Preliminaries

This appendix is the first of multiple appendices which present the code used to produce the results discussed in this document. There is a code appendix for each of the applicable chapters. This appendix presents the code used for the cosinor models chapter.

Firstly, the `CircadianTools` package was loaded. The count data (Laurasmappings) was read in and the CT18.4 column was removed. This column was removed as it is mostly zeroes due to a technical problem which occurred during the experiment. Any genes which showed zero activity for all readings were removed: as is customary. There were 91,311 genes in the dataset after removal.

```r
library(CircadianTools)
# Read in count data
Laurasmappings <- read.csv(
  "~/MEGA/Uni/Masters/Diss/Stats/Raw_Data/Laurasmappings.csv",
  stringsAsFactors = FALSE)


Laurasmappings$CT18.4 <- NULL # Remove column of zeroes
Laurasmappings <- GeneClean(Laurasmappings) # Remove genes which show no activity
nrow(Laurasmappings) # Print the number of genes in the reduced dataset
```

```
## [1] 91311
```

Cosinor 24 Hour (Circadian) Period

Cosinor models with 24 hour periods were fitted and F-tests were carried out on these models. The resultant p-values are saved alongside gene names in the `cosinor.24` object. The `system.time` function is used to show the speed of fitting cosinor models to all 91,311 genes.

```r
# Fit cosinor models with 24 hour period and calculate time taken.
system.time(cosinor.24 <- CosinorAnalysis(Laurasmappings, period = 24,
                                           progress = FALSE, print = FALSE)
            )
```

```
##    user  system elapsed
## 257.550   0.923 259.421
```

The p-values needed to be adjusted using a bonferroni correction due to the large number of genes being tested. After the adjustment, the genes which are significant at the 5% level were found.

```r
# Apply bonferroni correction
cosinor.24$pVal <- p.adjust(cosinor.24$pVal, method = "bonferroni")
# Sort by smallest p-value
cosinor.24 <- cosinor.24[order(cosinor.24$pVal, decreasing = FALSE), ]
# Find genes significant at the 5% level
sig.24 <- subset(cosinor.24, pVal <= 0.05)
sig.24 # Print these genes
```

```
##                  sample          pVal
## 90568    comp99801_c1_seq1 2.246277e-06
## 57405    comp80445_c1_seq1 3.783256e-05
## 66624    comp89211_c0_seq2 1.469563e-04
## 81165    comp96806_c0_seq1 4.191294e-04
## 86918    comp98714_c0_seq2 6.070344e-04
## 81      comp100026_c0_seq2 6.175014e-04
## 8815    comp102333_c0_seq8 1.638949e-03
## 8813    comp102333_c0_seq2 2.707901e-03
## 28969    comp23718_c0_seq1 5.662856e-03
```

```
## 8814   comp102333_c0_seq21 6.844270e-03
## 6583    comp101772_c1_seq2 1.034375e-02
## 50942    comp71855_c0_seq1 1.053243e-02
## 86530    comp98599_c2_seq1 1.281385e-02
## 43801       comp606_c0_seq1 1.745201e-02
## 8816     comp102333_c0_seq9 4.192379e-02
```

A FASTA file was created using the FastaSub function in `CircadianTools`. This FASTA file was created by including only the sequences for the samples which were found to be significant with 24 hour cosinor models from a FASTA file which detailed the transcriptome of *Talitrus saltaor*. This FASTA file was then 'BLASTed' to search for proteins.

```r
# Read in main FASTA file containing transcriptome
main.fasta <- seqinr::read.fasta(
  "~/MEGA/Uni/Masters/Diss/Stats/Raw_Data/JoesTranscriptomeMin300bp 2.fasta")

# Create a FASTA file which only contains sequences for the samples which were
# found to be significant with 24 hour cosinor models
sig.24.fasta <- FastaSub(gene.names = sig.24$sample, fasta.file = main.fasta,
          filename = "sig.24")
```

The below code saves plots of the cosinor models for these genes to a folder in the working directory called "cosinor_24".

```r
CosinorSignificantPlot(cosinor.24, Laurasmappings, number = nrow(sig.24),
                       print = FALSE, save = TRUE, path = "cosinor_24")
```

