Biochemical Analysis 20 Nov 2017

Chandra Jack

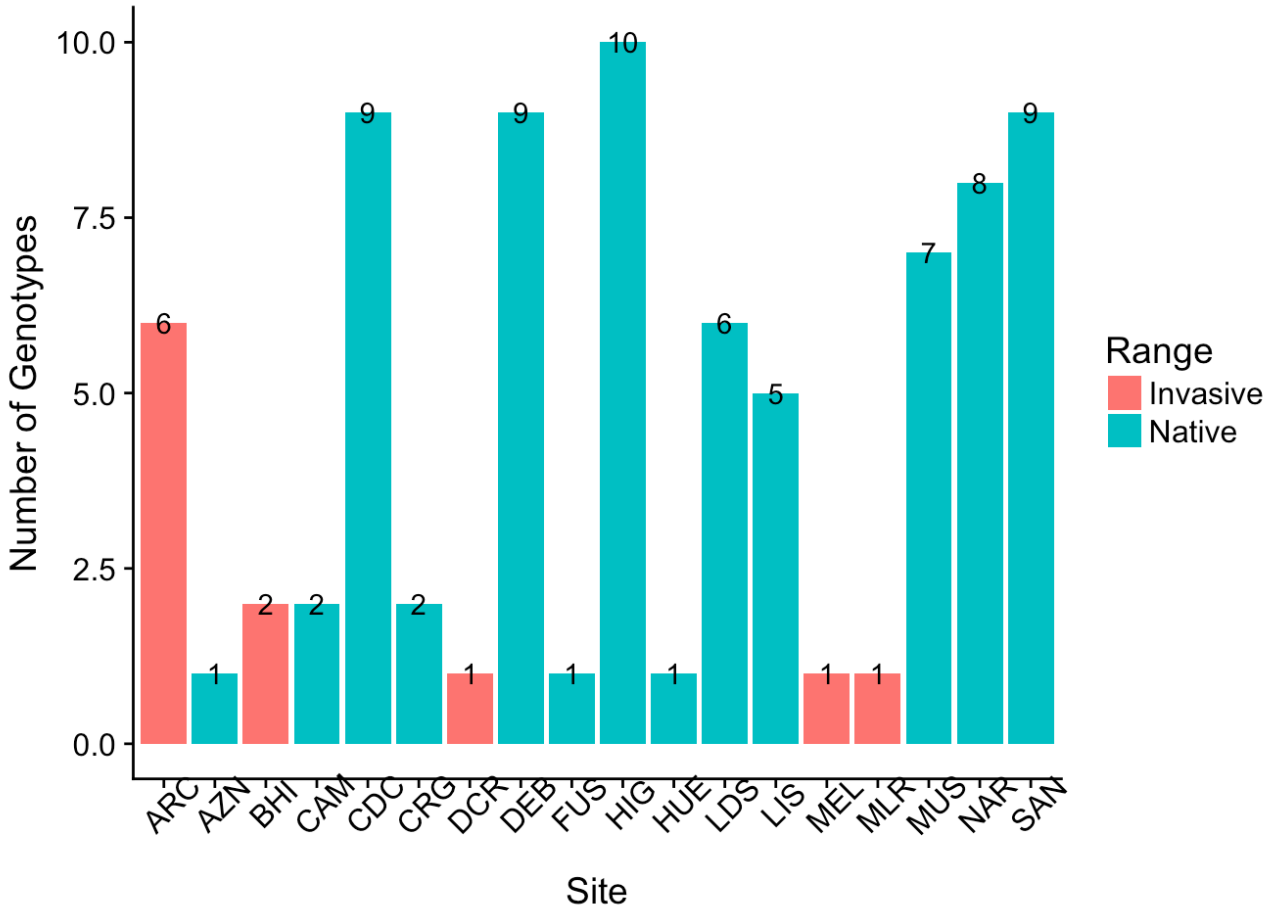
November 20, 2017

```
library(ggplot2)
library(data.table)
library(scales)
library(plyr)
library(dplyr)
library(reshape2)
library(lme4)
library(effects)
library(multcomp)
library(lmerTest)
library(piecewiseSEM)
library(car)
library(gridExtra)
library(cowplot)
library(Rmisc)
library(heplots)
library(Hmisc)
library(RcmdrMisc)
library(corrplot)
```

This is the set of packages that I regularly use. Not all of them are necessary to run the code below. This is a compilation of the key parts of the analysis that has been done.

