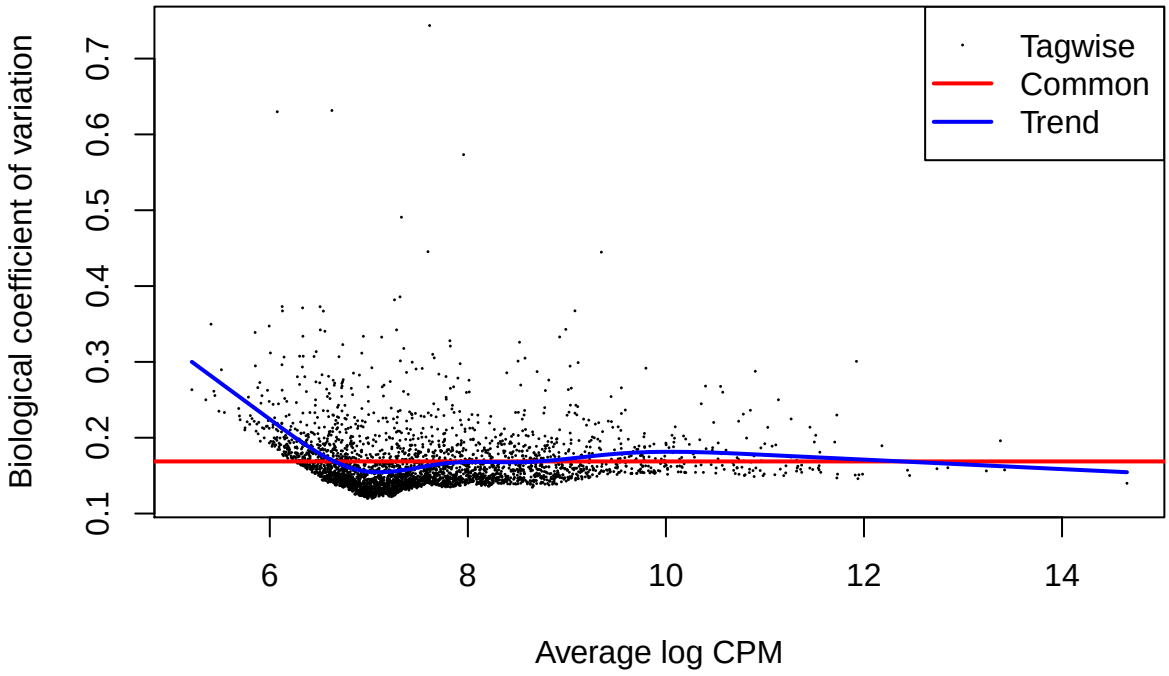
Samples do not tend to cluster based on growth medium

Create the design matrix

```
design.mat <- model.matrix(~ 0 + dge_norm$samples$group)
colnames(design.mat) <- levels(dge_norm$samples$group)
```

Estimate the dispersion with Cox-Reid profile-adjusted likelihood (CR) method in estimating dispersions with Generalized linear models (GLMs)

```
d2 <- estimateGLMCommonDisp(dge_norm,design.mat)
d2 <- estimateGLMTrendedDisp(d2,design.mat, method="auto")
# You can change method to "auto", "bin.spline", "power", "spline", "bin.loess".
# The default is "auto" which chooses "bin.spline" when > 200 tags and "power" otherwise.
d2 <- estimateGLMTagwiseDisp(d2,design.mat)
plotBCV(d2)
```



Calculate log2 CPM values

```
logcpm <- cpm(d2, log = TRUE)
```

Create a matrix of contrasts for anova-like testing (since data has different conditions that we want to compare)

ANOVA tests for DEGs between any set of groups with the null hypothesis that the mean gene expression is equal across all groups.

```
my_contrasts<-makeContrasts(
  HQst_vs_PWst = HQ_ST-PW_ST, #PW vs normal stationary growth samples

  HQE_vs_PWE = HQE-PWE,    # PW vs normal exponential growth samples,

  HQST_vs_HQE = HQ_ST-HQE, #normal exponential vs normal stationary

  PWE_vs_PWST = PWE-PW_ST, #100% exponential vs 100% stationary PW samples
```

```
  levels= design.mat
)
my_contrasts
```

```
##        Contrasts
## Levels  HQst_vs_PWst HQE_vs_PWE HQST_vs_HQE PWE_vs_PWST
##   HQ_ST            1          0           1           0
##   HQ10E            0          0           0           0
##   HQE              0          1          -1           0
##   PW_ST           -1          0           0          -1
##   PWE              0         -1           0           1
```

Fit a quasi-likelihood negative binomial generalized log-linear model to count data

```
fit <- glmQLFit(d2, design.mat)
```

Normal Medium vs 100% Produced Water in Stationary Growth Samples

```
HQst_vs_PWst <- glmQLFTest(fit, contrast = my_contrasts[,"HQst_vs_PWst"])
# top 50 DEGs by lowest adjusted p-values
HQst_vs_PWst_top50 <- topTags(HQst_vs_PWst,adjust.method = "BH", p.value = 0.05, n = 50)
HQst_vs_PWst_all <- topTags(HQst_vs_PWst,adjust.method = "BH", p.value = 0.05, n = nrow(HQst_vs_PWst$tal
HQst_vs_PWst_DEG <- HQst_vs_PWst_all[abs(HQst_vs_PWst_all$table$logFC) > 1, ]
HQst_vs_PWst_up <- HQst_vs_PWst_all[HQst_vs_PWst_all$table$logFC > 1, ]
HQst_vs_PWst_down <- HQst_vs_PWst_all[HQst_vs_PWst_all$table$logFC < -1, ]
head(HQst_vs_PWst_top50, n=10)
```

```
## Coefficient:  1*HQ_ST -1*PW_ST
##                     logFC    logCPM        F       PValue          FDR
## Phatr3_J47572    6.861896 10.355303 830.2771 7.879261e-10 1.438429e-06
## Phatr3_J50055   -4.296119  6.983754 711.6102 1.515469e-09 1.438429e-06
## Phatr3_J55010    5.536678  9.496214 633.6890 2.476507e-09 1.438429e-06
## Phatr3_J45193   -3.819702  9.779135 619.5537 2.724555e-09 1.438429e-06
## Phatr3_EG00333   5.960957  8.190521 614.3864 2.822823e-09 1.438429e-06
## Phatr3_J49202   -6.373400 12.846728 609.1734 2.926430e-09 1.438429e-06
## Phatr3_J10640  -10.070021  8.076858 601.0593 3.097202e-09 1.438429e-06
## Phatr3_J32747    4.480649  8.734087 537.9812 4.948332e-09 1.826148e-06
## Phatr3_J48882  -10.564538 13.377617 533.0443 5.144764e-09 1.826148e-06
## Phatr3_EG00065  -4.133901  9.136149 522.0666 5.617190e-09 1.826148e-06
```

Determine how many genes are up and down regulated for each pairwise comparison

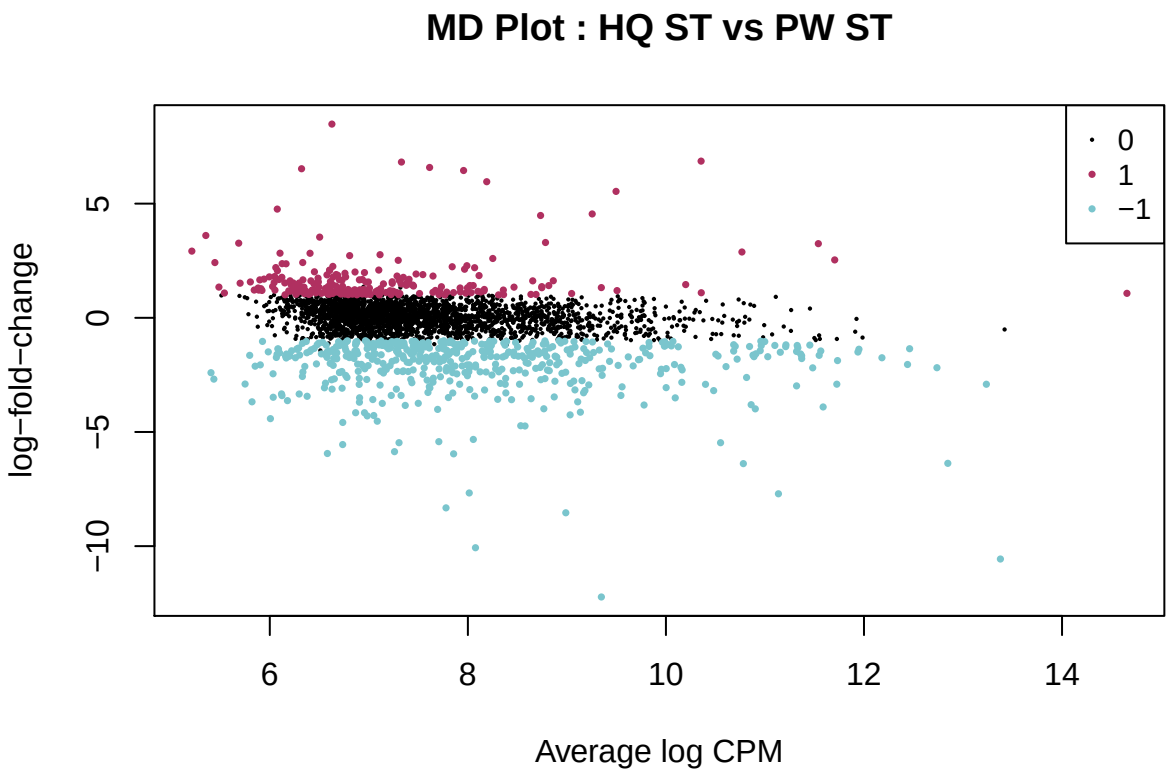
For differentially expressed genes set logFC and pvalue threshold to 0.05 and 1

lfc=1 sets a 2-fold change minimum

```
is.de <- decideTestsDGE(HQst_vs_PWst,adjust.method="BH",p.value=0.05,lfc=1)
summary(is.de)
```

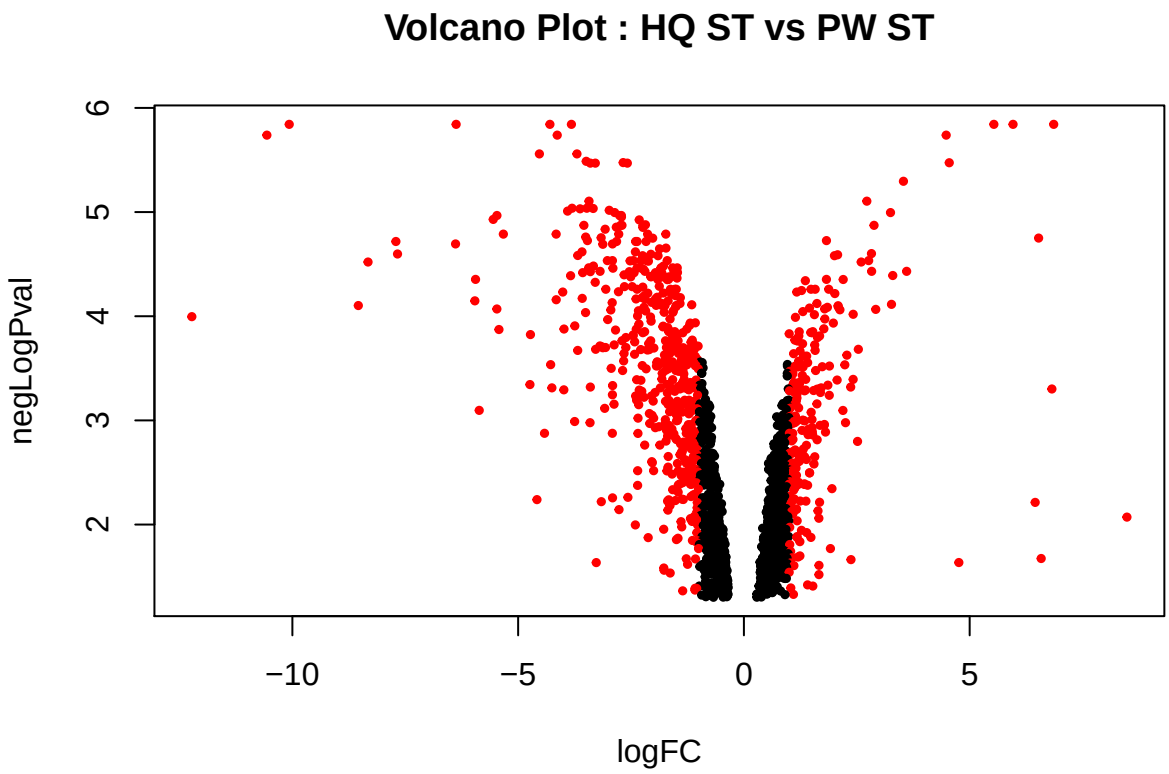
```
##        1*HQ_ST -1*PW_ST
## Down                509
## NotSig             2515
## Up                  227
```

```
plotMD(HQst_vs_PWst, status=is.de, values=c(1,-1), col=c("maroon","cadetblue3"),
       legend="topright", cex = .5, main = "MD Plot : HQ ST vs PW ST")
```