After viewing the plots, four plots were chosen for inclusion in the main document and were saved as print quality png files.

```r
gene.names <- c("comp99801_c1_seq1", "comp102333_c0_seq2", "comp98599_c2_seq1",
                "comp98714_c0_seq2" )

# Change the size of the text in the plots
txt.size <- ggplot2::theme(text = ggplot2::element_text(size = 7))


p1 <- CosinorPlot(gene.names[1], Laurasmappings) + txt.size
p2 <- CosinorPlot(gene.names[2], Laurasmappings) + txt.size
p3 <- CosinorPlot(gene.names[3], Laurasmappings) + txt.size
p4 <- CosinorPlot(gene.names[4], Laurasmappings) + txt.size

p <- gridExtra::grid.arrange(p1, p2, p3, p4, ncol = 2 , nrow = 2)
```
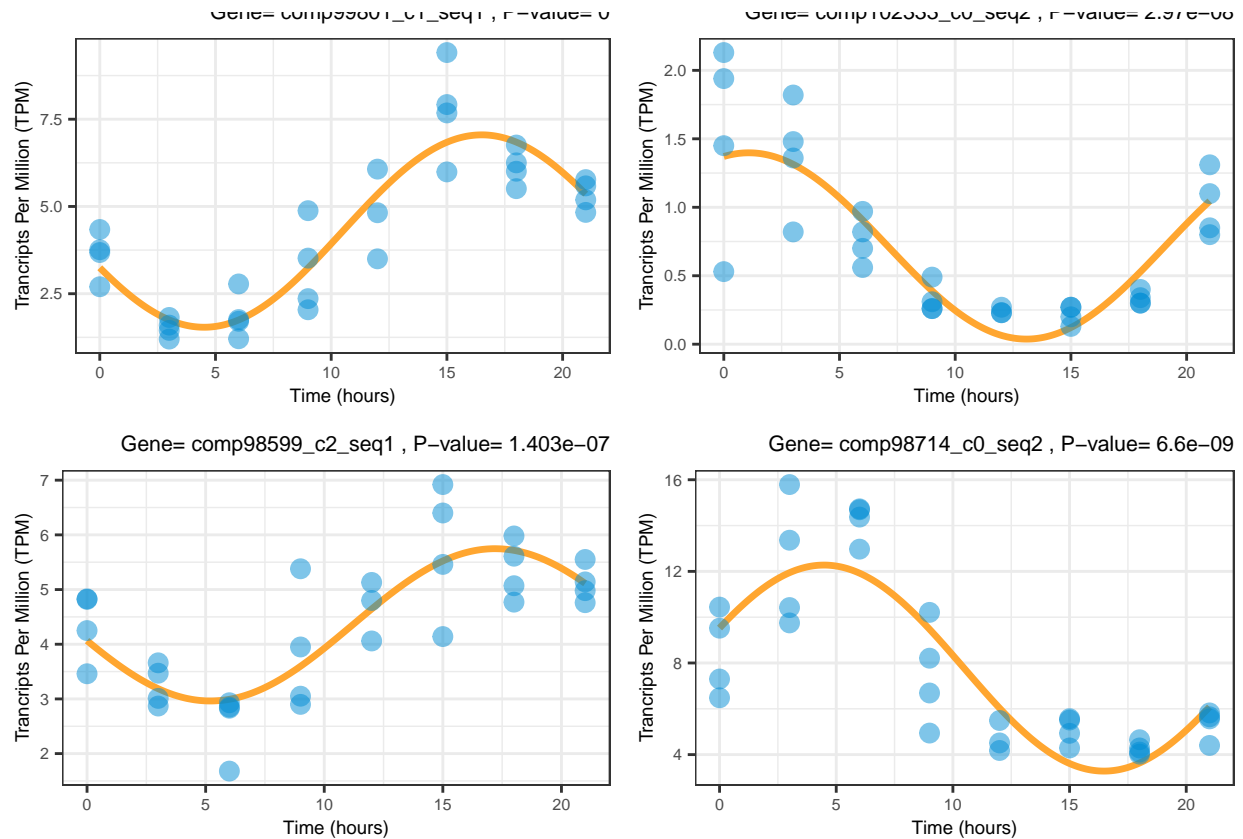
```
ggplot2::ggsave("cosinor_24.png", plot = p, width = 12, height = 10,
                units = "in")
```
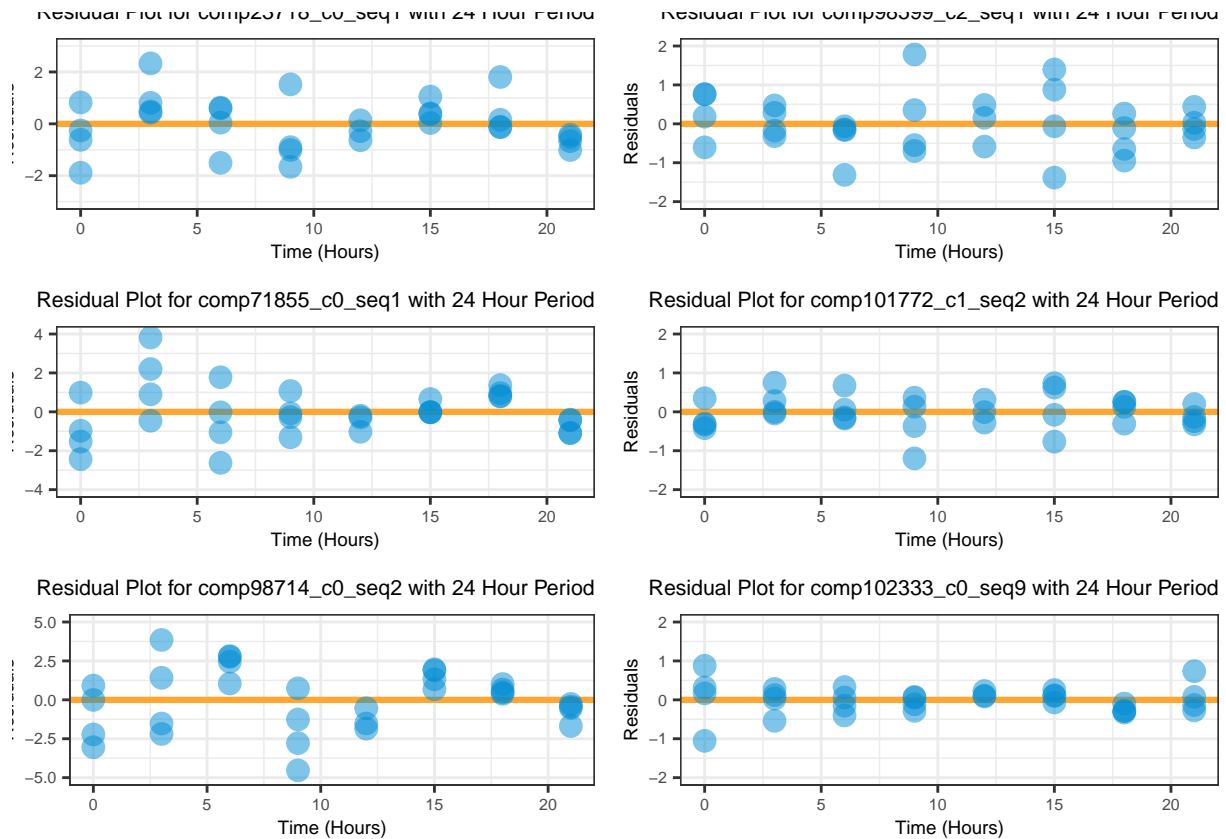
The residuals for these significant cosinor models were plotted and saved to a folder in the working directory called "residuals".

```
genes.sig <- GeneSub(sig.24, Laurasmappings)
CosinorResidualDatasetPlot(genes.sig, print = FALSE, save = TRUE,
                           path = "residuals")
```

