RM-ANOVA for coevolution with a seed bank

Percent Spores

Load organized data

Adjusting variable types for lme

In the lme models used below the fixed effects need to be specified as factors. In this experiment these are: *Phage, seed.bank and time*. However for the auto correlation specification time needs to be specified as an integer. To fulfill both requirements we use the experimental day rather than transfer as the time unit, since samples taken once a day but twice per transfer. This will be simply *time x 2*. From that we make a separate variable which will be the factor of the time.

```
d <-
  d%>%
  #make time an integer for correlation structure
  # by converting to days as time unit
  mutate(Time=as.integer(2*Time))%>%
  # for the lme model all fixed effects need to be factors
  mutate(time.fct=as.factor(Time))%>%
  mutate(phage=as.factor(phage))%>%
  mutate(seed.bank=as.factor(seed.bank))%>%
  #ajust factor order for seed bank
  mutate(seed.bank = fct_rev(seed.bank))
```

Select response variable to be analyzed

```
var.response <- "%spore/ml"

d <- d %>%
  #remove non-sporulating host sample
  filter(host=="WT")%>%
  mutate(response=100*spore.ml/cell.ml )
```

Here we analyze **%spore/ml** .

```
d%>%
ggplot(aes(x=Time, y=response))+
      geom_line(aes(group=flask,color=seed.bank), size=1, alpha=0.7)+
        facet_wrap(~phage)+
      theme_bw()+
      panel_border()+
      theme(legend.position = "bottom",
            text=element_text(size=14))+
      ylab(var.response)
```

Test data for homogeneity of variances

This is an assumption of ANOVA tests.

Based on : https://www.datanovia.com/en/lessons/homogeneity-of-variance-test-in-r/ Using "Levene's test" that according to website is the most commonly used test for this purpose. This test has a null hypothesis of equal variance. So getting P>0.05 suggests homogenic variance.

```
# Levene's test with multiple independent variables
car::leveneTest(response ~ phage*line, data = d)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
## group   5  5.3671 0.0001324 ***
##       168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The data does not fulfill the assumption of equal variance across test groups.

Log transform the data

```
d <- d%>%
  mutate(log.response=log10(response))


d%>%
ggplot(aes(x=Time, y=log.response))+
      geom_line(aes(group=flask,color=seed.bank), size=1, alpha=0.7)+
        facet_wrap(~phage)+
      theme_bw()+
      panel_border()+
      theme(legend.position = "bottom",
            text=element_text(size=14))+
      ylab(paste(var.response,"(log10)"))
```

Test transformed data for homogeneity of variances

```
# Levene's test with multiple independent variables
car::leveneTest(log.response ~ phage*line, data = d)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   5  2.6536 0.0245 *
```

```
##       168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The data still does not fulfill the assumption of equal variance across test groups.

Box-Cox transformation

> `powerTransform` uses the maximum likelihood-like approach of Box and Cox (1964) to select a transformation of a univariate or multivariate response for normality, linearity and/or constant variance.
> (help page for `car::powerTransform`)

```
# Multivariate transformation to normality within levels of treatments
bx.cx <- powerTransform(response ~ phage*line, d)
summary(bx.cx)
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.9256           1       0.6962       1.1551
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                             LRT df       pval
## LR test, lambda = (0) 113.4176  1 < 2.22e-16
##
## Likelihood ratio test that no transformation is needed
##                              LRT df    pval
## LR test, lambda = (1) 0.3893232  1 0.53266
```