Bootstrapping

There are 11 Invasive genotypes representing 5 populations but 70 genotypes 13 populations for the Native range. I ran a Type II MANOVA for each assay instead of a repeated measures ANOVA due to the unbalanced design and because it is less sensitive to heterogeneity of variances. However, I still wanted to check to see if differences in group size could change significance.



T test of bootstrap variance

I resampled from the native population, calcuated the variance and compared it to the variance from the full range set

Subset the native and invasive populations

```
BV_Nat <- subset(BiocAssaysWide, Range == "Native")
BV_Inv <- subset(BiocAssaysWide, Range == "Invasive")
```

```
N = 1000 # Number of times to resample
TestMat <- matrix(rep(0, 9*N), ncol = 9) # Define the matrix to hold bootstrapped
values and initiate with 0's
for (i in 1:N){
  pops <- sample(unique(BV_Nat$Genotype), 11) # Sample populations randomly keepi
ng the rows and replicates together
  BV_Nat2 <- BV_Nat[BV_Nat$Genotype %in% pops, ] # Narrow dataset
  TestMat[i,] <- apply(BV_Nat2[,6:ncol(BV_Nat2)], 2, function(x) var(x, na.rm=TRU
E)) # Calculate variance and fill matrix
}

VecVar <- apply(BV_Nat[,6:ncol(BV_Nat)], 2, function(x) var(x, na.rm=TRUE)) # Var
iance of all native genotypes
ColnNames <- colnames(BV_Nat)
ColnNames <- ColnNames[-(1:5)]

NewMat <- matrix(rep(0, 27), ncol = 3) # Place in its own matrix to call later
for (i in 1:9) {
  NewMat[i,1] <- ColnNames[i] # each variable tested
  NewMat[i,2] <- t.test(TestMat[,i], mu = VecVar[i])$p.value # p value comparing
vector of bootstrap variances to observed variance
  NewMat[i,3] <- mean(abs(TestMat[,i] -mean(TestMat[,i])) > VecVar[i]) # a way to
calculate a p value for the bootstrap
}
NewMat
```

```
##       [,1]          [,2]                 [,3]
##  [1,] "ProteinHr0"  "0.170660464718401"  "0.146"
##  [2,] "ProteinHr4"  "0.0424580720160553" "0.125"
##  [3,] "ProteinHr24" "0.430437166002364"  "0.165"
##  [4,] "PODHr0"      "0.540646931200331"  "0.191"
##  [5,] "PODHr4"      "0.50298994687696"   "0.165"
##  [6,] "PODHr24"     "0.124187437998476"  "0.164"
##  [7,] "PPOHr0"      "0.392950708183783"  "0.173"
##  [8,] "PPOHr4"      "0.307691882221562"  "0.165"
##  [9,] "PPOHr24"     "0.290872993105146"  "0.165"
```

According to this metric, the variances do not differ between the bootstrapped test and the observed variance.