After investigating the plots of the residuals, some genes appeared to have sinusoidal residuals whilst others did not. As a result, the plots of residuals for 6 genes were were saved for use as a figure. 3 of the genes were judged to have sinusoidal residuals whilst the remaining three were judged to not have this attribute.

```
p1 <- CosinorResidualPlot("comp23718_c0_seq1", Laurasmappings) + txt.size
p2 <- CosinorResidualPlot("comp71855_c0_seq1", Laurasmappings) + txt.size
p3 <- CosinorResidualPlot("comp98714_c0_seq2", Laurasmappings) + txt.size
p4 <- CosinorResidualPlot("comp98599_c2_seq1", Laurasmappings) + txt.size
p5 <- CosinorResidualPlot("comp101772_c1_seq2", Laurasmappings) + txt.size
p6 <- CosinorResidualPlot("comp102333_c0_seq9", Laurasmappings) + txt.size

arranged <- gridExtra::grid.arrange(p1,p4,p2,p5,p3,p6, ncol=2)
```

Residual Plot for comp71855_c0_seq1 with 24 Hour Period

Residual Plot for comp101772_c1_seq2 with 24 Hour Period

Residual Plot for comp98714_c0_seq2 with 24 Hour Period

Residual Plot for comp102333_c0_seq9 with 24 Hour Period

```
ggplot2::ggsave("residuals.png", arranged, width=12, height=10, units="in",
                dpi=300)
```

Investigating Genes with Sinusoidal Residuals

The three genes which were judged to have sinsoidal characteristics were then fitted with cosinor models with both circadian and circatidal terms. ANOVA was used to see if the addition of these extra terms resulted in a signficantly better fit.

Comp23718_c0_seq1

```
MultiCosinorTest("comp23718_c0_seq1", Laurasmappings, period.1 = 24,
                 period.2 = 12.4 )
```

```
## The anova table for the simple model is given by:
## Analysis of Variance Table
##
## Response: activity
##           Df Sum Sq Mean Sq F value      Pr(>F)
## rrr.1      1 50.411  50.411  50.044 1.073e-07 ***
## sss.1      1 13.673  13.673  13.573 0.0009732 ***
## Residuals 28 28.205   1.007
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The anova table for the more complex model is given by:
## Analysis of Variance Table
##
## Response: activity
```

```
##             Df Sum Sq Mean Sq F value    Pr(>F)
## rrr.1      1 50.411  50.411 65.7513 1.372e-08 ***
## sss.1      1 13.673  13.673 17.8338 0.0002608 ***
## rrr.2      1  0.076   0.076  0.0997 0.7546551
## sss.2      1  8.195   8.195 10.6886 0.0030321 **
## Residuals 26 19.934   0.767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Carrying out ANOVA on both models produces the following table :
## Analysis of Variance Table
##
## Model 1: activity ~ rrr.1 + sss.1
## Model 2: activity ~ rrr.1 + sss.1 + rrr.2 + sss.2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 28.205
## 2     26 19.934  2    8.2714 5.3942 0.01098 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The ANOVA results have been saved.
```

Comp71855_c0_seq1

```
MultiCosinorTest("comp71855_c0_seq1", Laurasmappings, period.1 = 24,
                 period.2 = 12.4)
```

```
## The anova table for the simple model is given by:
## Analysis of Variance Table
##
## Response: activity
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## rrr.1      1 104.718 104.718 55.1337 4.364e-08 ***
## sss.1      1   8.569   8.569  4.5117   0.04263 *
## Residuals 28  53.182   1.899
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The anova table for the more complex model is given by:
## Analysis of Variance Table
##
## Response: activity
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## rrr.1      1 104.718 104.718 65.4088 1.441e-08 ***
## sss.1      1   8.569   8.569  5.3525   0.02886 *
## rrr.2      1   1.284   1.284  0.8020   0.37872
## sss.2      1  10.272  10.272  6.4163   0.01768 *
## Residuals 26  41.625   1.601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Carrying out ANOVA on both models produces the following table :
## Analysis of Variance Table
##
## Model 1: activity ~ rrr.1 + sss.1
## Model 2: activity ~ rrr.1 + sss.1 + rrr.2 + sss.2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 53.182
## 2     26 41.625  2    11.556 3.6091 0.04138 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The ANOVA results have been saved.
```