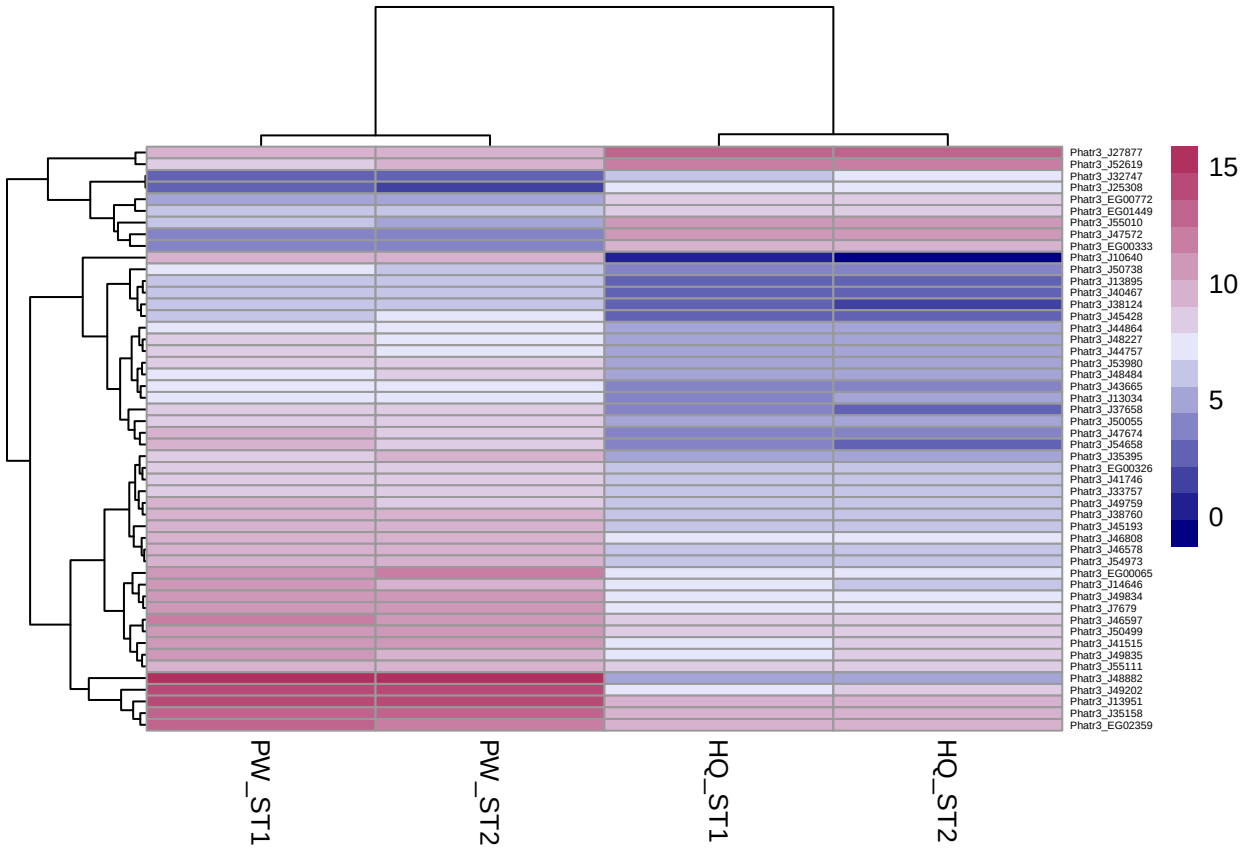
Volcano Plot

```
volcanoData <- cbind(HQst_vs_PWst_all$table$logFC, -log10(HQst_vs_PWst_all$table$FDR))
colnames(volcanoData) <- c("logFC", "negLogPval")
DEGs <- HQst_vs_PWst_all$table$FDR < 0.05 & abs(HQst_vs_PWst_all$table$logFC) > 1
point.col <- ifelse(DEGs, "red", "black")
plot(volcanoData, pch=19, cex = .5, col = point.col, main = "Volcano Plot : HQ ST vs PW ST")
```




```
HQst_vs_PWst_top50_log2_cpm <- logcpm[rownames(HQst_vs_PWst_top50$table),]

pheatmap(subset(HQst_vs_PWst_top50_log2_cpm,select=c(HQ_ST1,HQ_ST2,PW_ST1,PW_ST2)),
         color=colorRampPalette(c("navy", "lavender", "maroon"))(15),fontsize_row=4)
```



Normal Medium vs 100% Produced Water in Exponential Growth Samples

```
HQE_vs_PWE <- glmQLFTest(fit, contrast = my_contrasts[,"HQE_vs_PWE"])
HQE_vs_PWE_top50 = topTags(HQE_vs_PWE,adjust.method = "BH", p.value = 0.05, n = 50)
HQE_vs_PWE_all = topTags(HQE_vs_PWE,adjust.method = "BH", p.value = 0.05, n = nrow(HQE_vs_PWE$table))
HQE_vs_PWE_DEG <- HQE_vs_PWE_all[abs(HQE_vs_PWE_all$table$logFC) > 1, ]
HQE_vs_PWE_up <- HQE_vs_PWE_all[HQE_vs_PWE_all$table$logFC > 1, ]
HQE_vs_PWE_down <- HQE_vs_PWE_all[HQE_vs_PWE_all$table$logFC < -1, ]
head(HQE_vs_PWE_top50, n=10)
```

```
## Coefficient:  1*HQE -1*PWE
##                     logFC    logCPM        F       PValue          FDR
## Phatr3_J34132  -8.609962  8.008770 727.2847 1.381801e-09 4.492234e-06
## Phatr3_J55010  -5.111485  9.496214 561.2096 4.139153e-09 6.323839e-06
## Phatr3_J49151   3.607201 10.200579 498.9501 6.800095e-09 6.323839e-06
## Phatr3_EG00333 -5.001485  8.190521 483.2628 7.780792e-09 6.323839e-06
## Phatr3_J15393  -5.145494  7.119785 417.1914 1.445075e-08 9.395875e-06
## Phatr3_J31433  -3.291023 10.002898 296.6263 6.035273e-08 3.270112e-05
## Phatr3_EG02360  6.246511  9.360258 268.4197 9.154341e-08 4.251538e-05
## Phatr3_J50500  -2.371713  9.723243 241.8646 1.411624e-07 5.619997e-05
## Phatr3_J51092  -3.844156  7.877635 236.2610 1.555828e-07 5.619997e-05
## Phatr3_J31619  -1.785251  7.409644 217.5707 2.188898e-07 7.116106e-05
```

Determine how many genes are up and down regulated for each pairwise comparison

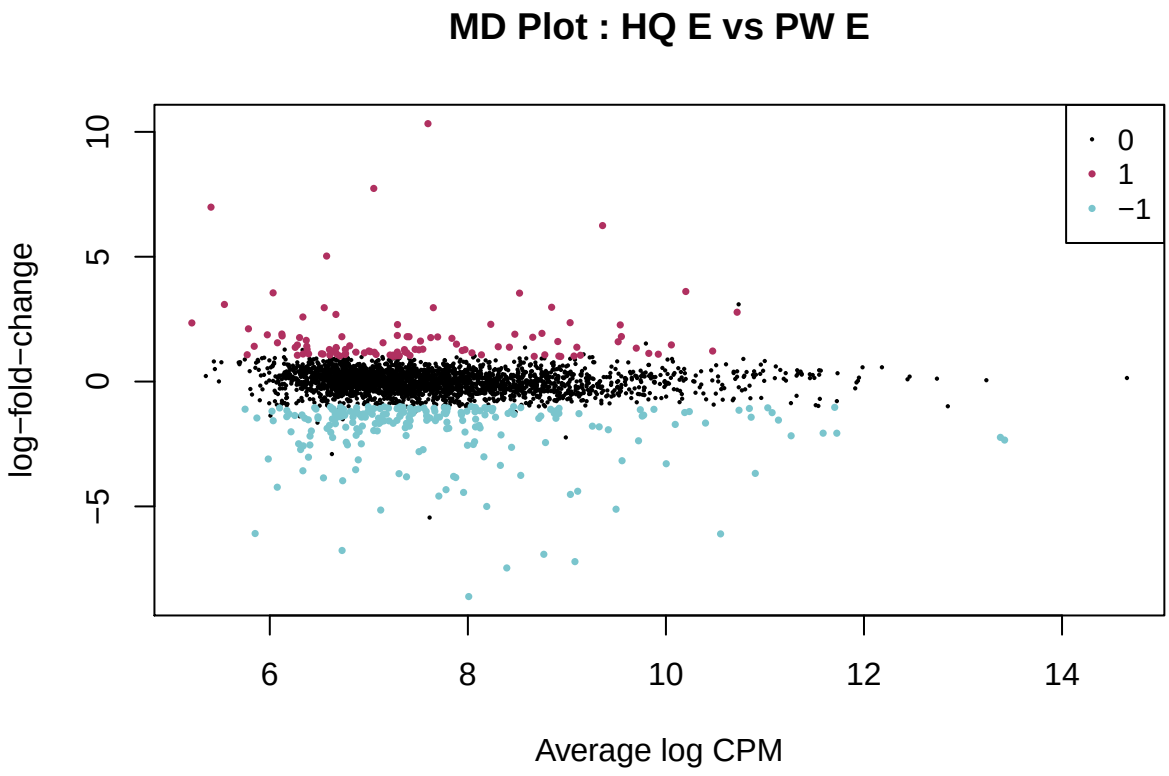
For differentially expressed genes set logFC and pvalue threshold to 0.05 and 1

lfc=1 sets a 2-fold change minimum

```
is.de <- decideTestsDGE(HQE_vs_PWE,adjust.method="BH",p.value=0.05,lfc=1)
summary(is.de)
```

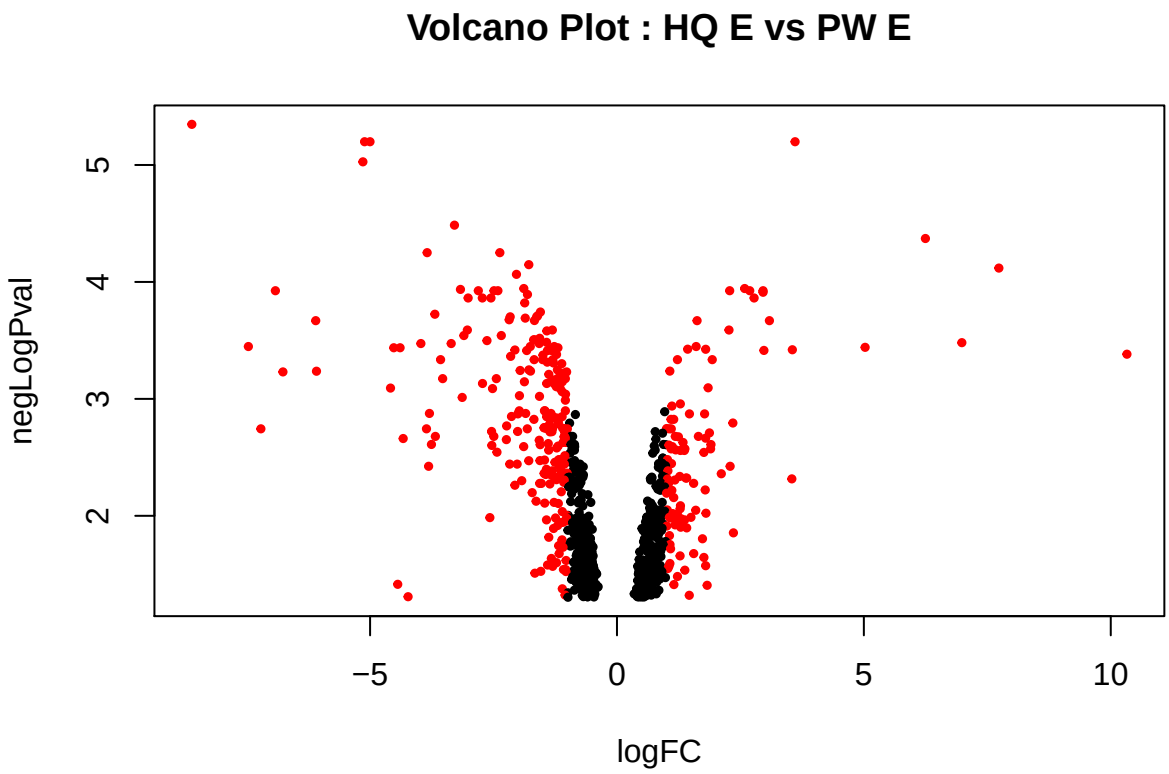
```
##        1*HQE -1*PWE
## Down             223
## NotSig          2926
## Up               102
```

```
plotMD(HQE_vs_PWE, status=is.de, values=c(1,-1), col=c("maroon","cadetblue3"),
       legend="topright", cex = .5, main = "MD Plot : HQ E vs PW E")
```