Levene's Test resample

I resampled the native range, combined it with the invasive range dataset and then ran the LeveneTest to compare variances

```
N = 1000
TestMat <- matrix(rep(0, 9*N), ncol = 9)
for (i in 1:N){
  pops <- sample(unique(BV_Nat$Genotype), 11, replace = T) # Sample populations r
andomly
  BV_Nat2 <- BV_Nat[BV_Nat$Genotype %in% pops, ] # Narrow dataset
  FullBV <- rbind(BV_Nat2, BV_Inv)
  test <- leveneTests(FullBV[,6:14], FullBV$Range)
  FStat <- test[,3] # Gets the F statistic from the Levene Test
  for (j in 1:9){
    TestMat[i, j] <- FStat[j]
  }
}

ColnNames <- colnames(BV_Nat)
ColnNames <- ColnNames[-(1:5)]
testObs <- leveneTests(BiocAssaysWide[,6:14], BiocAssaysWide$Range)
FStatObs <- testObs[,3] # Observed F statistic

FMat <- matrix(rep(0, 18), ncol = 2)
for (i in 1:9) {
  FMat[i,1] <- ColnNames[i]
  FMat[i,2] <- mean(abs(TestMat[,i] -mean(TestMat[,i])) > FStatObs[i]) # p value
for bootstrapped difference between groups
}
FMat
```

```
##       [,1]          [,2]
##  [1,] "ProteinHr0"  "0.142"
##  [2,] "ProteinHr4"  "0.789"
##  [3,] "ProteinHr24" "0.987"
##  [4,] "PODHr0"      "0.104"
##  [5,] "PODHr4"      "0.373"
##  [6,] "PODHr24"     "0.134"
##  [7,] "PPOHr0"      "0.09"
##  [8,] "PPOHr4"      "0.079"
##  [9,] "PPOHr24"     "0.483"
```

Under this metric there is also not a significant difference between group variances

Bootstrap univariate MANOVAs

I resampled the native range and ran univariate MANOVAs. I'm going to run all of the code and then show the outputs together after.

This section of code is also used to run the repeated measures MANOVA on the full dataset

```
block <- ordered(c("Hr0", "Hr4", "Hr24"), levels = c("Hr0", "Hr4", "Hr24"))
contrasts(block) <- matrix(c(-1,1,0,0,-1,1), ncol = 2)
idata <- data.frame(block)
```

Bootstrap Peroxidase Assay

```
PODWide_Nat <- subset(PODWide, Range == "Native")
PODWide_Inv <- subset(PODWide, Range == "Invasive")

N = 1000
PodMat <- matrix(rep(0, 6*N), ncol = 6)

for (i in 1:N) {
  outtests <- car:::print.Anova.mlm
  body(outtests)[[16]] <- quote(invisible(tests))
  body(outtests)[[15]] <- NULL
  pops <- sample(unique(PODWide_Nat$Genotype), 11, replace = T)
  PODWide_Nat2 <- PODWide_Nat[PODWide_Nat$Genotype %in% pops, ] # Narrow dataset
  FullPODWide <- rbind(PODWide_Nat2, PODWide_Inv)
  PodMod1 <- lm(cbind(PODHr0, PODHr4, PODHr24) ~ Range, data = FullPODWide)
  tab <- lapply(c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),  function(i)  ou
ttests(Anova(PodMod1, idata = idata, idesign = ~ block, test.statistic=i)))
  tab <- do.call(rbind, tab)
  Fval <- tab[,3]
  Pval <- tab[,6]
  for (j in 1:3){
    X = j + 1
    Y = j + 3
    PodMat[i,j] <- Fval[X]
    PodMat[i,Y] <- Pval[X]
  }
}
```