Comp98714_c0_seq2

```
MultiCosinorTest("comp98714_c0_seq2", Laurasmappings, period.1 = 24,
                  period.2 = 12.4)
```

```
## The anova table for the simple model is given by:
## Analysis of Variance Table
##
## Response: activity
##            Df Sum Sq Mean Sq F value    Pr(>F)
## rrr.1       1 274.78 274.776  68.132 5.54e-09 ***
## sss.1       1  45.69  45.690  11.329  0.00223 **
## Residuals 28 112.92   4.033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The anova table for the more complex model is given by:
## Analysis of Variance Table
##
## Response: activity
##            Df  Sum Sq Mean Sq  F value     Pr(>F)
## rrr.1       1 274.776 274.776 106.0460 1.147e-10 ***
## sss.1       1  45.690  45.690  17.6335 0.0002776 ***
## rrr.2       1  12.303  12.303   4.7482 0.0385810 *
## sss.2       1  33.252  33.252  12.8332 0.0013752 **
## Residuals 26  67.369   2.591
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Carrying out ANOVA on both models produces the following table :
## Analysis of Variance Table
##
## Model 1: activity ~ rrr.1 + sss.1
## Model 2: activity ~ rrr.1 + sss.1 + rrr.2 + sss.2
##   Res.Df     RSS Df Sum of Sq      F   Pr(>F)
## 1     28 112.924
## 2     26  67.369  2    45.555 8.7907 0.001213 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The ANOVA results have been saved.
```

Investigating Genes With Non-Sinusoidal Residuals

The Three genes which were judged to not have sinusodial residuals were then investigated.

Comp98599_c2_seq1

```
MultiCosinorTest("comp98599_c2_seq1", Laurasmappings, period.1 = 24,
                  period.2 = 12.4)
```

```
## The anova table for the simple model is given by:
## Analysis of Variance Table
##
```

```
## Response: activity
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## rrr.1       1 29.6726 29.6726 56.0481 3.735e-08 ***
## sss.1       1  1.2592  1.2592  2.3784    0.1342
## Residuals 28 14.8236  0.5294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The anova table for the more complex model is given by:
## Analysis of Variance Table
##
## Response: activity
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## rrr.1       1 29.6726 29.6726 56.7985 5.317e-08 ***
## sss.1       1  1.2592  1.2592  2.4103    0.1326
## rrr.2       1  1.1592  1.1592  2.2189    0.1484
## sss.2       1  0.0815  0.0815  0.1560    0.6960
## Residuals 26 13.5829  0.5224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Carrying out ANOVA on both models produces the following table :
## Analysis of Variance Table
##
## Model 1: activity ~ rrr.1 + sss.1
## Model 2: activity ~ rrr.1 + sss.1 + rrr.2 + sss.2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 14.824
## 2     26 13.583  2    1.2407 1.1875  0.321
## The ANOVA results have been saved.
```

Comp101772_c1_seq2

```
MultiCosinorTest("comp101772_c1_seq2", Laurasmappings, period.1 = 24,
                 period.2 = 12.4)
```

```
## The anova table for the simple model is given by:
## Analysis of Variance Table
##
## Response: activity
##             Df Sum Sq Mean Sq F value    Pr(>F)
## rrr.1       1 4.5071  4.5071  22.872 5.027e-05 ***
## sss.1       1 7.2688  7.2688  36.887 1.502e-06 ***
## Residuals 28 5.5176  0.1971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The anova table for the more complex model is given by:
## Analysis of Variance Table
##
## Response: activity
##             Df Sum Sq Mean Sq F value    Pr(>F)
## rrr.1       1 4.5071  4.5071 24.3783 3.964e-05 ***
## sss.1       1 7.2688  7.2688 39.3165 1.231e-06 ***
## rrr.2       1 0.0211  0.0211  0.1142   0.73808
## sss.2       1 0.6896  0.6896  3.7301   0.06441 .
## Residuals 26 4.8069  0.1849
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Carrying out ANOVA on both models produces the following table :
## Analysis of Variance Table
##
## Model 1: activity ~ rrr.1 + sss.1
## Model 2: activity ~ rrr.1 + sss.1 + rrr.2 + sss.2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 5.5176
## 2     26 4.8069  2   0.71074 1.9222 0.1665
## The ANOVA results have been saved.
```