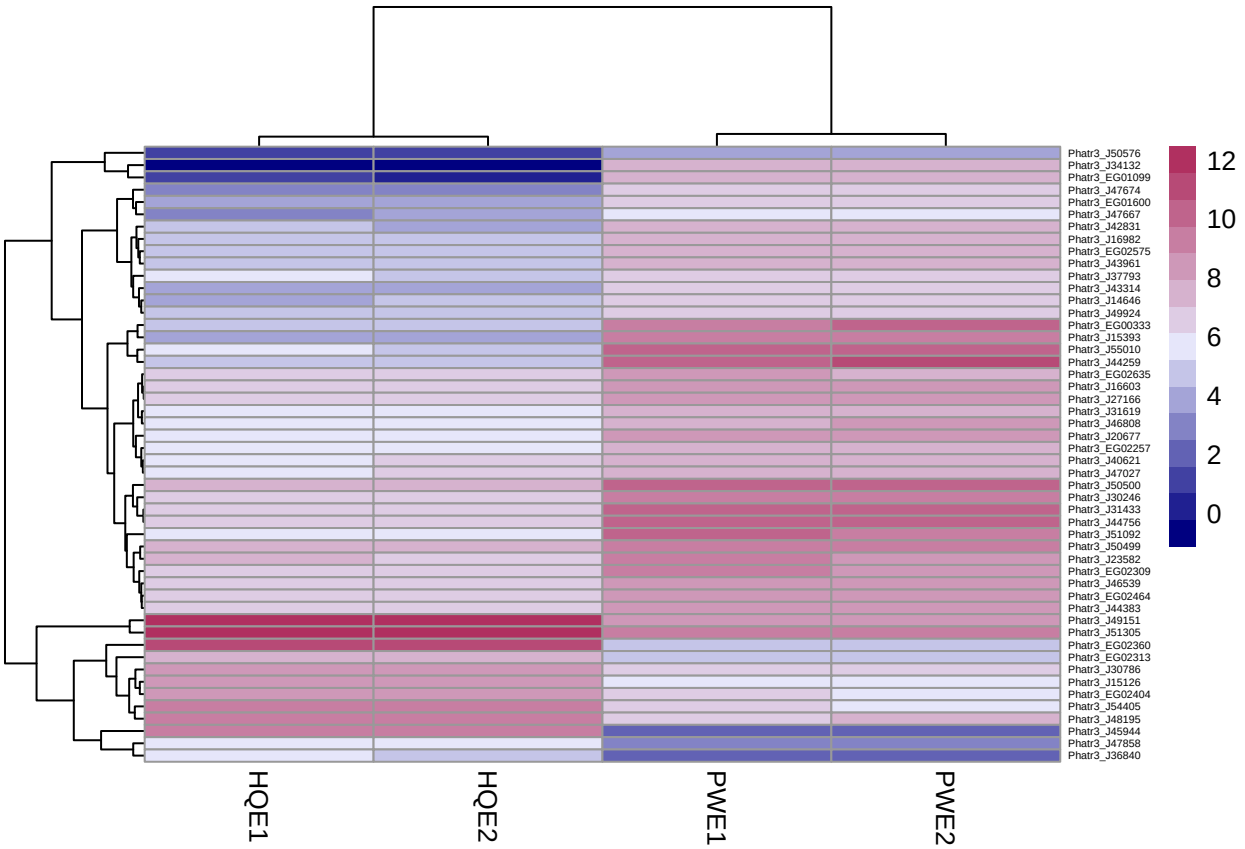
Volcano Plot

```
volcanoData <- cbind(HQE_vs_PWE_all$table$logFC, -log10(HQE_vs_PWE_all$table$FDR))
colnames(volcanoData) <- c("logFC", "negLogPval")
DEGs <- HQE_vs_PWE_all$table$FDR < 0.05 & abs(HQE_vs_PWE_all$table$logFC) > 1
point.col <- ifelse(DEGs, "red", "black")
plot(volcanoData, pch=19, cex = .5, col = point.col, main = "Volcano Plot : HQ E vs PW E")
```




```
HQE_vs_PWE_top50_log2_cpm <- logcpm[rownames(HQE_vs_PWE_top50$table),]

pheatmap(subset(HQE_vs_PWE_top50_log2_cpm,select=c(HQE1,HQE2,PWE1,PWE2)),
         color=colorRampPalette(c("navy", "lavender", "maroon"))(15),fontsize_row=4)
```



Normal Medium Exponential Growth vs Normal Medium in Stationary Growth

```
HQST_vs_HQE <- glmQLFTest(fit, contrast = my_contrasts[,"HQST_vs_HQE"])
HQST_vs_HQE_top50 = topTags(HQST_vs_HQE,adjust.method = "BH", p.value = 0.05, n = 50)
HQST_vs_HQE_all = topTags(HQST_vs_HQE,adjust.method = "BH", p.value = 0.05, n = nrow(HQST_vs_HQE$table))
HQST_vs_HQE_DEG <- HQST_vs_HQE_all[abs(HQST_vs_HQE_all$table$logFC) > 1, ]
HQST_vs_HQE_up <- HQST_vs_HQE_all[HQST_vs_HQE_all$table$logFC > 1, ]
HQST_vs_HQE_down <- HQST_vs_HQE_all[HQST_vs_HQE_all$table$logFC < -1, ]
head(HQST_vs_HQE_top50, n=10)
```

```
## Coefficient:  1*HQ_ST -1*HQE
##                    logFC    logCPM         F       PValue          FDR
## Phatr3_J23830   6.286260  8.103721 1546.1159 5.594732e-11 1.818847e-07
## Phatr3_J47667  11.239932 13.420073 1249.2782 1.387318e-10 2.014103e-07
## Phatr3_J40433   7.752019 14.655511 1162.2726 1.886329e-10 2.014103e-07
## Phatr3_J15126 -11.632476  6.667419  971.6138 4.041445e-10 2.014103e-07
## Phatr3_J40467  -5.679432  7.803590  965.0569 4.159431e-10 2.014103e-07
## Phatr3_J48511   4.590450  9.653790  930.4874 4.856776e-10 2.014103e-07
## Phatr3_J46796   9.157803  9.775273  924.0527 5.002087e-10 2.014103e-07
## Phatr3_J46395  -6.934256  8.292089  885.7125 5.988517e-10 2.014103e-07
## Phatr3_J10068  -6.734953  6.486422  878.3977 6.203162e-10 2.014103e-07
## Phatr3_J48834   4.924547  6.469906  861.1280 6.748816e-10 2.014103e-07
```

Determine how many genes are up and down regulated for each pairwise comparison

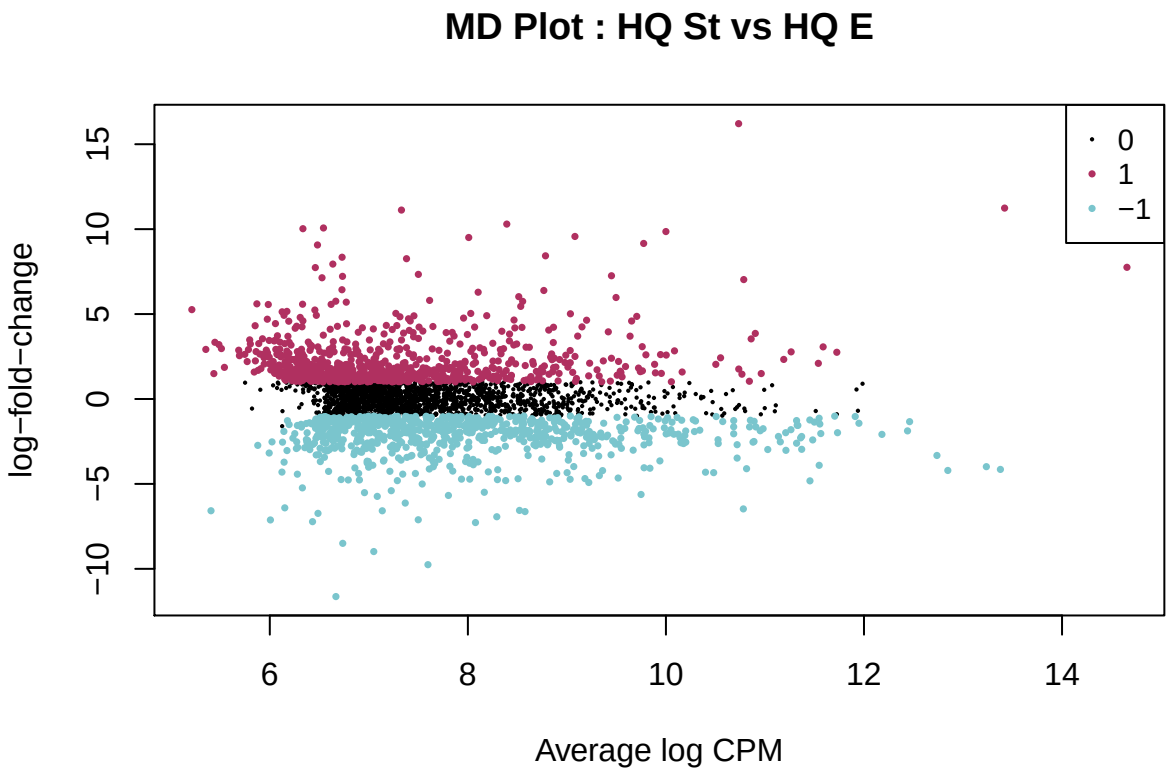
For differentially expressed genes set logFC and pvalue threshold to 0.05 and 1

lfc=1 sets a 2-fold change minimum

```
is.de <- decideTestsDGE(HQST_vs_HQE,adjust.method="BH",p.value=0.05,lfc=1)
summary(is.de)
```

