

Levin *et al* reanalyses

This document can be executed to a pdf or html file in RStudio (or with the `knitr` package at the command line).

Reproducing previous results

The first goal is to reproduce the previous results to ensure that I understand their methods

Figure 4c data

```
Phase = as.matrix(read.table("data_processed/Figure_4c_mode.txt", header=T))
Ortho = as.matrix(read.table("data_processed/Figure_4c_orthology.txt", header=T))

dim(Phase)
```

```
## [1] 24717    10
```

```
dim(Ortho)
```

```
## [1] 24717    10
```

First, check to see if there are any elements that are 0 in one table but not the other.

```
sum((Phase>0) != (Ortho>0))
```

```
## [1] 0
```

There aren't. This verifies that every annotated ortholog present in each species is categorized as belonging to one of the three phases.

```
table(Phase)
```

```
## Phase
##      0      1      2      3
## 173779 37429 11697 24265
```

```
nrow(Phase[rowSums(Phase)>0,]) # number of nonzero rows
```

```
## [1] 21463
```

From Levin et al:

To compare gene expression across these species, we delineated 11,139 orthologous protein families, with each orthologous family having representatives from an average of six species.

Question: What is the source of the discrepancy between the reported number 11,139 in the manuscript and the number of genes (ie, rows) in this Figure_4c matrix, ie 21463? Were there additional filters applied?

C. elegans and D. melanogaster comparison

Here I try to replicate the example described by Levin et al.:

Finally, we measured the extent of evolutionary change within the two conserved phases and the mid-developmental transition by determining whether orthologues annotated for a particular temporal category in one species are also annotated to the same temporal category in another species. Figure 4c shows an example of this analysis for *D. melanogaster* and *C. elegans*. For 4,395 orthologues delineated between these two species, the early phase, mid-developmental transition, and the late phase expression account for 51%, 14%, and 35% of the *C. elegans* orthologues, respectively. A total of 28% of the orthologues are annotated to the early phase in both *C. elegans* and *D. melanogaster*, while by chance only 22% are expected given the fraction of genes in each category across the species (Fig. 4c). In contrast, 3% were expected to be conserved at the mid-developmental transition at random, and 3% were observed.

```
# Sample the columns for these two species
DC = Phase[,c(2,3)]
colnames(DC)
```

```
## [1] "Nematodes" "Arthropodes"
```

```
# sample only rows that have phase annotations for both
```

```
DC = DC[(DC[,1]>0) & (DC[,2]>0),]
nrow(DC)
```

```
## [1] 4395
```

So the number of genes selected is consistent with the reported number.

Now calculate the fraction of genes in each phase for each species.

```
colnames(DC)[1]
```

```
## [1] "Nematodes"
```

```
table(DC[,1])/sum(DC[,1])
```

```
##
##           1           2           3
## 0.27491367 0.07634435 0.19079921
```

Question: These values do not correspond to the 51%, 14%, and 35% reported in the manuscript for *C. elegans*. What is the source of the discrepancy?

```
colnames(DC)[2]
```

```
## [1] "Arthropodes"
```

```
table(DC[,2])/sum(DC[,2])
```

```
##
##           1           2           3
## 0.2343675 0.1167629 0.1773689
```

Figure 4d data

```
D = read.table("data_processed/Figure_4d_data.txt", header=T)
ks.test(D$Early, D$Mid, alternative="less")
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: D$Early and D$Mid
## D^- = 0.37778, p-value = 0.001625
## alternative hypothesis: the CDF of x lies below that of y
```

```
ks.test(D$Late, D$Mid, alternative="less")
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: D$Late and D$Mid
## D^- = 0.53333, p-value = 2.761e-06
## alternative hypothesis: the CDF of x lies below that of y
```

The KS tests as reproduced here are still significant, but the p values are much lower than those published. Levin et al. stated:

we found that the mid-developmental transition profiles are significantly less conserved than the early and late phase expression (Fig. 4d, $P < 10^{-6}$ compared with the early phase and $P < 10^{-12}$ with the late phase, Kolmogorov-Smirnov test)

Question: How was the published P value calculated? Was it two sided? Was it on a different data set?

Additional analyses

I conducted several additional analyses to address specific methodological concerns.