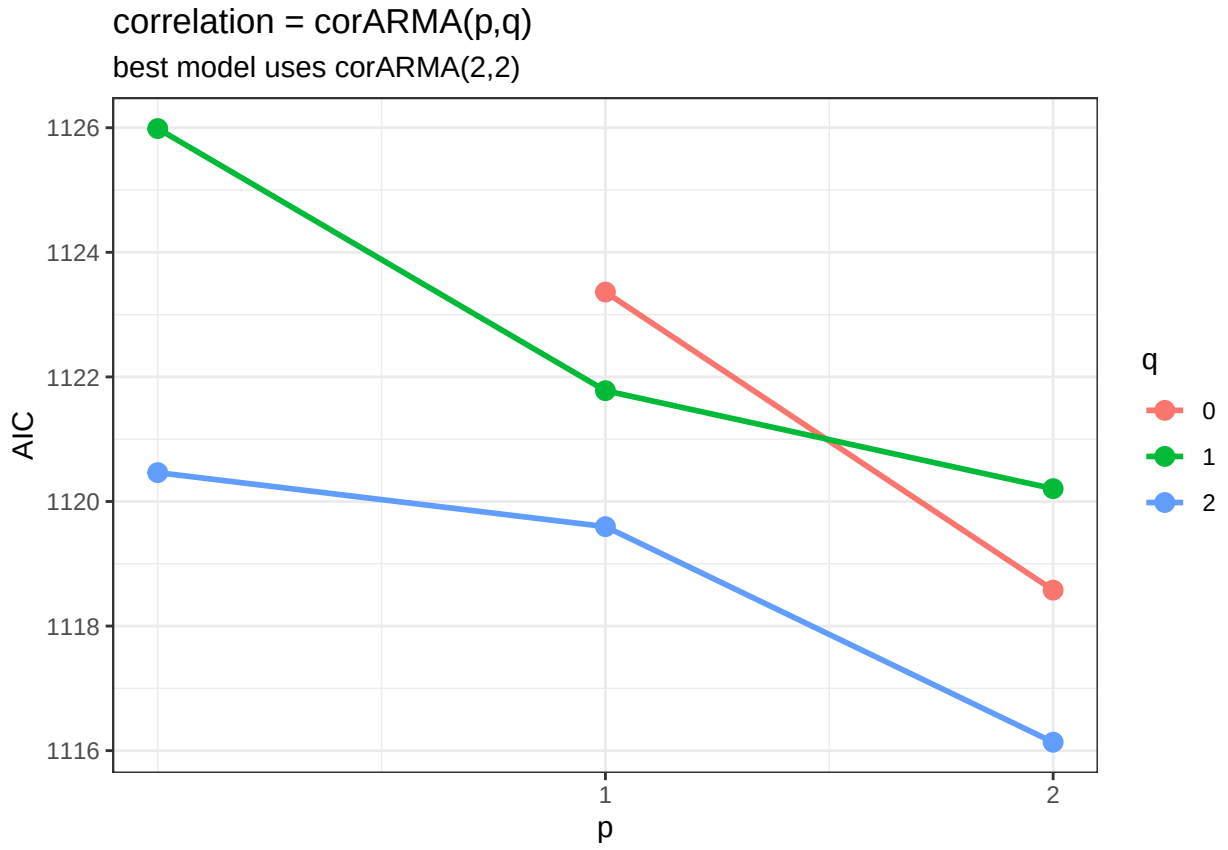
Tranformation is not required by Box-Cox test