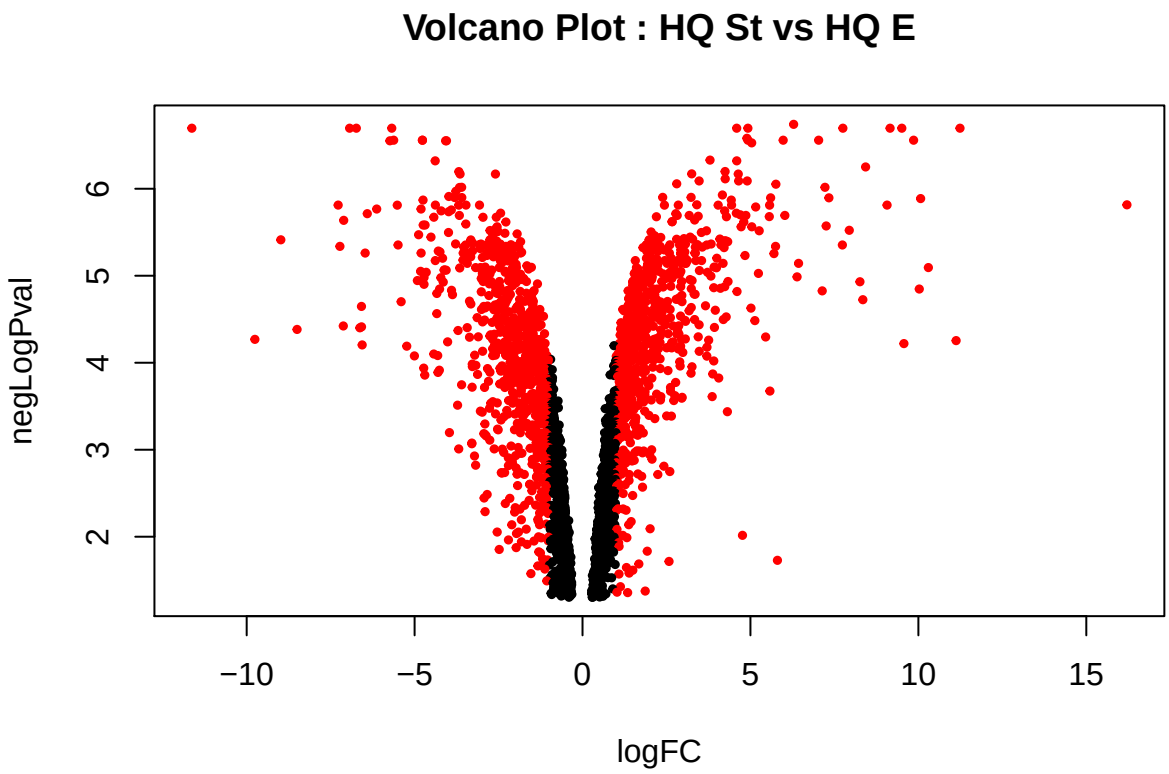
```
##        1*HQ_ST -1*HQE
## Down               765
## NotSig            1734
## Up                 752
```

```
plotMD(HQST_vs_HQE, status=is.de, values=c(1,-1), col=c("maroon","cadetblue3"),
       legend="topright", cex = .5, main = "MD Plot : HQ St vs HQ E")
```



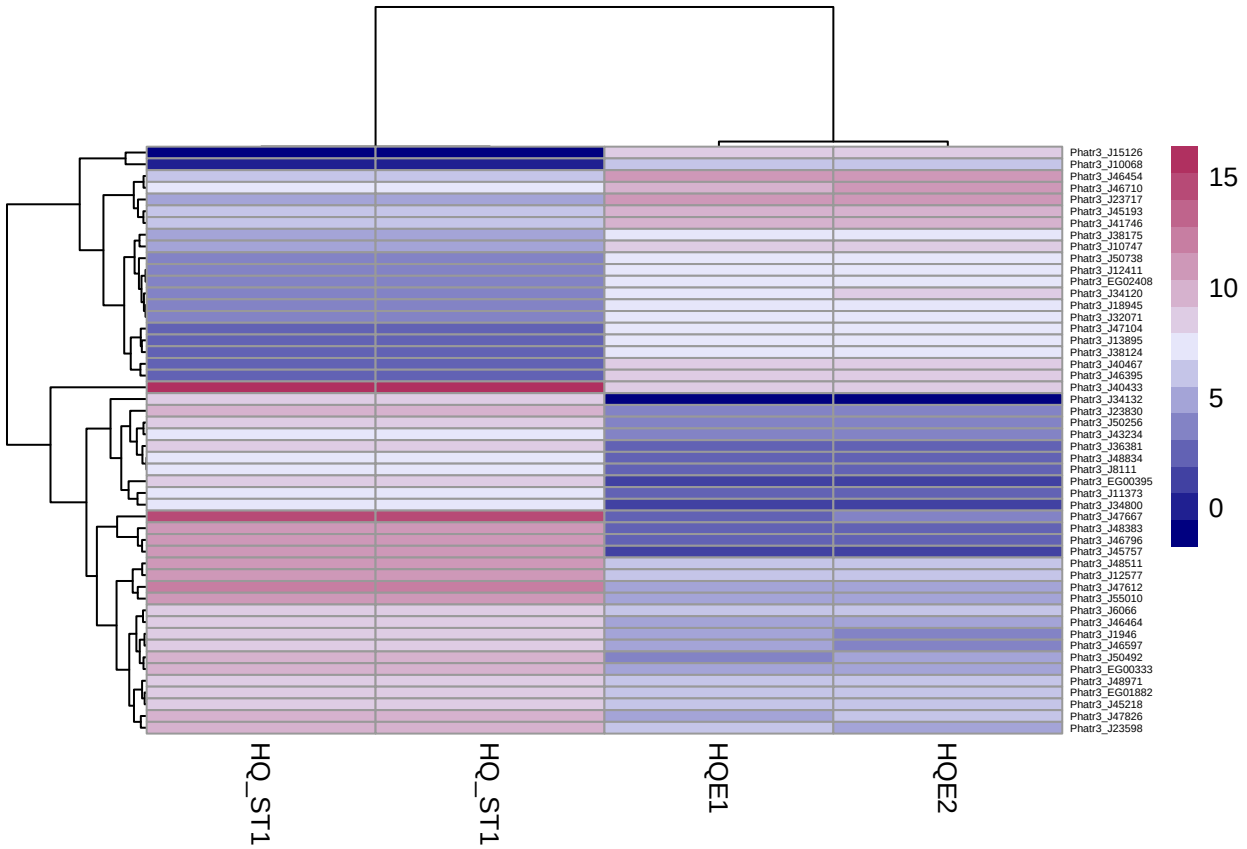
Volcano Plot

```
volcanoData <- cbind(HQST_vs_HQE_all$table$logFC, -log10(HQST_vs_HQE_all$table$FDR))
colnames(volcanoData) <- c("logFC", "negLogPval")
DEGs <- HQST_vs_HQE_all$table$FDR < 0.05 & abs(HQST_vs_HQE_all$table$logFC) > 1
point.col <- ifelse(DEGs, "red", "black")
plot(volcanoData, pch=19, cex = .5, col = point.col, main = "Volcano Plot : HQ St vs HQ E")
```




```
HQST_vs_HQE_top50_log2_cpm <- logcpm[rownames(HQST_vs_HQE_top50$table),]

pheatmap(subset(HQST_vs_HQE_top50_log2_cpm,select=c(HQ_ST1, HQ_ST1, HQE1, HQE2)),
         color=colorRampPalette(c("navy", "lavender", "maroon"))(15),fontsize_row=4)
```



100% Produced Water Exponential Growth vs 100% Produced Water in Stationary Growth

```
PWE_vs_PWST <- glmQLFTest(fit, contrast = my_contrasts[,"PWE_vs_PWST"])
PWE_vs_PWST_top50 = topTags(PWE_vs_PWST,adjust.method = "BH", p.value = 0.05, n = 50)
PWE_vs_PWST_all = topTags(PWE_vs_PWST,adjust.method = "BH", p.value = 0.05, n = nrow(PWE_vs_PWST$table))
PWE_vs_PWST_DEG <- PWE_vs_PWST_all[abs(PWE_vs_PWST_all$table$logFC) > 1, ]
PWE_vs_PWST_up <- PWE_vs_PWST_all[PWE_vs_PWST_all$table$logFC > 1, ]
PWE_vs_PWST_down <- PWE_vs_PWST_all[PWE_vs_PWST_all$table$logFC < -1, ]
head(PWE_vs_PWST_top50, n=10)
```

```
## Coefficient:  -1*PW_ST 1*PWE
##                    logFC    logCPM         F       PValue          FDR
## Phatr3_J23830  -6.170073  8.103721 1468.9586 6.958730e-11 1.449607e-07
## Phatr3_J46597  -7.421010  9.152332 1204.6116 1.619890e-10 1.449607e-07
## Phatr3_J10068   7.728042  6.486422 1155.1607 1.936245e-10 1.449607e-07
## Phatr3_J47104   6.502905  6.792173 1145.7683 2.004689e-10 1.449607e-07
## Phatr3_J32747   7.098911  8.734087 1110.3420 2.291282e-10 1.449607e-07
## Phatr3_J48834  -5.807797  6.469906 1070.6144 2.675374e-10 1.449607e-07
## Phatr3_J47667  -9.408787 13.420073 1016.6111 3.333997e-10 1.503106e-07
## Phatr3_J40433  -6.822467 14.655511  977.0339 3.947017e-10 1.503106e-07
## Phatr3_EG00065 -6.173506  9.136149  953.6820 4.374371e-10 1.503106e-07
## Phatr3_J25308   7.803495  9.255371  941.3297 4.623518e-10 1.503106e-07
```

Determine how many genes are up and down regulated for each pairwise comparison

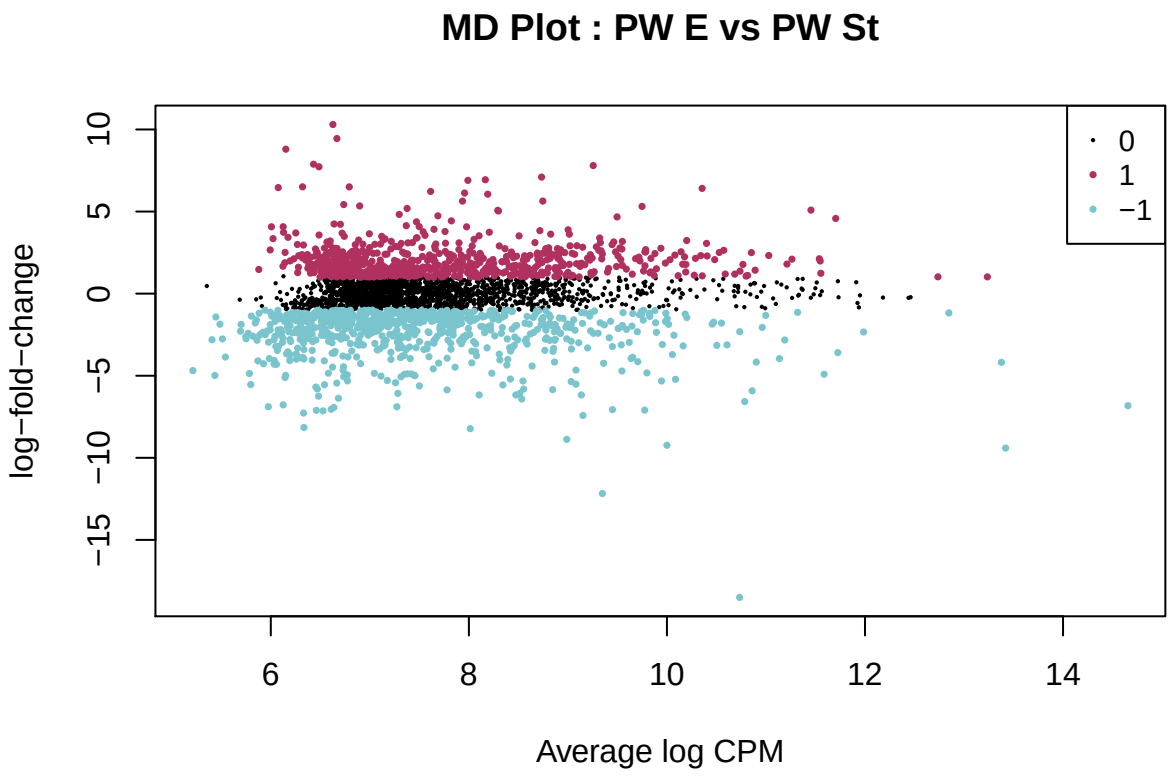
For differentially expressed genes set logFC and pvalue threshold to 0.05 and 1

lfc=1 sets a 2-fold change minimum

```
is.de <- decideTestsDGE(PWE_vs_PWST,adjust.method="BH",p.value=0.05,lfc=1)
summary(is.de)
```