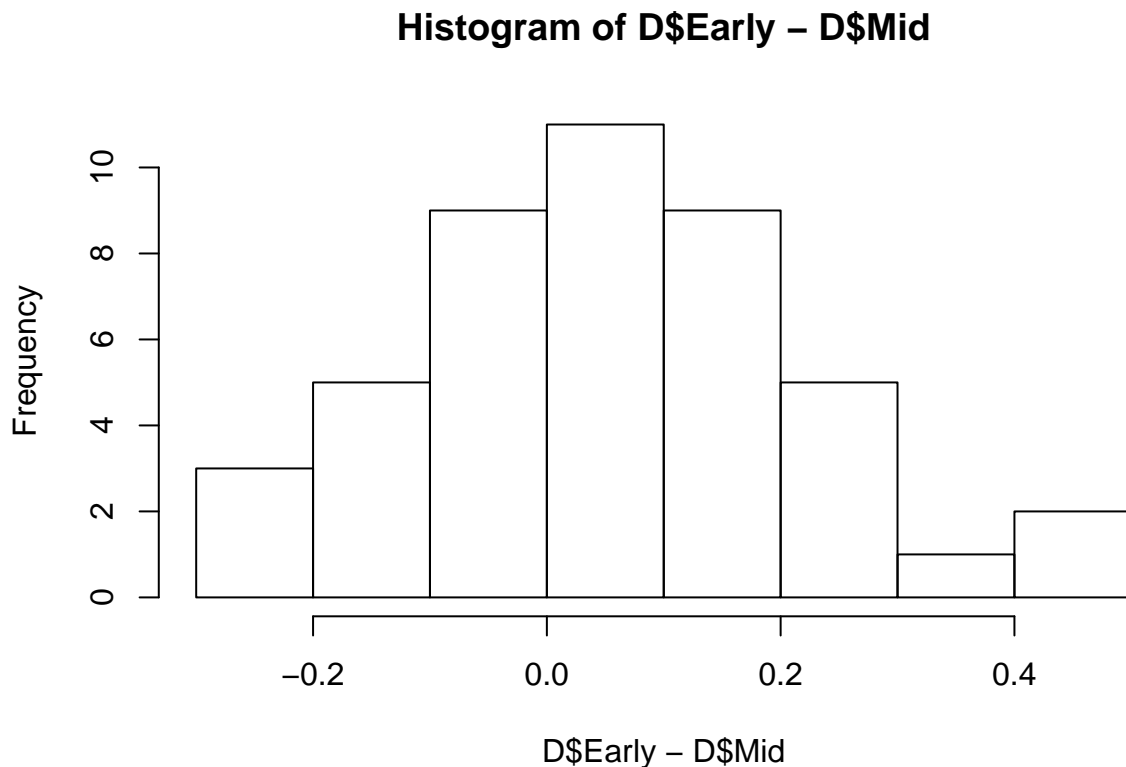
Test method

The primary results are based on the box plot in Figure 4d and the corresponding KS test. The box plots for each temporal phase largely overlap and there is little differentiation between the distributions apparent in this figure. The published KS test did reject the hypothesis that each is drawn from the same distribution. There are a couple problems with using the KS test in this context, though:

- The KS test doesn't just test whether one distribution is greater than the other, it also tests whether the shape of the distributions are the same. This is a much more specific null hypothesis than what is discussed in the paper. The null hypothesis would be rejected, for example, if the two distributions had the same mean but different variance.
- The samples in this dataset are matched (ie, for each pairwise comparison there is a early, mid, and late expression value). The KS test is for unmatched samples. In this instance, the question isn't whether the distribution of pairwise values as a higher cumulative distribution function for one phase verses the other, as the KS test can evaluate, but whether each value in one phases tends to be higher or lower than its matched value in another phase. Nonparametric tests for matched samples include the Sign test and the Wilcoxon Signed Rank Test. See, for example, this explanation.

Here I rerun the analyses taking the matched samples into account.

```
hist(D$Early - D$Mid)
```