RM-ANOVA model selection

To account for time series auto-correlation we will specify correlation structure to the lme model. We will test various corARMA correlation structures and choose by lowest AIC. This is a combined auto-regressive model, AR(p), and moving average model, MA(q). I here take a model selection approach to choose these parameters (lowest AIC). Since we have a seasonality of lag 2 due to transfer we will look at lag up to 2 for both AR and MA. Note that a corARMA(p=1, q=0) is the same as corAR1.

```
# initalise empty list to save models
l.rm <- list()

# initalise empty table to collect model data
pq.aic <- tibble()



for(Q in c(0:2)){
```

```
  for (P in c(0:2)){

    #skip corARMA(0,0)
    if (P==0 & Q==0) next

    #run model
    cur.model <- d%>%
      lme(response ~ phage * time.fct ,
          random = ~1|flask,
          correlation = corARMA(form = ~ Time | flask, p=P,q=Q),
            data = .)
    l.rm[[paste0("ARMA_P",P,"Q",Q)]] <- cur.model

    cur.tbl <- broom.mixed::glance(cur.model)%>%
      mutate(p=P)%>%
      mutate(q=Q)


      pq.aic <- bind_rows(pq.aic,cur.tbl)

  }
}

# get list order by AIC
pq.aic <-
  pq.aic%>%arrange(AIC)%>%
  # model name
  mutate(arma=paste0("ARMA_P",p,"Q",q))

# save the best model
m.best <- l.rm[[pq.aic$arma[1]]]
best <- paste0("corARMA(",pq.aic$p[1],",",pq.aic$q[1],")")

pq.aic%>%
  mutate(q=as.character(q))%>%
  ggplot(aes(p,AIC))+
  geom_line(aes(color=q),size=1)+
  geom_point(aes(color=q),size=3)+
  theme_bw()+
  scale_x_continuous(breaks = 1:10)+
  ggtitle("correlation = corARMA(p,q)",paste("best model uses", best))