Bootstrap Polyphenol Oxidase Assay

```
PPOWide_Nat <- subset(PPOWide, Range == "Native")
PPOWide_Inv <- subset(PPOWide, Range == "Invasive")
N = 1000
PpoMat <- matrix(rep(0, 6*N), ncol = 6)

for (i in 1:N) {
  outtests <- car:::print.Anova.mlm
  body(outtests)[[16]] <- quote(invisible(tests))
  body(outtests)[[15]] <- NULL
  pops <- sample(unique(PPOWide_Nat$Genotype), 11, replace = T)
  PPOWide_Nat2 <- PPOWide_Nat[PPOWide_Nat$Genotype %in% pops, ] # Narrow dataset
  FullPPOWide <- rbind(PPOWide_Nat2, PPOWide_Inv)
  PpoMod1 <- lm(cbind(PPOHr0, PPOHr4, PPOHr24) ~ Range, data = FullPPOWide)
  tab <- lapply(c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),  function(i)  ou
ttests(Anova(PpoMod1, idata = idata, idesign = ~ block, test.statistic=i)))
  tab <- do.call(rbind, tab)
  Fval <- tab[,3]
  Pval <- tab[,6]
  for (j in 1:3){
    X = j + 1
    Y = j + 3
    PpoMat[i,j] <- Fval[X]
    PpoMat[i,Y] <- Pval[X]
  }
}
```

Bootstrap Protein Quantification Assay

```
ProWide_Nat <- subset(ProWide, Range == "Native")
ProWide_Inv <- subset(ProWide, Range == "Invasive")
N = 1000
ProMat <- matrix(rep(0, 6*N), ncol = 6)

for (i in 1:N) {
  outtests <- car:::print.Anova.mlm
  body(outtests)[[16]] <- quote(invisible(tests))
  body(outtests)[[15]] <- NULL
  pops <- sample(unique(PODWide_Nat$Genotype), 11, replace = T)
  ProWide_Nat2 <- ProWide_Nat[ProWide_Nat$Genotype %in% pops, ] # Narrow dataset
  FullProWide <- rbind(ProWide_Nat2, ProWide_Inv)
  ProMod1 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24) ~ Range, data = FullPr
oWide)
  tab <- lapply(c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),  function(i)  ou
ttests(Anova(ProMod1, idata = idata, idesign = ~ block, test.statistic=i)))
  tab <- do.call(rbind, tab)
  Fval <- tab[,3]
  Pval <- tab[,6]
  for (j in 1:3){
    X = j + 1
    Y = j + 3
    ProMat[i,j] <- Fval[X]
    ProMat[i,Y] <- Pval[X]
  }
}
```