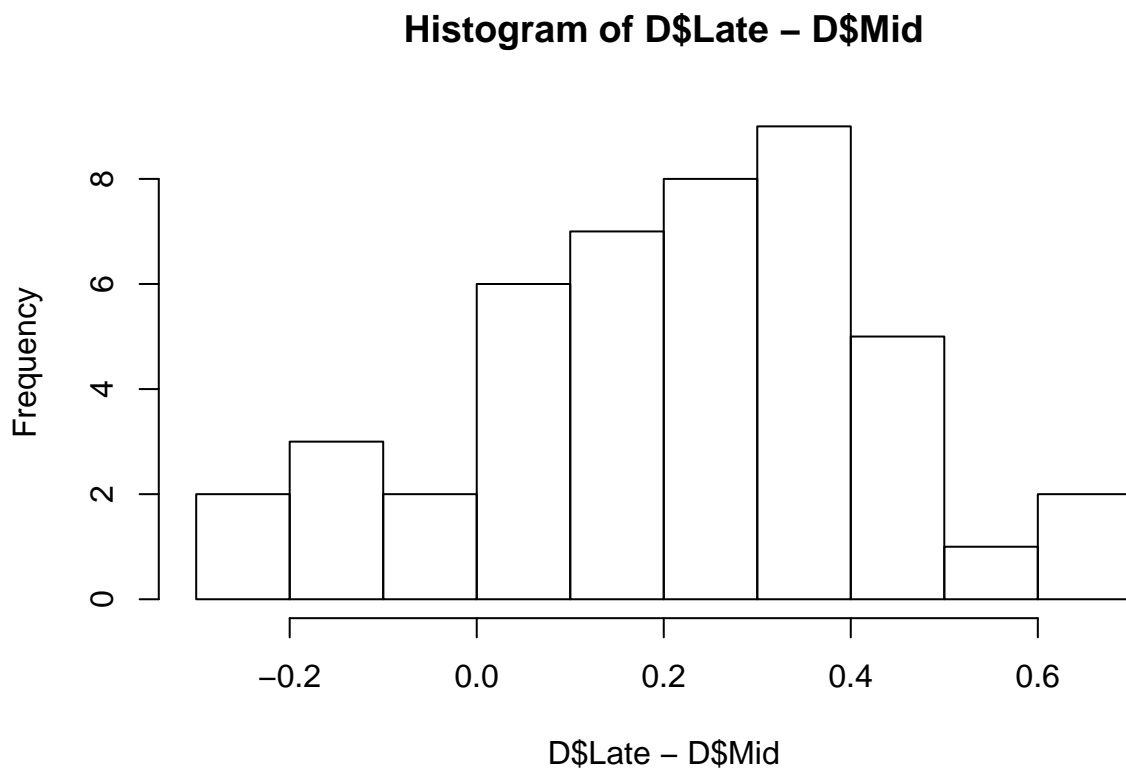
```
wilcox.test(D$Early, D$Mid, alternative="greater", paired=T)
```

```
##
## Wilcoxon signed rank test
##
## data: D$Early and D$Mid
## V = 699, p-value = 0.02011
## alternative hypothesis: true location shift is greater than 0
```

```
binom.test(sum((D$Early - D$Mid)>0), length(D$Early - D$Mid)) # sign test
```

```
##
## Exact binomial test
##
## data: sum((D$Early - D$Mid) > 0) and length(D$Early - D$Mid)
## number of successes = 28, number of trials = 45, p-value = 0.1352
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4654101 0.7623218
## sample estimates:
## probability of success
##      0.6222222
```

```
hist(D$Late - D$Mid)
```



```
wilcox.test(D$Late, D$Mid, alternative="greater", paired=T)
```

```
##
## Wilcoxon signed rank test
##
## data: D$Late and D$Mid
## V = 934, p-value = 1.498e-07
## alternative hypothesis: true location shift is greater than 0
```

```
binom.test(sum((D$Late - D$Mid)>0), length(D$Late - D$Mid)) # sign test
```

```
##
## Exact binomial test
```

```
##
## data:  sum((D$Late - D$Mid) > 0) and length(D$Late - D$Mid)
## number of successes = 38, number of trials = 45, p-value =
## 3.121e-06
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.7054484 0.9350908
## sample estimates:
## probability of success
##           0.8444444
```

The Early-Mid comparison is barely significant to not significant, while the Late-Mid comparison is still significant.

Software versions

This manuscript was computed on Thu May 26 08:39:16 2016 with the following R package versions.

```
R version 3.2.4 (2016-03-10)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.11.4 (El Capitan)
```

```
locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods   base
```

```
loaded via a namespace (and not attached):
[1] magrittr_1.5      formatR_1.3      tools_3.2.4      htmltools_0.3.5
[5] yaml_2.1.13       Rcpp_0.12.4      stringi_1.0-1     rmarkdown_0.9.5
[9] knitr_1.12.3      stringr_1.0.0    digest_0.6.9     evaluate_0.8.3
```