```




```
# compare best 5 models by AIC order
t.models <- anova(l.rm[[pq.aic$arma[5]]],
                  l.rm[[pq.aic$arma[4]]],
                  l.rm[[pq.aic$arma[3]]],
                  l.rm[[pq.aic$arma[2]]],
                  l.rm[[pq.aic$arma[1]]])

t.models%>%
  tibble()%>%
  select(-call)%>%
  mutate(arma=pq.aic$arma[1:5])
```

```
## # A tibble: 5 x 9
##   Model    df   AIC   BIC logLik Test     L.Ratio `p-value` arma
##   <int> <dbl> <dbl> <dbl>  <dbl> <fct>      <dbl>     <dbl> <chr>
## 1     1    62 1120. 1291.  -498. ""         NA       NA     ARMA_P2Q2
## 2     2    63 1120. 1294.  -497. "1 vs 2"   2.26     0.133  ARMA_P2Q0
## 3     3    63 1120. 1293.  -497. ""         NA       NA     ARMA_P1Q2
## 4     4    62 1119. 1289.  -497. "3 vs 4"   0.981    0.322  ARMA_P2Q1
## 5     5    64 1116. 1292.  -494. "4 vs 5"   6.44     0.0399 ARMA_P0Q2
```

best model is corARMA(2,2)

Though it is not significantly better than mostof the other models.

Results of selected model

```
# best model data
pq.aic%>%
  slice_min(AIC)
```

```
## # A tibble: 1 x 7
##   sigma logLik   AIC   BIC     p     q arma
##   <dbl>  <dbl> <dbl> <dbl> <int> <int> <chr>
## 1  14.1  -494. 1116. 1292.     2     2 ARMA_P2Q2
```

```
#display best model results
  anova(m.best)
```

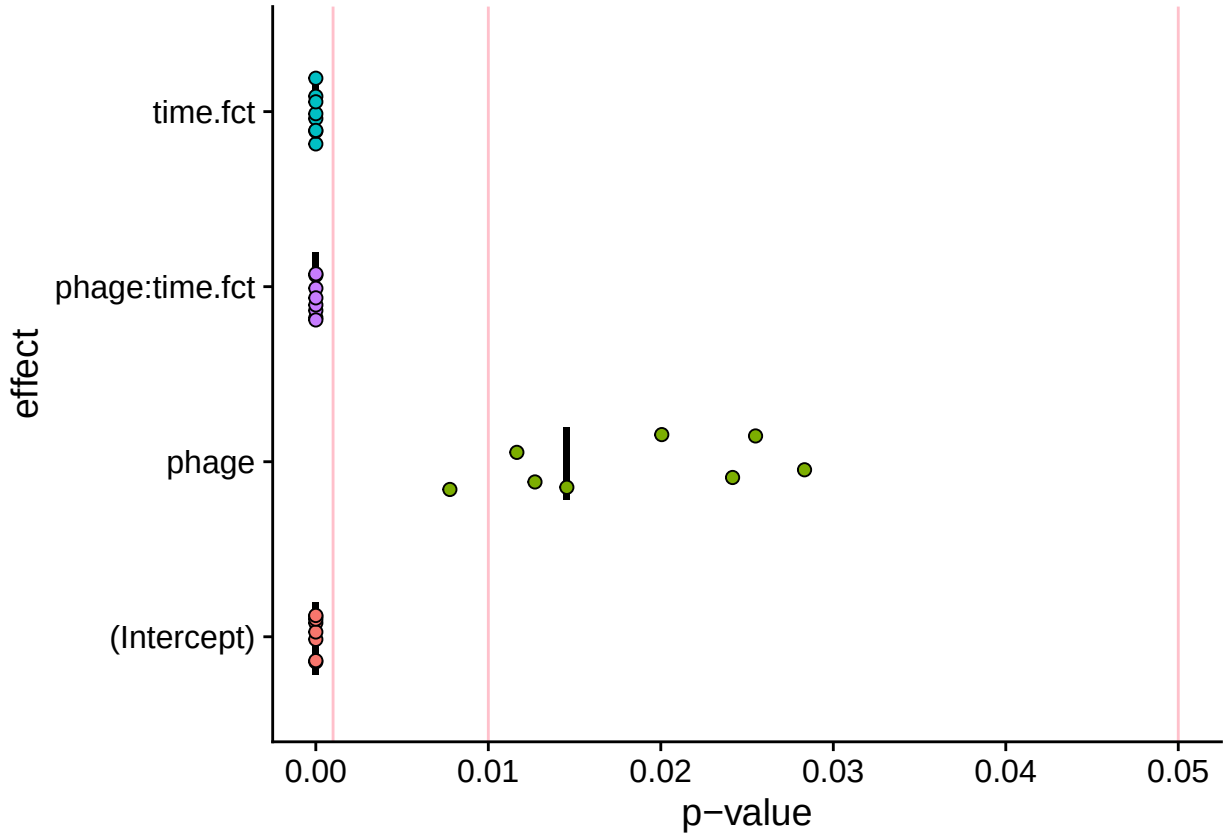
```
##                numDF denDF   F-value p-value
## (Intercept)        1   112 1418.8745  <.0001
## phage              1     4   17.0163  0.0146
## time.fct          28   112   11.5187  <.0001
## phage:time.fct    28   112    3.7704  <.0001
```