Comp102333__c0__seq9

```
MultiCosinorTest("comp102333_c0_seq9", Laurasmappings, period.1 = 24,
                 period.2 = 12.4)
```

```
## The anova table for the simple model is given by:
## Analysis of Variance Table
##
## Response: activity
##           Df Sum Sq Mean Sq F value    Pr(>F)
## rrr.1      1 0.2780  0.2780  2.0105    0.1672
## sss.1      1 6.8318  6.8318 49.3998 1.208e-07 ***
## Residuals 28 3.8723  0.1383
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## The anova table for the more complex model is given by:
## Analysis of Variance Table
##
## Response: activity
##           Df Sum Sq Mean Sq F value    Pr(>F)
## rrr.1      1 0.2780  0.2780  1.9827    0.1710
## sss.1      1 6.8318  6.8318 48.7158 2.068e-07 ***
## rrr.2      1 0.2071  0.2071  1.4768    0.2352
## sss.2      1 0.0190  0.0190  0.1356    0.7157
## Residuals 26 3.6462  0.1402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Carrying out ANOVA on both models produces the following table :
## Analysis of Variance Table
##
## Model 1: activity ~ rrr.1 + sss.1
## Model 2: activity ~ rrr.1 + sss.1 + rrr.2 + sss.2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 3.8723
## 2     26 3.6462  2   0.22611 0.8062 0.4574
## The ANOVA results have been saved.
```

Cosinor 12.4 (Circatidal) Period

Cosinor models with periods of 12.4 hours were then be fitted. The 25 best fitting models were then found.

```
cosinor.12.4 <- CosinorAnalysis(Laurasmappings, period = 12.4, progress = FALSE,
                                print= FALSE )
cosinor.12.4 <- cosinor.12.4[order(cosinor.12.4$pVal, decreasing = FALSE), ]
```

```
sig.12.4 <- cosinor.12.4[1:25, ]
sig.12.4
```

```
##                     sample         pVal
## 84017  comp97780_c0_seq11 0.0001188788
## 40770   comp53017_c0_seq2 0.0001813181
## 47742   comp66784_c0_seq1 0.0001975727
## 32217   comp28197_c0_seq1 0.0002257871
## 13449 comp1187601_c0_seq1 0.0003098192
## 40902  comp533893_c0_seq1 0.0003403444
## 60758   comp84247_c0_seq1 0.0003462262
## 20107   comp15545_c0_seq1 0.0003619025
## 82934   comp97430_c3_seq8 0.0003650143
## 39427  comp489773_c0_seq1 0.0003777484
## 89458   comp99487_c0_seq2 0.0003795692
## 57101   comp80084_c0_seq1 0.0003968281
## 36320   comp37769_c0_seq1 0.0003979793
## 71437   comp92563_c0_seq4 0.0004666427
## 23690   comp17963_c0_seq1 0.0004721746
## 54310   comp76855_c0_seq2 0.0005164138
## 88158   comp99093_c0_seq2 0.0006021586
## 78647   comp95889_c0_seq2 0.0006182605
## 85027  comp98126_c0_seq33 0.0006260849
## 13526 comp1189457_c0_seq1 0.0006328214
## 63453   comp86536_c0_seq1 0.0006437563
## 46227   comp64680_c0_seq1 0.0006444333
## 16688 comp1372561_c0_seq1 0.0006836282
## 44567   comp617341_c0_seq1 0.0007331367
## 86925   comp98716_c0_seq1 0.0007550429
```

A FASTA file was created using the FastaSub function in `CircadianTools` similar to the signficant circadian genes was used for a BLAST search.

```
sig.12.4.fasta <- FastaSub(gene.names = sig.12.4$sample, fasta.file = main.fasta,
        filename = "sig.24")
```

Plots of the best fitting genes were plotted and saved to the `cosinor_12.4` folder in the working directory so they could be judged for circatidal characteristics.

```
CosinorSignificantPlot(cosinor.12.4, Laurasmappings, period = 12.4 ,number = 25,
                    print = FALSE, save = TRUE, path = "cosinor_12.4")
```