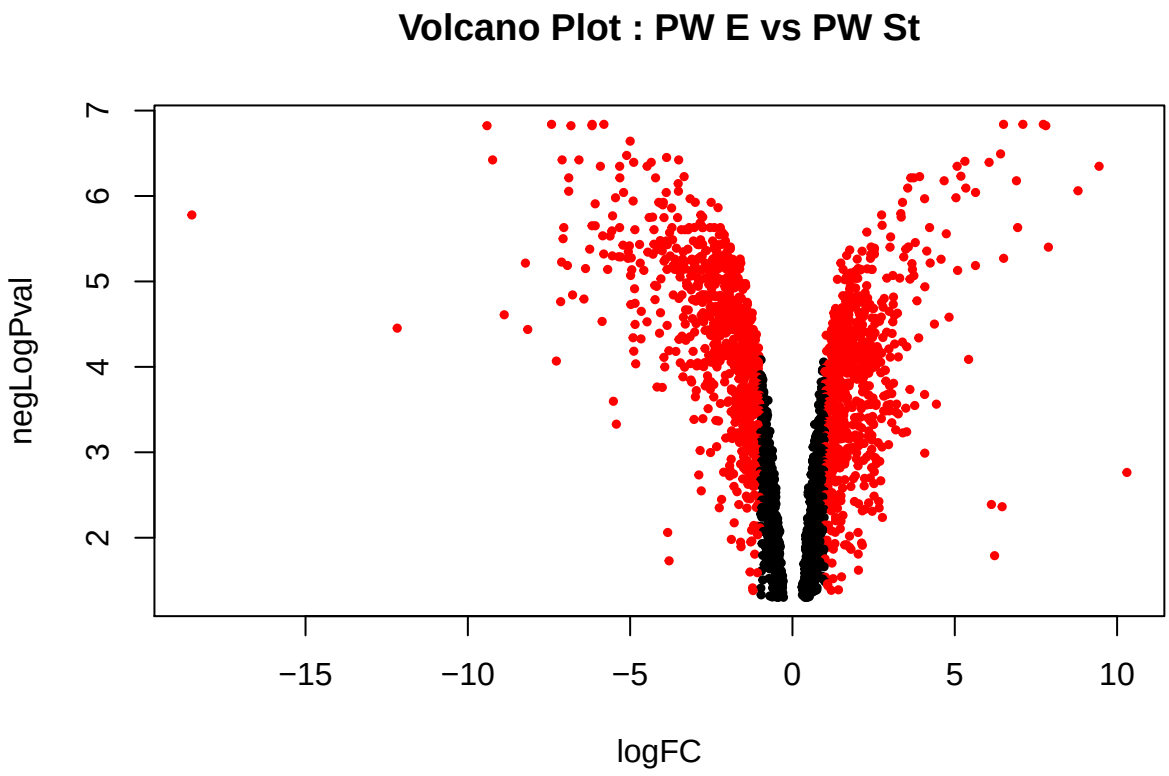
```
##        -1*PW_ST 1*PWE
## Down        785
## NotSig     1776
## Up          690
```

```
plotMD(PWE_vs_PWST, status=is.de, values=c(1,-1), col=c("maroon","cadetblue3"),
       legend="topright", cex = .5, main = "MD Plot : PW E vs PW St")
```



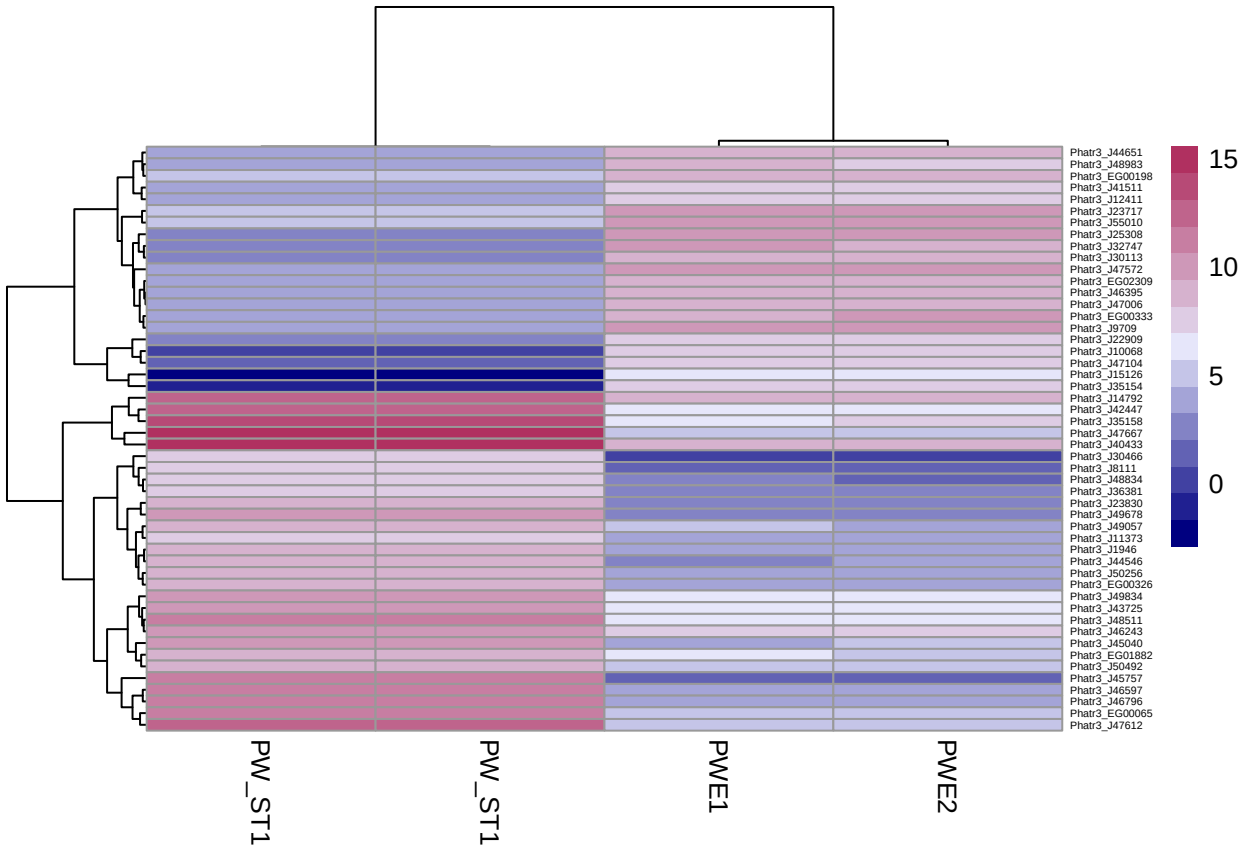
Volcano Plot

```
volcanoData <- cbind(PWE_vs_PWST_all$table$logFC, -log10(PWE_vs_PWST_all$table$FDR))
colnames(volcanoData) <- c("logFC", "negLogPval")
DEGs <- PWE_vs_PWST_all$table$FDR < 0.05 & abs(PWE_vs_PWST_all$table$logFC) > 1
point.col <- ifelse(DEGs, "red", "black")
plot(volcanoData, pch=19, cex = .5, col = point.col, main = "Volcano Plot : PW E vs PW St")
```




```
PWE_vs_PWST_top50_log2_cpm <- logcpm[rownames(PWE_vs_PWST_top50$table),]

pheatmap(subset(PWE_vs_PWST_top50_log2_cpm,select=c(PWE1, PWE2, PW_ST1, PW_ST1)),
         color=colorRampPalette(c("navy", "lavender", "maroon"))(15),fontsize_row=4)
```



Exponential and Stationary Samples as the Only Contrasts

Groups the normal media and PW and contrasts samples by their growth stage only

```
group<-c("exp","exp","exp","exp","st","st","exp","exp","st","st")

dge2<-DGEList(counts=star_data,group=group) #creates a DGE list object
dim(dge2)
```

```
## [1] 12392    10
```

```
full_dge<-dge2 #store original data just in case

# filtering & normalizing data
#keep only 100 counts per mil in at least 2 samples
head(cpm(dge2))
```

```
##                   HQ10E1     HQ10E2       HQE1       HQE2     HQ_ST1     HQ_ST2
## Phatr3_J31400  0.05087787  0.2236625  0.1336435  0.4334001   0.00000  0.2161654
## Phatr3_J42422  7.32641286  7.6939896  9.5332341  9.4728883  26.56036 24.3726541
## Phatr3_J31402  0.00000000  0.0000000  0.0000000  0.0000000   0.00000  0.0000000
## Phatr3_J42423  1.37370241  1.9682299  3.4301824  2.6004007  36.80130 27.4530117
## Phatr3_J42424 92.54684020 97.1589854 83.5717162 77.6405351 100.69153 99.4901468
## Phatr3_J7430  23.14942952 22.8135739 27.6196503 27.4280359  30.85495 27.7232185
```

```
##                     PWE1        PWE2     PW_ST1    PW_ST2
## Phatr3_J31400  0.1070611  0.05289409  0.5624836  0.910445
## Phatr3_J42422  6.2630738  5.34230276 11.4184165 11.517130
## Phatr3_J31402  0.0000000  0.00000000  0.0000000  0.000000
## Phatr3_J42423  2.4624051  2.06286938  6.1873193 13.838764
## Phatr3_J42424 81.2593682 79.76428279 58.2170497 71.105757
## Phatr3_J7430  17.8256717 24.11970355 18.9556964 16.570100
```

```
apply(dge2$counts, 2, sum)
```

```
##   HQ10E1   HQ10E2     HQE1     HQE2   HQ_ST1   HQ_ST2     PWE1     PWE2
## 19654912 22355112 22447786 16151357 15135335 18504345 18680923 18905705
##   PW_ST1   PW_ST2
## 17778297 21967279
```

```
keep2 <- rowSums(cpm(dge2)>100) >=2
dge2 <- dge2[keep2,]
dim(dge2)  #check number of genes left after filtering
```

```
## [1] 3251   10
```

```
#resetting the library size
dge2$samples$lib.size <- colSums(dge2$counts)
dge2$samples
```

```
##        group lib.size norm.factors
## HQ10E1   exp 15289361            1
## HQ10E2   exp 17319367            1
## HQE1     exp 17111487            1
## HQE2     exp 12568283            1
## HQ_ST1    st 11248709            1
## HQ_ST2    st 13869364            1
## PWE1     exp 14606604            1
## PWE2     exp 14925823            1
## PW_ST1    st 14619994            1
## PW_ST2    st 17756624            1
```