Phage and time both significantly effect spore fraction, as main effects and in their interaction. Before looking into the potential drivers of these effects we evaluate the model.

How sensitive would the result be to model selected?

looking at different correlation structures.

```
# all model results to tibble
all.models <-
  map(l.rm, anova.lme) %>%
  map(rownames_to_column)%>%
  bind_rows(.id = "arma")%>%
  rename(effect=rowname)


arma.best <- paste0("ARMA_P",pq.aic$p[1],"Q",pq.aic$q[1])
selected <-
  all.models%>%
  filter(arma==arma.best)


all.models%>%
  mutate(effect=fct_inorder(effect))%>%
  ggplot(aes(x=effect, y=`p-value`))+
  geom_point(data = selected, color="black", size=10, shape="|")+
  geom_hline(yintercept = c(0.05,0.01,1e-3), color="pink")+
  geom_jitter(aes(fill=effect),width = 0.2, height = 0,
              shape=21, size=2, show.legend = F)+
  coord_flip()+
  theme_cowplot()
```

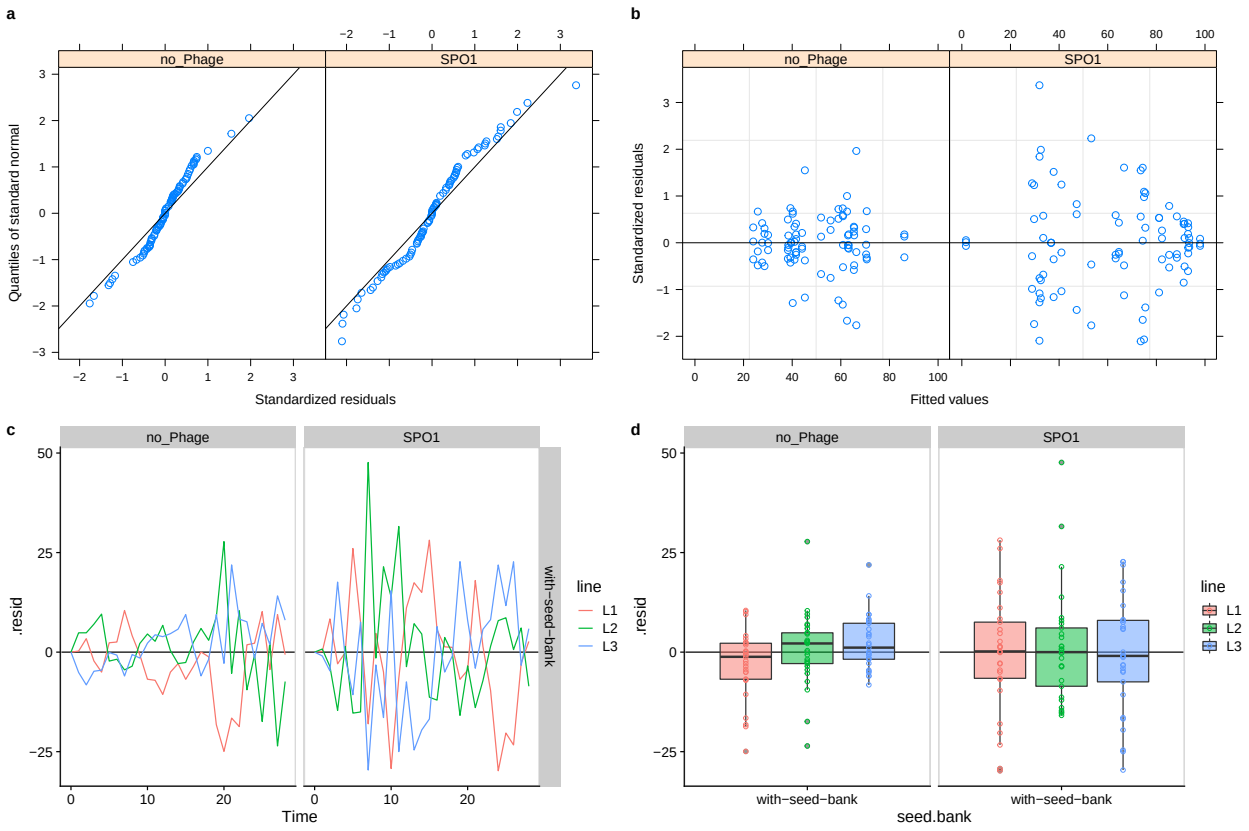


The line marks the p-values of the selected model, and the points show p-values of all models tested (including best). All the models have the same results, even though the exact significance of the phage effect varies.

Distribution of model residuals For the best model.

```
p1 <-
  #qqplot by phage
  qqnorm(m.best,~ resid(., type = "p")|phage, abline = c(0, 1))

p2 <-
  # standardized residuals versus fitted values by phage
plot(m.best, resid(., type = "p") ~ fitted(.) | phage, abline = 0)

p3 <-
broom.mixed::augment(m.best)%>%
  ggplot(aes(Time,.resid)) +
  geom_hline(yintercept = 0)+
  geom_line(aes(color=line))+
  facet_grid(seed.bank~phage)+
  theme_cowplot()+panel_border()

p4 <-
broom.mixed::augment(m.best)%>%
  ggplot(aes(seed.bank,.resid)) +
  geom_hline(yintercept = 0)+
```

```
  geom_boxplot(aes(fill=line),alpha=.5, position = position_dodge(width = .9))+
  geom_point(aes(color=line), position = position_dodge(width = .9), shape=21)+
  facet_wrap(~phage)+
  theme_cowplot()+panel_border()



plot_grid(p1,p2,p3,p4, nrow = 2, labels = 'auto')
```



conclusions:
a. model residuals are very close to normal distribution. Sign of good fit.
b. The residuals are evenly distributed around 0, suggesting equal variance.
c. Equal residual variance holds across time.
d. Equal residual variance holds across experimental units (flasks).