After viewing the plots, 4 were chosen for inclusion in the main document.

```
p1 <- CosinorPlot("comp97780_c0_seq11", Laurasmappings, period=12.4) +
        ggplot2::ggtitle("comp97780_c0_seq11, unadjusted P-Value: 1x10^-4")  +
        txt.size
p2 <- CosinorPlot("comp5046_c0_seq2", Laurasmappings, period=12.4) +
        ggplot2::ggtitle("comp5046_c0_seq2, unadjusted P-Value: 8x10^-4") +
        txt.size
p3 <- CosinorPlot("comp53017_c0_seq2", Laurasmappings, period=12.4) +
        ggplot2::ggtitle("comp53017_c0_seq2, unadjusted P-Value: 1x10^-4") +
        txt.size
p4 <- CosinorPlot("comp84247_c0_seq1", Laurasmappings, period=12.4) +
        ggplot2::ggtitle("comp84247_c0_seq1, unadjusted P-Value: 3x10^-4") +
        txt.size
```

```
arranged <- gridExtra::grid.arrange(p1,p2,p3,p4, ncol=2)
```
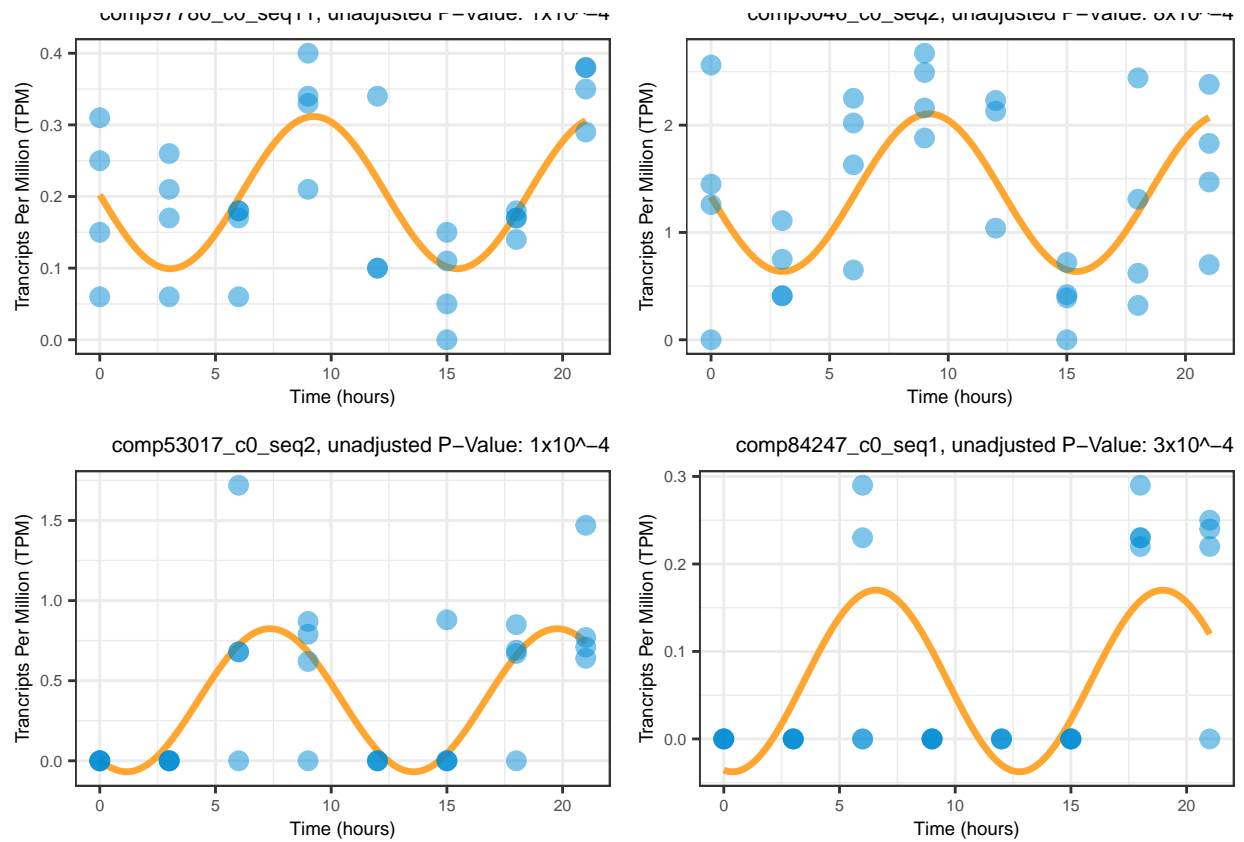


```
ggplot2::ggsave("cosinor_12.4.png", arranged,width=12, height=10, units="in",
                dpi=300)
```