This gives the observed output. It is arranged by F statistic of the observed model, p value of the observed model, and p value for the F statistic of the bootstrap

```
#pod
outtests <- car:::print.Anova.mlm
body(outtests)[[16]] <- quote(invisible(tests))
body(outtests)[[15]] <- NULL
PodMod1 <- lm(cbind(PODHr0, PODHr4, PODHr24) ~ Range, data = PODWide)
tab <- lapply(c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),  function(i)  outt
ests(Anova(PodMod1, idata = idata, idesign = ~ block, test.statistic=i)))
tab <- do.call(rbind, tab)

Fval <- tab[,3]
Pval <- tab[,6]
PODtrue <- data.frame(Fval[2:4], Pval[2:4])
rownames(PODtrue) <- c("Range", "block", "Range:block")
names(PODtrue)[names(PODtrue) == 'Fval.2.4.'] <- 'test.statistic'
names(PODtrue)[names(PODtrue) == 'Pval.2.4.'] <- 'p.value'

for (i in 1:3) {
  PODtrue$BootF[i] <-mean(abs(PodMat[,i] - mean(PodMat[,i])) > PODtrue$test.stati
stic[i])
}

# PPO
outtests <- car:::print.Anova.mlm
body(outtests)[[16]] <- quote(invisible(tests))
body(outtests)[[15]] <- NULL
PpoMod1 <- lm(cbind(PPOHr0, PPOHr4, PPOHr24) ~ Range, data = PPOWide)
tab <- lapply(c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),  function(i)  outt
ests(Anova(PpoMod1, idata = idata, idesign = ~ block, test.statistic=i)))
tab <- do.call(rbind, tab)

Fval <- tab[,3]
Pval <- tab[,6]
PPOtrue <- data.frame(Fval[2:4], Pval[2:4])
rownames(PPOtrue) <- c("Range", "block", "Range:block")
names(PPOtrue)[names(PPOtrue) == 'Fval.2.4.'] <- 'test.statistic'
names(PPOtrue)[names(PPOtrue) == 'Pval.2.4.'] <- 'p.value'

for (i in 1:3) {
  PPOtrue$BootF[i] <-mean(abs(PpoMat[,i] - mean(PpoMat[,i])) > PPOtrue$test.stati
stic[i])
}

# PROTEIN
outtests <- car:::print.Anova.mlm
body(outtests)[[16]] <- quote(invisible(tests))
body(outtests)[[15]] <- NULL
ProMod1 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24) ~ Range, data = ProWide)
tab <- lapply(c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),  function(i)  outt
```

```
ests(Anova(ProMod1, idata = idata, idesign = ~ block, test.statistic=i)))
tab <- do.call(rbind, tab)

Fval <- tab[,3]
Pval <- tab[,6]
PROtrue <- data.frame(Fval[2:4], Pval[2:4])
rownames(PROtrue) <- c("Range", "block", "Range:block")
names(PROtrue)[names(PROtrue) == 'Fval.2.4.'] <- 'test.statistic'
names(PROtrue)[names(PROtrue) == 'Pval.2.4.'] <- 'p.value'

for (i in 1:3) {
  PROtrue$BootF[i] <-mean(abs(ProMat[,i] - mean(ProMat[,i])) > PROtrue$test.stati
stic[i])
}

PODtrue
```

```
##             test.statistic      p.value BootF
## Range            0.2437189 6.219826e-01 0.959
## block          496.8575493 4.781492e-86 0.000
## Range:block      4.4457017 1.271047e-02 0.097
```

```
PPOtrue
```

```
##             test.statistic      p.value BootF
## Range           15.6101679 1.022742e-04 0.025
## block           10.2693754 5.256175e-05 0.032
## Range:block      0.3582317 6.992842e-01 0.846
```

```
PROtrue
```

```
##             test.statistic      p.value BootF
## Range            4.6333571 3.234941e-02 0.511
## block           37.8962384 4.978924e-15 0.006
## Range:block      0.4679875 6.268316e-01 0.837
```

p values for the bootstrapped values seem elevated compared to the observed value.

Repeated measures MANOVAs for each assay.

Diagnostic steps for these assays can be found in the most Induced vs constitutive pdf.

Peroxidase

1. Does the invasive range differ in POD activity significantly from the native range?

2. Do the two ranges respond similarly over time?

```
PodMod1 <- lm(cbind(PODHr0, PODHr4, PODHr24) ~ Range, data = PODWide)
Anova(PodMod1, idata = idata, idesign = ~ block)
```

```
##
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
##             Df test stat approx F num Df den Df  Pr(>F)
## (Intercept)  1   0.96493   6630.6      1    241 < 2e-16 ***
## Range        1   0.00101      0.2      1    241 0.62198
## block        1   0.80547    496.9      2    240 < 2e-16 ***
## Range:block  1   0.03572      4.4      2    240 0.01271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Range is not significant but there are differences between the two over time.

Polyphenol Oxidase

1. Does the invasive range differ in PPO activity significantly from the native range?
2. Do the two ranges respond similarly over time?

```
PpoMod1 <- lm(cbind(PPOHr0, PPOHr4, PPOHr24) ~ Range, data = PPOWide)
Anova(PpoMod1, idata = idata, idesign = ~ block)
```

```
##
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
##             Df test stat approx F num Df den Df    Pr(>F)
## (Intercept)  1   0.88625  1877.64      1    241 < 2.2e-16 ***
## Range        1   0.06083    15.61      1    241 0.0001023 ***
## block        1   0.07883    10.27      2    240 5.256e-05 ***
## Range:block  1   0.00298     0.36      2    240 0.6992842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PPO activity varies between the ranges but both respond similarly over time