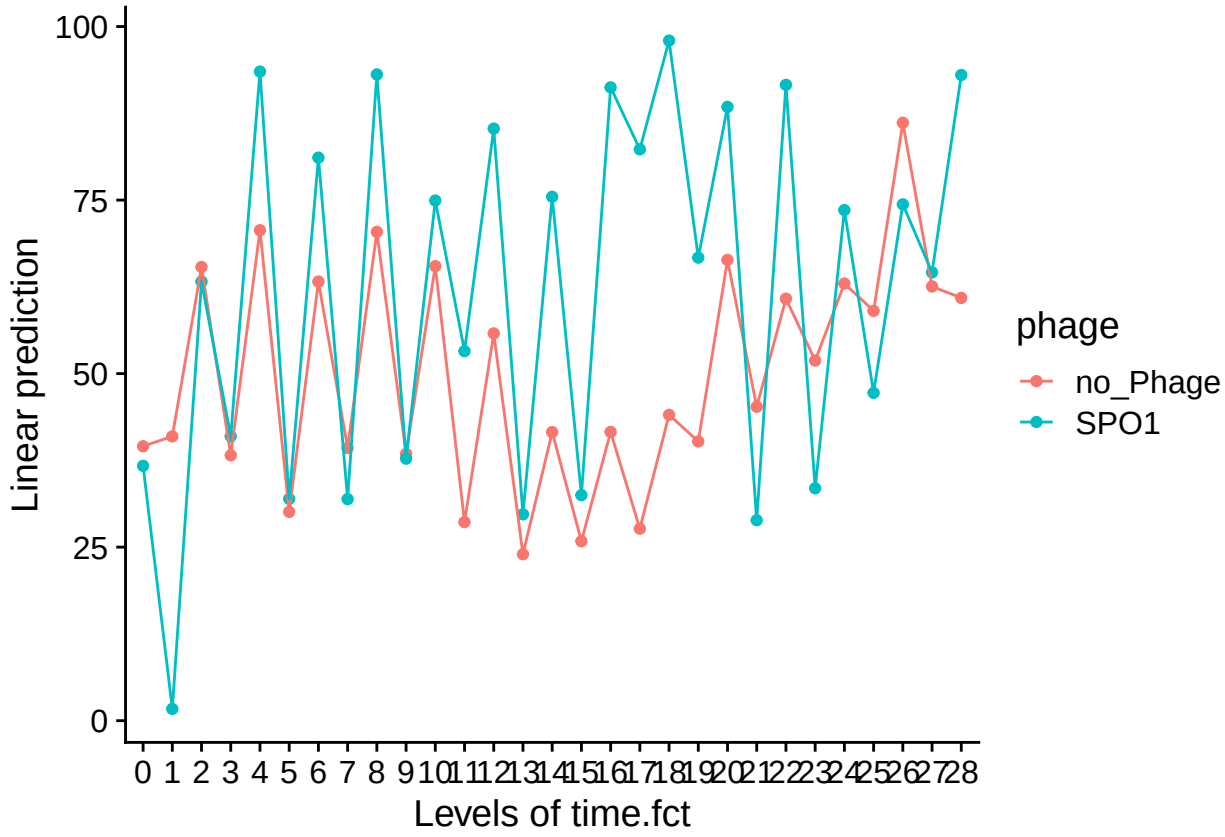
post hoc

What in the phage X time interaction is causing a significant effect on spore fraction dynamics?
We analyze the *estimated marginal means (EMM)*, following examples from: https://cran.r-project.org/web/packages/emmeans/vignettes/interactions.html

Visualize EMM of interactions

```
# plot
emmip(m.best, phage~time.fct)+
  theme_cowplot()
```

```
## Warning in sweep(X, 1, sqrt(weights), "*"): STATS is longer than the extent of
## 'dim(x)[MARGIN]'
```