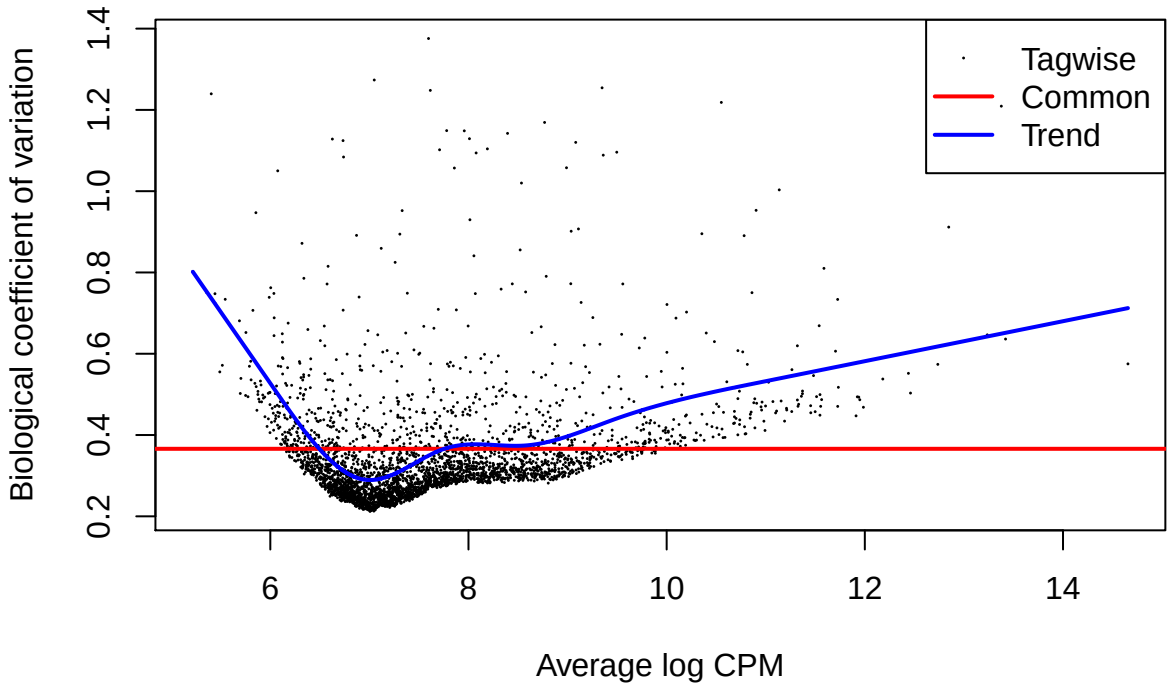
```
#now we can normalize the data
dge_norm2=calcNormFactors(dge2, method="TMM")
dge_norm2
```

```
## An object of class "DGEList"
## $counts
##                HQ10E1 HQ10E2 HQE1 HQE2 HQ_ST1 HQ_ST2 PWE1 PWE2 PW_ST1 PW_ST2
## Phatr3_J42426    8520   9594 8689 7289   5396   7052 6769 7244   5813   8494
## Phatr3_EG02408   3806   4483 3948 2637    224    238 3580 3084    216    368
## Phatr3_J31409    6448   7590 9257 5980   1008   1394 4863 4622   1381   1864
## Phatr3_J42429    1747   1981 2134 1377   2150   2386 1650 1699   1427   1398
## Phatr3_J4937     6608   7884 7049 4583   3737   5446 6837 6271   2997   3700
## 3246 more rows ...
##
```

```
## $samples
##        group lib.size norm.factors
## HQ10E1   exp 15289361    1.0347646
## HQ10E2   exp 17319367    1.0442405
## HQE1     exp 17111487    1.0884649
## HQE2     exp 12568283    1.0460172
## HQ_ST1    st 11248709    1.0977999
## HQ_ST2    st 13869364    1.0696829
## PWE1     exp 14606604    1.0251548
## PWE2     exp 14925823    1.0196206
## PW_ST1    st 14619994    0.8029116
## PW_ST2    st 17756624    0.8247657
```

```
# create design matrix
design.mat2 <- model.matrix(~ 0 + dge_norm2$samples$group)
colnames(design.mat2) <- levels(dge_norm2$samples$group)

#estimate the dispersion
d3 <- estimateGLMCommonDisp(dge_norm2,design.mat2)
d3 <- estimateGLMTrendedDisp(d3,design.mat2, method="auto")
# You can change method to "auto", "bin.spline", "power", "spline", "bin.loess".
# The default is "auto" which chooses "bin.spline" when > 200 tags and "power" otherwise.
d3 <- estimateGLMTagwiseDisp(d3,design.mat2)
plotBCV(d3)
```



```
#create a matrix of contrasts for anova-like testing
my_contrasts2<-makeContrasts(
 #exponential vs stationary
  exp_vs_st = exp-st,
  levels= design.mat2
)
my_contrasts2
```

```
##       Contrasts
## Levels exp_vs_st
##    exp         1
##    st         -1
```

```
fit2 <- glmQLFit(d3, design.mat2)
```

```
E_vs_ST <- glmQLFTest(fit2, contrast = my_contrasts2)
E_vs_ST_top100 = topTags(E_vs_ST,adjust.method = "BH", p.value = 0.05, n = 100)
E_vs_ST_all = topTags(E_vs_ST,adjust.method = "BH", p.value = 0.05, n = nrow(E_vs_ST$table))
E_vs_ST_DEG <- E_vs_ST_all[abs(E_vs_ST_all$table$logFC) > 1, ]
E_vs_ST_up <- E_vs_ST_all[E_vs_ST_all$table$logFC > 1, ]
E_vs_ST_down <- E_vs_ST_all[E_vs_ST_all$table$logFC < -1, ]
head(E_vs_ST_top100, n=10)
```

```
## Coefficient:  1*exp -1*st
##                     logFC    logCPM         F       PValue          FDR
## Phatr3_J45757  -9.563363 10.000099 1354.2665 1.684748e-12 5.477116e-09
## Phatr3_J23717   5.472037  9.748688  885.9707 1.554619e-11 2.527033e-08
## Phatr3_J47612  -6.769131 10.785967  716.4619 4.710233e-11 3.908231e-08
## Phatr3_J36381  -5.753589  6.668757  713.6253 4.808651e-11 3.908231e-08
## Phatr3_J47006   4.918942  8.298303  673.7905 6.486447e-11 4.217488e-08
## Phatr3_J48511  -4.332105  9.654110  560.0490 1.696963e-10 9.194711e-08
## Phatr3_J12411   3.673927  6.993582  536.4989 2.120974e-10 9.850411e-08
## Phatr3_J44651   3.881203  7.976553  520.8250 2.473700e-10 1.005250e-07
## Phatr3_J49907  -3.531604  7.214636  484.8683 3.584177e-10 1.179003e-07
## Phatr3_EG02408  3.525080  7.151882  480.6009 3.752105e-10 1.179003e-07
```

Determine how many genes are up and down regulated for each pairwise comparison

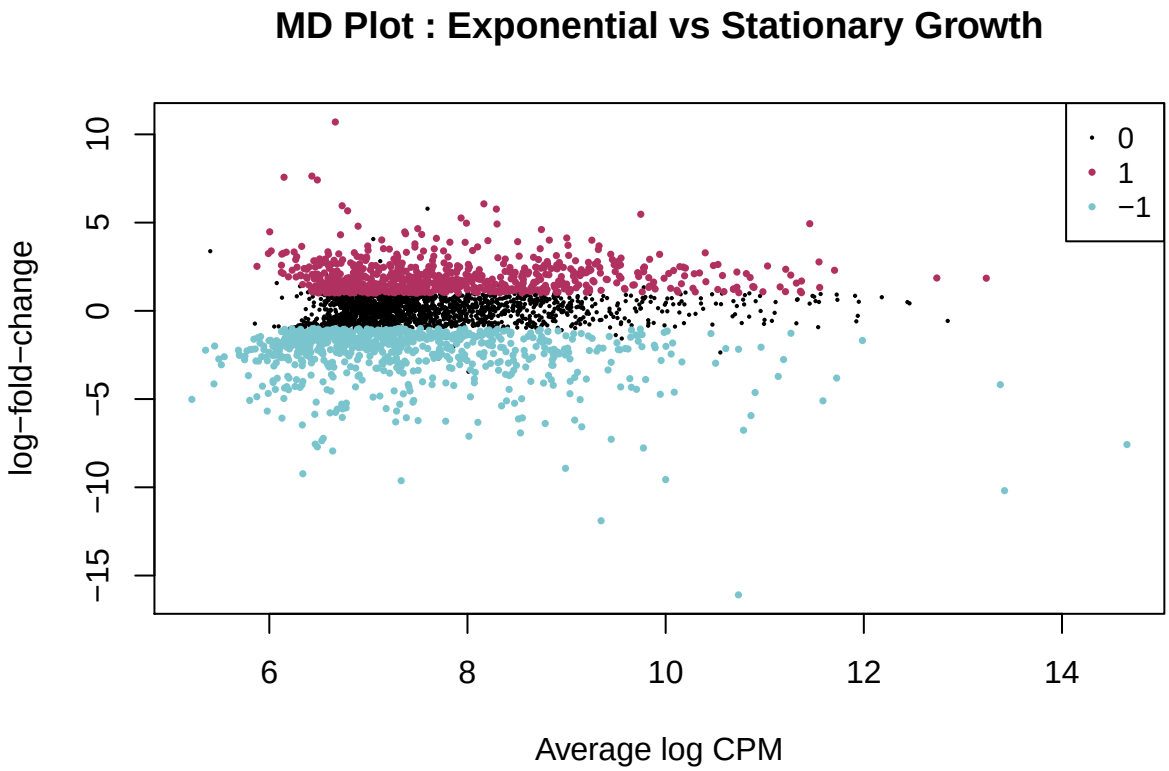
For differentially expressed genes set logFC and pvalue threshold to 0.05 and 1

lfc=1 sets a 2-fold change minimum

```
is.de <- decideTestsDGE(E_vs_ST,adjust.method="BH",p.value=0.05,lfc=1)
summary(is.de)
```

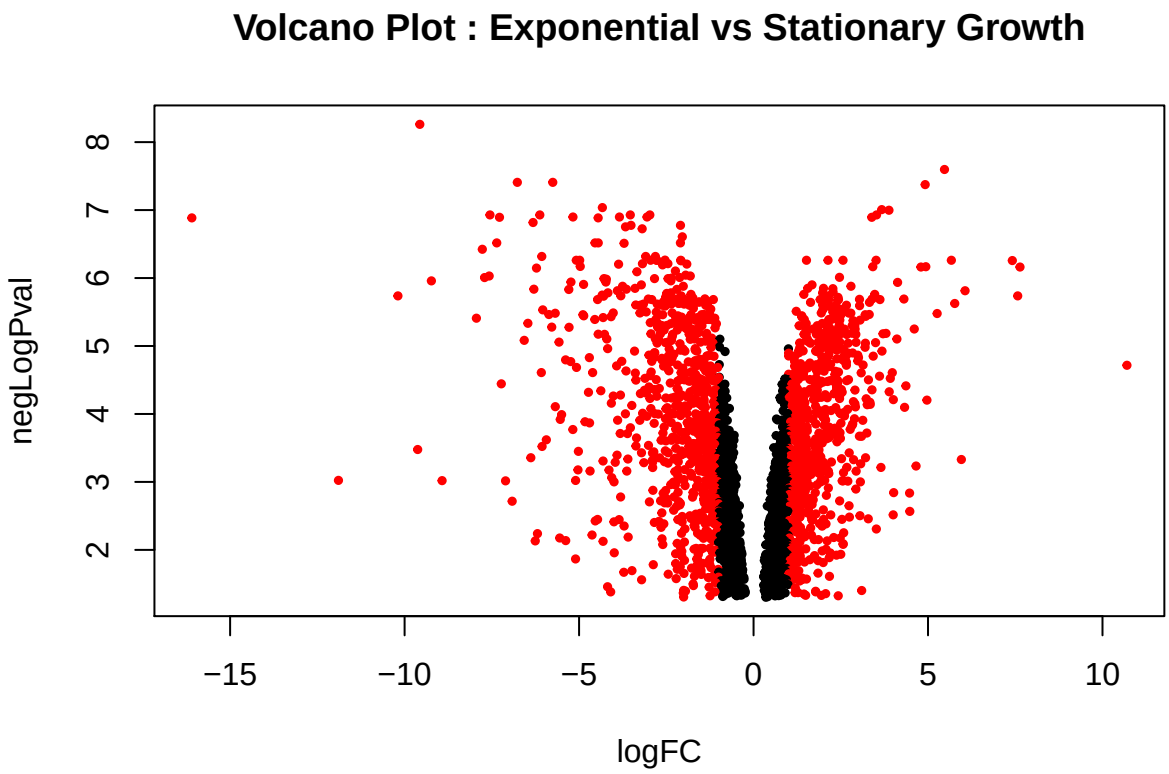
```
##        1*exp -1*st
## Down           771
## NotSig        1792
## Up             688
```

```
plotMD(E_vs_ST, status=is.de, values=c(1,-1), col=c("maroon","cadetblue3"),
       legend="topright", cex = .5, main = "MD Plot : Exponential vs Stationary Growth")
```