Protein Quantification

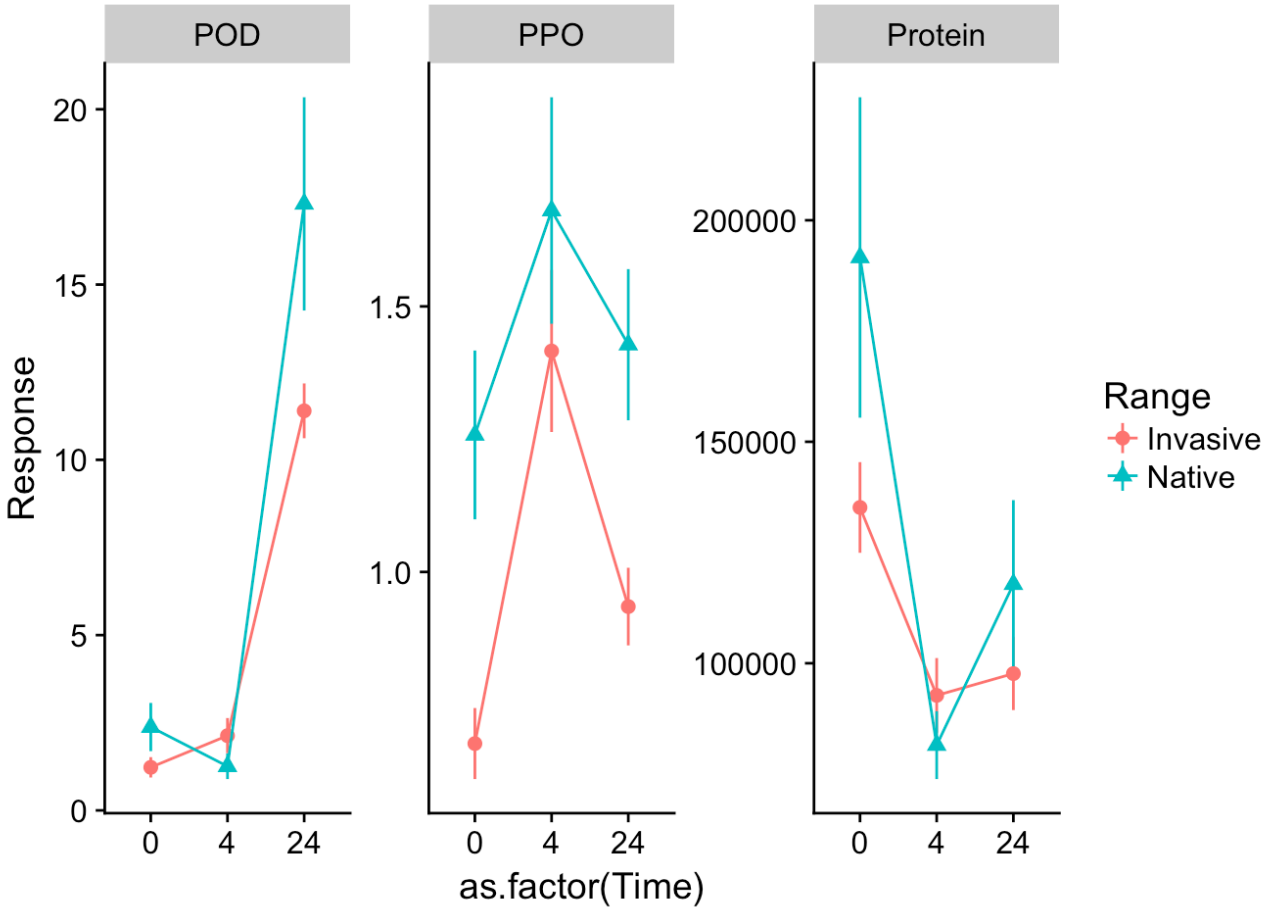
1. Does the invasive range differ in protein production significantly from the native range?
2. Do the two ranges respond similarly over time?

```
ProMod1 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24) ~ Range, data = ProWide)
Anova(ProMod1, idata = idata, idesign = ~ block)
```

```
##
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
##             Df test stat approx F num Df den Df    Pr(>F)
## (Intercept)  1   0.99824   136641      1    241 < 2.2e-16 ***
## Range        1   0.01886        5      1    241   0.03235 *
## block        1   0.24001       38      2    240 4.979e-15 ***
## Range:block  1   0.00388        0      2    240   0.62683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

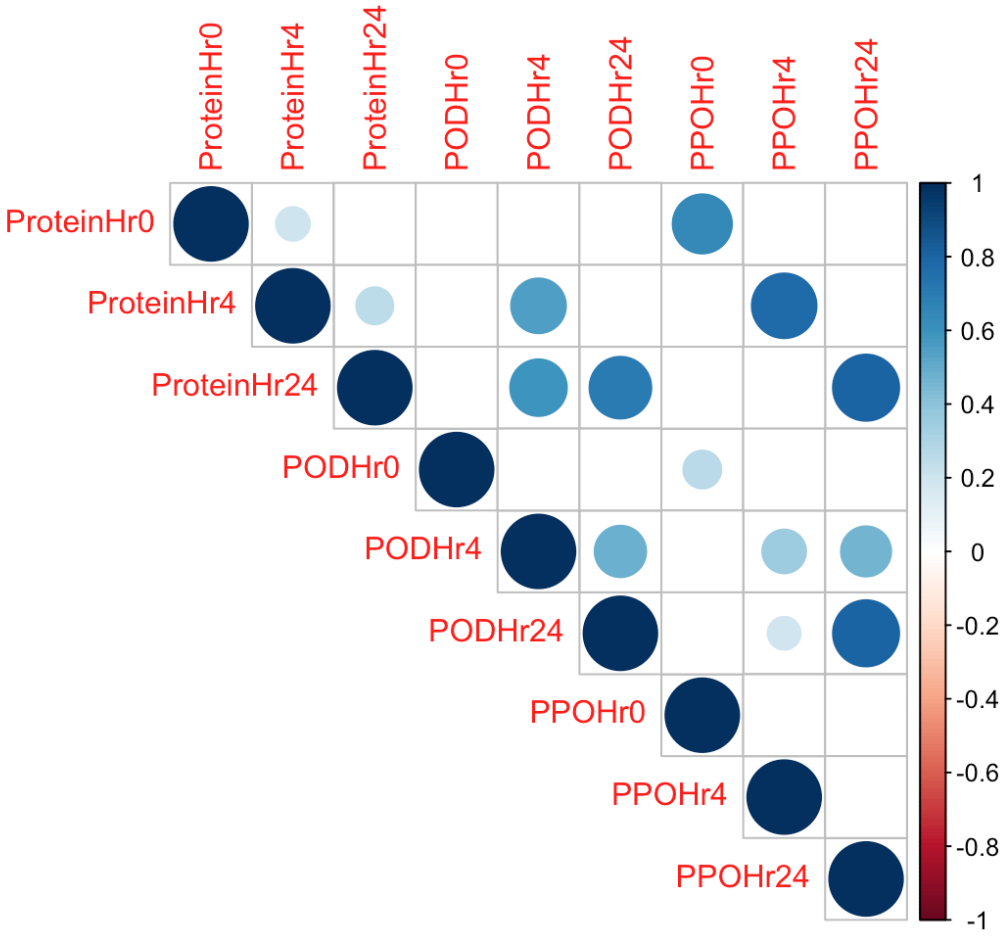
Protein production varies between the ranges but both respond similarly over time

Plot of assay results



Correlations between all the assays and times

A pearson correlation plot where the intensity of the color and size of the circle indicate correlations. Blue are positive and Red are negative. Only significant correlations are shown (Holm's adjusted p values < 0.05)


ProteinHr0
ProteinHr4
ProteinHr24
PODHr0
PODHr4
PODHr24
PPOHr0
PPOHr4
PPOHr24
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1