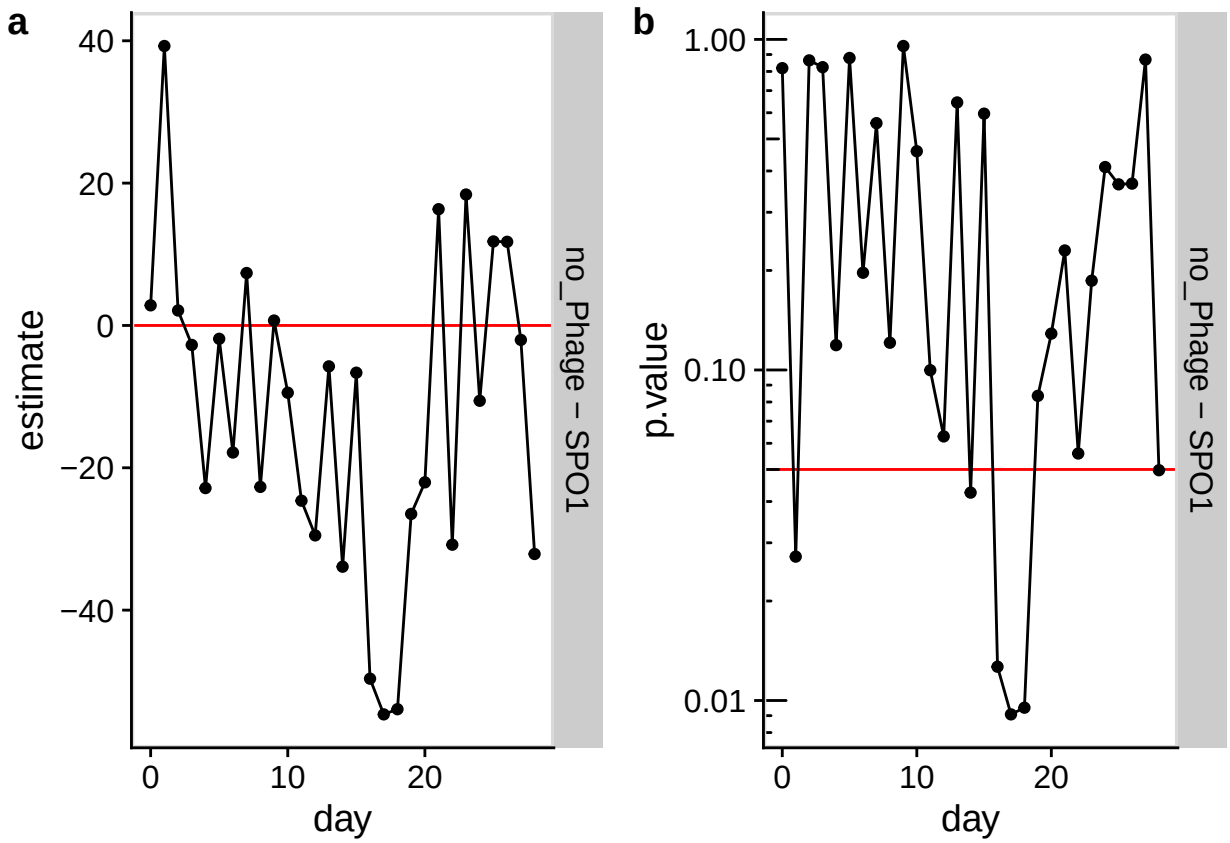
An increase in percent of spores is seen for a short period in the phage infected cultures, an not in the absence of phage.

```
coevo.emm <- emmeans(m.best, ~ phage * time.fct)
```

```
## Warning in sweep(X, 1, sqrt(weights), "*"): STATS is longer than the extent of
## 'dim(x)[MARGIN]'
```

```
coevo.pairs <- pairs(regrid(coevo.emm), simple="phage")%>%
  tidy

p1 <-
  coevo.pairs%>%
  # focus on phage infected
  # filter(phage=="SPO1")%>%
  # make time continuous for plotting.
  mutate(day=as.numeric(time.fct))%>%
```

```
  # # arrange panel order
  # mutate(contrast=fct_relevel(contrast, "long - short", after = 0))%>%
  #plot
  ggplot(aes(x=day, y=estimate))+
    #add 0 line
  geom_hline(yintercept = 0, color="red")+
  geom_point()+
  geom_line()+
  facet_grid(contrast~.)+
  theme_cowplot()+
  panel_border()

p2 <- coevo.pairs%>%
  # focus on phage infected
  # filter(phage=="SPO1")%>%
  # make time continuous for plotting.
  mutate(day=as.numeric(time.fct))%>%

  #plot
  ggplot(aes(x=day, y=p.value))+
    #add 0.05 signifcance thrshold
  geom_hline(yintercept = 0.05, color="red")+
  geom_point()+
  geom_line()+
  facet_grid(contrast~.)+
  theme_cowplot()+
  panel_border()+
  scale_y_log10()+
  annotation_logticks(sides = "l")



plot_grid(p1,p2, labels = "auto")
```



For a few days (days 16-18) the phage infected cultures have significantly greater fraction of spores.

Summary

??