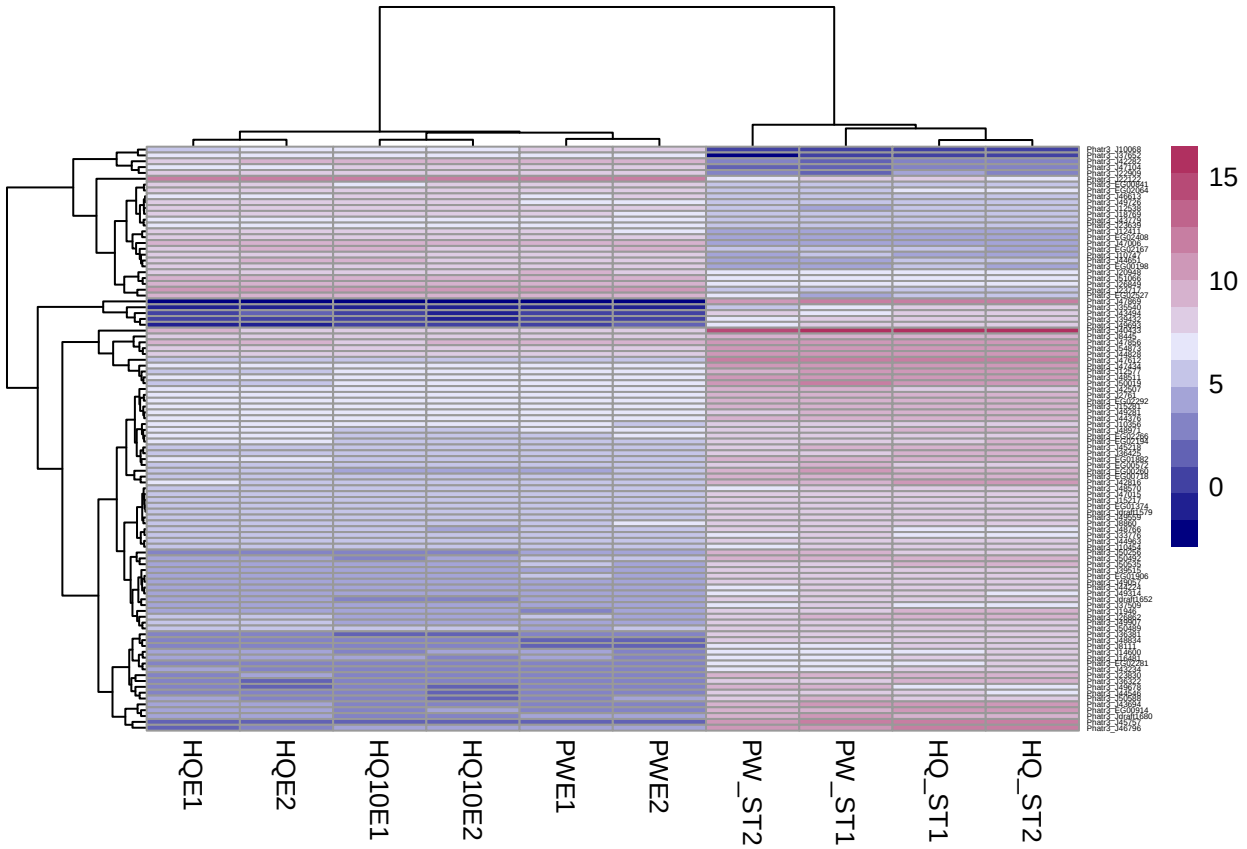


Volcano Plot

```
volcanoData <- cbind(E_vs_ST_all$table$logFC, -log10(E_vs_ST_all$table$FDR))
colnames(volcanoData) <- c("logFC", "negLogPval")
DEGs <- E_vs_ST_all$table$FDR < 0.05 & abs(E_vs_ST_all$table$logFC) > 1
point.col <- ifelse(DEGs, "red", "black")
plot(volcanoData, pch=19, cex = .5, col = point.col, main = "Volcano Plot : Exponential vs Stationary G
```




```
E_vs_ST_top100_log2_cpm <- logcpm[rownames(E_vs_ST_top100$table),]
pheatmap(E_vs_ST_top100_log2_cpm,color=colorRampPalette(c("navy", "lavender", "maroon"))(15),fontsize_r
```



Find intersect of DEGs found between HQst_vs_PWst and HQE_vs_PWE

```
common_HQ_vs_PW_DEGs <- intersect(row.names(HQst_vs_PWst_DEG$table), row.names(HQE_vs_PWE_DEG$table))
common_HQ_vs_PW_DEGs <- data.frame(common_HQ_vs_PW_DEGs)
```

Write all DEGs results to a Excel file

```
library(openxlsx)
list_of_datasets <- list("HQst_vs_PWst" = HQst_vs_PWst_DEG$table,
                         "HQE_vs_PWE" = HQE_vs_PWE_DEG$table,
                         "HQST_vs_HQE" = HQST_vs_HQE_DEG$table,
                         "PWE_vs_PWST" = PWE_vs_PWST_DEG$table,
                         "E_vs_ST" = E_vs_ST_DEG$table,
                         "HQ_vs_PW" = common_HQ_vs_PW_DEGs
                         )
write.xlsx(list_of_datasets, file = "PW_DEGs.xlsx", row.names = TRUE)
```

Gene Ontology

Perform gene ontology of the common DEGs between HQst_vs_PWst and HQE_vs_PWE with gprofiler2

```
library(gprofiler2)

HQst_vs_PWst_over_rep <- gost(query = list("HQst vs PWst" = row.names(HQst_vs_PWst_DEG$table)),
                          organism = "ptricornutum",
                          ordered_query = TRUE,
                          measure_underrepresentation = FALSE,
                          evcodes = TRUE)

HQE_vs_PWE_over_rep <- gost(query = list("HQE vs PWE" = row.names(HQE_vs_PWE_DEG$table)),
                          organism = "ptricornutum",
                          ordered_query = TRUE,
                          measure_underrepresentation = FALSE,
                          evcodes = TRUE)

HQst_vs_PWst_under_rep <- gost(query = list("HQst vs PWst" = row.names(HQst_vs_PWst_DEG$table)),
                          organism = "ptricornutum",
                          ordered_query = TRUE,
                          measure_underrepresentation = TRUE,
                          evcodes = TRUE)

HQE_vs_PWE_under_rep <- gost(query = list("HQE vs PWE" = row.names(HQE_vs_PWE_DEG$table)),
                          organism = "ptricornutum",
                          ordered_query = TRUE,
                          measure_underrepresentation = TRUE,
                          evcodes = TRUE)
```

HQst vs PWst

Visualize the over-represented GO terms of HQst vs PWst

```
gostplot(HQst_vs_PWst_over_rep, capped = TRUE, interactive = TRUE)
```

Top 20 GO terms of over-represented DEGs in HQst vs PWst by lowest p-values

```
publish_gosttable(HQst_vs_PWst_over_rep,
                  highlight_terms = HQst_vs_PWst_over_rep$result[order(HQst_vs_PWst_over_rep$result$p_va
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:BP	GO:0019684	photosynthesis, light reaction	51	32		9.0e−13
2	GO:MF	GO:0016168	chlorophyll binding	41	25		1.9e−12
3	GO:BP	GO:0015979	photosynthesis	65	36		6.0e−12
4	GO:MF	GO:0046906	tetrapyrrole binding	100	43		7.0e−12
5	GO:BP	GO:0009765	photosynthesis, light harvesting	42	21		2.2e−10
6	GO:BP	GO:0006091	generation of precursor metabolites and energy	175	47		1.7e−09
7	GO:BP	GO:0018298	protein−chromophore linkage	46	21		2.1e−09
8	GO:CC	GO:0030076	light−harvesting complex	41	21		6.7e−09
9	GO:CC	GO:0009579	thylakoid	54	30		2.6e−07
10	GO:CC	GO:0034357	photosynthetic membrane	50	28		7.1e−07
11	GO:CC	GO:0009521	photosystem	46	26		1.9e−06
12	GO:CC	GO:0009523	photosystem II	46	26		1.9e−06
13	GO:BP	GO:0006096	glycolytic process	45	11		1.2e−04
14	GO:BP	GO:0006757	ATP generation from ADP	45	11		1.2e−04
15	GO:BP	GO:0046031	ADP metabolic process	46	11		1.6e−04
16	GO:BP	GO:0009185	ribonucleoside diphosphate metabolic process	48	11		2.5e−04
17	GO:BP	GO:0006165	nucleoside diphosphate phosphorylation	48	11		2.5e−04
18	GO:BP	GO:0009135	purine nucleoside diphosphate metabolic process	48	11		2.5e−04
19	GO:BP	GO:0009179	purine ribonucleoside diphosphate metabolic process	48	11		2.5e−04
20	GO:BP	GO:0016052	carbohydrate catabolic process	51	11		4.8e−04

g:Profiler (biit.cs.ut.ee/gprofiler)

Visualize the under-represented GO terms of HQst vs PWst

```
gostplot(HQst_vs_PWst_under_rep, capped = TRUE, interactive = TRUE)
```

Top 20 GO terms of under-represented DEGs in HQst vs PWst by lowest p-values

```
publish_gosttable(HQst_vs_PWst_under_rep,
                  highlight_terms = HQst_vs_PWst_under_rep$result[order(HQst_vs_PWst_under_rep$result$p
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:BP	GO:0008150	biological_process	4450	277		4.9e−324
2	GO:CC	GO:0005575	cellular_component	3635	1		4.9e−324
3	GO:CC	GO:0110165	cellular anatomical entity	3570	36		4.9e−324
4	GO:MF	GO:0003674	molecular_function	5580	20		4.9e−324
5	KEGG	KEGG:01100	Metabolic pathways	413	41		4.9e−324
6	KEGG	KEGG:00000	KEGG root term	984	1		4.9e−324
7	GO:BP	GO:0009987	cellular process	3397	248		1.7e−115
8	GO:BP	GO:0008152	metabolic process	3488	261		1.4e−114
9	GO:CC	GO:0016020	membrane	2315	212		9.2e−102
10	GO:CC	GO:0005622	intracellular anatomical structure	1551	88		1.2e−86
11	GO:CC	GO:0031224	intrinsic component of membrane	2094	191		2.4e−82
12	GO:CC	GO:0016021	integral component of membrane	2094	191		2.4e−82
13	GO:BP	GO:0043170	macromolecule metabolic process	1930	82		2.1e−75
14	KEGG	KEGG:01110	Biosynthesis of secondary metabolites	174	23		4.8e−74
15	GO:BP	GO:0071704	organic substance metabolic process	2717	187		1.7e−73
16	GO:BP	GO:0044238	primary metabolic process	2535	162		6.6e−73
17	GO:BP	GO:0006807	nitrogen compound metabolic process	2348	147		2.3e−65
18	GO:BP	GO:0044237	cellular metabolic process	2506	177		2.3e−62
19	GO:CC	GO:0043226	organelle	1198	65		4.2e−62
20	GO:CC	GO:0043229	intracellular organelle	1179	64		9.6e−61

g:Profiler (biit.cs.ut.ee/gprofiler)

HQE vs PWE

Visualize the over-represented GO terms of HQE vs PWE

```
gostplot(HQE_vs_PWE_over_rep, capped = TRUE, interactive = TRUE)
```

Top 20 GO terms of over-represented DEGs in HQE vs PWE by lowest p-values

```
publish_gosttable(HQE_vs_PWE_over_rep,
                  highlight_terms = HQE_vs_PWE_over_rep$result[order(HQE_vs_PWE_over_rep$result$p_value
                  )
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:CC	GO:0005852	eukaryotic translation initiation factor 3 complex	13	9		4.9e−07
2	GO:MF	GO:0008135	translation factor activity, RNA binding	55	13		3.3e−05
3	GO:MF	GO:0045182	translation regulator activity	55	13		3.3e−05
4	GO:MF	GO:0090079	translation regulator activity, nucleic acid binding	55	13		3.3e−05
5	GO:MF	GO:0003743	translation initiation factor activity	35	10		1.3e−04
6	GO:CC	GO:0033290	eukaryotic 48S preinitiation complex	4	4		7.1e−04
7	GO:CC	GO:0016282	eukaryotic 43S preinitiation complex	4	4		7.1e−04
8	GO:CC	GO:0070993	translation preinitiation complex	4	4		7.1e−04
9	GO:BP	GO:0043436	oxoacid metabolic process	327	44		1.1e−03
10	GO:BP	GO:0001732	formation of cytoplasmic translation initiation complex	4	4		1.1e−03
11	GO:BP	GO:0019752	carboxylic acid metabolic process	324	43		1.3e−03
12	GO:MF	GO:0031369	translation initiation factor binding	3	3		1.4e−03
13	GO:BP	GO:0006413	translational initiation	34	8		2.4e−03
14	GO:BP	GO:0006082	organic acid metabolic process	337	44		2.4e−03
15	GO:BP	GO:0001731	formation of translation preinitiation complex	5	4		5.5e−03
16	GO:BP	GO:0002183	cytoplasmic translational initiation	5	4		5.5e−03
17	GO:BP	GO:0006446	regulation of translational initiation	5	4		5.5e−03
18	GO:BP	GO:0006090	pyruvate metabolic process	53	10		2.2e−02
19	GO:MF	GO:0005452	inorganic anion exchanger activity	8	4		3.4e−02

g:Profiler (biit.cs.ut.ee/gprofiler)

Visualize the under-represented GO terms of HQE vs PWE

```
gostplot(HQE_vs_PWE_under_rep, capped = TRUE, interactive = TRUE)
```

Top 20 GO terms of under-represented DEGs in HQE vs PWE by lowest p-values

```
publish_gosttable(HQE_vs_PWE_under_rep,
                  highlight_terms = HQE_vs_PWE_under_rep$result[order(HQE_vs_PWE_under_rep$result$p_valu
                  )
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:CC	GO:0110165	cellular anatomical entity	3570	41		4.9e−324
2	GO:CC	GO:0005575	cellular_component	3635	1		4.9e−324
3	GO:MF	GO:0003674	molecular_function	5580	33		4.9e−324
4	KEGG	KEGG:00000	KEGG root term	984	1		4.9e−324
5	GO:BP	GO:0008150	biological_process	4450	181		4.6e−88
6	GO:CC	GO:0016020	membrane	2315	58		4.8e−65
7	GO:CC	GO:0016021	integral component of membrane	2094	53		6.8e−53
8	GO:CC	GO:0031224	intrinsic component of membrane	2094	53		6.8e−53
9	KEGG	KEGG:01100	Metabolic pathways	413	35		1.4e−47
10	GO:BP	GO:0009987	cellular process	3397	133		7.9e−29
11	GO:BP	GO:0008152	metabolic process	3488	146		1.7e−24
12	GO:BP	GO:0043170	macromolecule metabolic process	1930	46		1.2e−20
13	GO:CC	GO:0043226	organelle	1198	28		1.0e−19
14	GO:CC	GO:0043229	intracellular organelle	1179	27		1.4e−17
15	GO:CC	GO:0005622	intracellular anatomical structure	1551	55		7.2e−17
16	GO:BP	GO:0071704	organic substance metabolic process	2717	108		1.7e−15
17	GO:BP	GO:0044260	cellular macromolecule metabolic process	1452	34		6.1e−14
18	GO:MF	GO:0005488	binding	3541	136		8.4e−14
19	GO:CC	GO:0043227	membrane−bounded organelle	1028	27		1.5e−12
20	GO:BP	GO:0006807	nitrogen compound metabolic process	2348	91		1.6e−12

g:Profiler (biit.cs.ut.ee/gprofiler)

Common DEGs between HQst vs PWst and HQE and PWE

```
common_HQ_vs_PW_over_rep <- gost(query = list("Common Over-rep DEGs: HQ vs PW" = common_HQ_vs_PW_DEGs[,
                          organism = "ptricornutum",
                          ordered_query = TRUE,
                          measure_underrepresentation = FALSE,
                          evcodes = TRUE)

common_HQ_vs_PW_under_rep <- gost(query = list("Common Under-rep DEGs: HQ vs PW" = common_HQ_vs_PW_DEGs
                          organism = "ptricornutum",
                          ordered_query = TRUE,
                          measure_underrepresentation = TRUE,
                          evcodes = TRUE)
```

```
gostplot(common_HQ_vs_PW_over_rep, capped = TRUE, interactive = TRUE)
```

```
publish_gosttable(common_HQ_vs_PW_over_rep,
                  highlight_terms = common_HQ_vs_PW_over_rep$result[order(common_HQ_vs_PW_over_rep$resul
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	KEGG	KEGG:00910	Nitrogen metabolism	11	7		4.9e−03
2	GO:BP	GO:0009084	glutamine family amino acid biosynthetic process	18	5		1.7e−02

g:Profiler (biit.cs.ut.ee/gprofiler)

```
gostplot(common_HQ_vs_PW_under_rep, capped = TRUE, interactive = TRUE)
```

```
publish_gosttable(common_HQ_vs_PW_under_rep,
                  highlight_terms = common_HQ_vs_PW_under_rep$result[order(common_HQ_vs_PW_under_rep$res
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:CC	GO:0110165	cellular anatomical entity	3570	38		4.9e−324
2	GO:CC	GO:0005575	cellular_component	3635	1		4.9e−324
3	GO:MF	GO:0003674	molecular_function	5580	25		4.9e−324
4	KEGG	KEGG:00000	KEGG root term	984	1		4.9e−324
5	GO:BP	GO:0008150	biological_process	4450	63		5.6e−56
6	GO:BP	GO:0009987	cellular process	3397	40		8.9e−25
7	GO:BP	GO:0008152	metabolic process	3488	43		2.4e−22
8	KEGG	KEGG:01100	Metabolic pathways	413	13		2.7e−19
9	GO:CC	GO:0005622	intracellular anatomical structure	1551	11		4.4e−18
10	GO:CC	GO:0016020	membrane	2315	36		1.0e−17
11	GO:BP	GO:0043170	macromolecule metabolic process	1930	10		3.0e−17
12	GO:CC	GO:0043226	organelle	1198	4		4.2e−17
13	GO:CC	GO:0043229	intracellular organelle	1179	4		1.2e−16
14	GO:BP	GO:0071704	organic substance metabolic process	2717	30		6.5e−16
15	GO:MF	GO:0005488	binding	3541	40		7.4e−15
16	GO:BP	GO:0044237	cellular metabolic process	2506	27		4.1e−14
17	GO:BP	GO:0044238	primary metabolic process	2535	28		6.0e−14
18	GO:CC	GO:0016021	integral component of membrane	2094	34		9.8e−14
19	GO:CC	GO:0031224	intrinsic component of membrane	2094	34		9.8e−14
20	GO:CC	GO:0043227	membrane−bounded organelle	1028	4		2.3e−13

g:Profiler (biit.cs.ut.ee/gprofiler)

Write gene ontology results to an Excel file

```
columns <- c("term_id",
             "source",
             "term_name",
             "p_value",
             "effective_domain_size",
```

```
             "intersection_size",
             "intersection")

GO_results <- list("HQst_vs_PWst_over_rep" = HQst_vs_PWst_over_rep$result[, columns],
                   "HQst_vs_PWst_under_rep" = HQst_vs_PWst_under_rep$result[, columns],
                   "HQE_vs_PWE_over_rep" = HQE_vs_PWE_over_rep$result[, columns],
                   "HQE_vs_PWE_under_rep" = HQE_vs_PWE_under_rep$result[, columns],
                   "common_HQ_vs_PW_over_rep" = common_HQ_vs_PW_over_rep$result[, columns],
                   "common_HQ_vs_PW_under_rep" = common_HQ_vs_PW_under_rep$result[, columns]
                   )

write.xlsx(GO_results, file = "PW_GO_results.xlsx", row.names = TRUE)
```

Explore Gene Ontology of Up-regulated vs Down-regulated DEGs in HQst vs PWst

```
HQst_vs_PWst_up_reg_over_rep <- gost(query = list("HQst vs PWst Up-regulated" = row.names(HQst_vs_PWst_u
                                     organism = "ptricornutum",
                                     ordered_query = TRUE,
                                     measure_underrepresentation = FALSE,
                                     evcodes = TRUE)
```

Over-represented GO terms of Up-regulated DEGs in HQst vs PWst

```
gostplot(HQst_vs_PWst_up_reg_over_rep, capped = TRUE, interactive = TRUE)
```

```
publish_gosttable(HQst_vs_PWst_up_reg_over_rep,
                  highlight_terms = HQst_vs_PWst_up_reg_over_rep$result[order(HQst_vs_PWst_up_reg_over_r
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:BP	GO:0006757	ATP generation from ADP	45	5		2.8e−03
2	GO:BP	GO:0006096	glycolytic process	45	5		2.8e−03
3	GO:BP	GO:0046031	ADP metabolic process	46	5		3.1e−03
4	GO:BP	GO:0009185	ribonucleoside diphosphate metabolic process	48	5		3.9e−03
5	GO:BP	GO:0009179	purine ribonucleoside diphosphate metabolic process	48	5		3.9e−03
6	GO:BP	GO:0006165	nucleoside diphosphate phosphorylation	48	5		3.9e−03
7	GO:BP	GO:0009135	purine nucleoside diphosphate metabolic process	48	5		3.9e−03
8	GO:BP	GO:0016052	carbohydrate catabolic process	51	5		5.2e−03
9	GO:BP	GO:0006090	pyruvate metabolic process	53	5		6.3e−03
10	GO:BP	GO:0009132	nucleoside diphosphate metabolic process	54	5		6.9e−03
11	GO:BP	GO:0046939	nucleotide phosphorylation	56	5		8.3e−03
12	GO:BP	GO:0046034	ATP metabolic process	60	5		1.2e−02
13	GO:MF	GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	21	2		1.8e−02
14	GO:BP	GO:0006006	glucose metabolic process	18	2		3.2e−02
15	GO:MF	GO:0015204	urea transmembrane transporter activity	3	2		4.3e−02
16	GO:MF	GO:0050661	NADP binding	33	2		4.6e−02
17	GO:MF	GO:0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	34	2		4.9e−02

g:Profiler (biit.cs.ut.ee/gprofiler)

Over-represented GO terms of Down-regulated DEGs in HQst vs PWst

```
HQst_vs_PWst_down_reg_over_rep <- gost(query = list("HQst vs PWst Down-regulated" = row.names(HQst_vs_P
                            organism = "ptricornutum",
                            ordered_query = TRUE,
                            measure_underrepresentation = FALSE,
                            evcodes = TRUE)
```

```
gostplot(HQst_vs_PWst_down_reg_over_rep, capped = TRUE, interactive = TRUE)
```

```
publish_gosttable(HQst_vs_PWst_down_reg_over_rep,
                  highlight_terms = HQst_vs_PWst_down_reg_over_rep$result[order(HQst_vs_PWst_down_reg_o
```

id	source	term_id	term_name	term_size	intersection_size		p_value
1	GO:BP	GO:0019684	photosynthesis, light reaction	51	27		2.5e−15
2	GO:BP	GO:0015979	photosynthesis	65	34		6.9e−15
3	GO:MF	GO:0016168	chlorophyll binding	41	22		8.3e−15
4	GO:MF	GO:0046906	tetrapyrrole binding	100	38		2.2e−13
5	GO:BP	GO:0009765	photosynthesis, light harvesting	42	22		8.3e−13
6	GO:BP	GO:0018298	protein−chromophore linkage	46	22		9.8e−12
7	GO:CC	GO:0030076	light−harvesting complex	41	22		2.9e−11
8	GO:CC	GO:0009579	thylakoid	54	28		1.5e−09
9	GO:BP	GO:0006091	generation of precursor metabolites and energy	175	40		3.2e−09
10	GO:CC	GO:0009521	photosystem	46	25		4.7e−09
11	GO:CC	GO:0009523	photosystem II	46	25		4.7e−09
12	GO:CC	GO:0034357	photosynthetic membrane	50	26		7.2e−09
13	GO:CC	GO:0009536	plastid	108	28		2.1e−05
14	GO:BP	GO:0006779	porphyrin−containing compound biosynthetic process	25	12		3.3e−04
15	GO:BP	GO:0033014	tetrapyrrole biosynthetic process	31	13		7.9e−04
16	GO:BP	GO:0006782	protoporphyrinogen IX biosynthetic process	12	8		8.4e−04
17	GO:BP	GO:0046501	protoporphyrinogen IX metabolic process	12	8		8.4e−04
18	GO:BP	GO:0046148	pigment biosynthetic process	32	13		1.2e−03
19	GO:BP	GO:0006783	heme biosynthetic process	17	6		1.6e−03
20	GO:CC	GO:0009507	chloroplast	100	23		2.2e−03

g:Profiler (biit.cs.ut.ee/gprofiler)

```
sessionInfo()
```

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19042)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
```

```
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] gprofiler2_0.2.0 openxlsx_4.2.3   pheatmap_1.0.12  edgeR_3.32.1
## [5] limma_3.46.0     readr_1.4.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.6         locfit_1.5-9.4     lattice_0.20-41    tidyr_1.1.3
##  [5] assertthat_0.2.1   digest_0.6.27      utf8_1.2.1         mime_0.10
##  [9] R6_2.5.0           evaluate_0.14      httr_1.4.2         ggplot2_3.3.3
## [13] highr_0.9          pillar_1.6.0       rlang_0.4.10       lazyeval_0.2.2
## [17] curl_4.3           rstudioapi_0.13    data.table_1.14.0  rmarkdown_2.8
## [21] splines_4.0.4      stringr_1.4.0      htmlwidgets_1.5.3  RCurl_1.98-1.3
## [25] munsell_0.5.0      shiny_1.6.0        compiler_4.0.4     httpuv_1.6.1
## [29] xfun_0.22          pkgconfig_2.0.3    htmltools_0.5.1.1  tidyselect_1.1.1
## [33] gridExtra_2.3      tibble_3.1.0       fansi_0.4.2        viridisLite_0.4.0
## [37] crayon_1.4.1       dplyr_1.0.5        later_1.2.0        bitops_1.0-7
## [41] grid_4.0.4         xtable_1.8-4       jsonlite_1.7.2     gtable_0.3.0
## [45] lifecycle_1.0.0    DBI_1.1.1          magrittr_2.0.1     scales_1.1.1
## [49] zip_2.1.1          cli_2.5.0          stringi_1.5.3      promises_1.2.0.1
## [53] ellipsis_0.3.1     generics_0.1.0     vctrs_0.3.7        RColorBrewer_1.1-2
## [57] tools_4.0.4        glue_1.4.2         purrr_0.3.4        hms_1.0.0
## [61] crosstalk_1.1.1    fastmap_1.1.0      yaml_2.2.1         colorspace_2.0-0
## [65] plotly_4.9.3       knitr_1.33
```