RM-ANOVA for coevolution with a seed bank

Phage density

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.

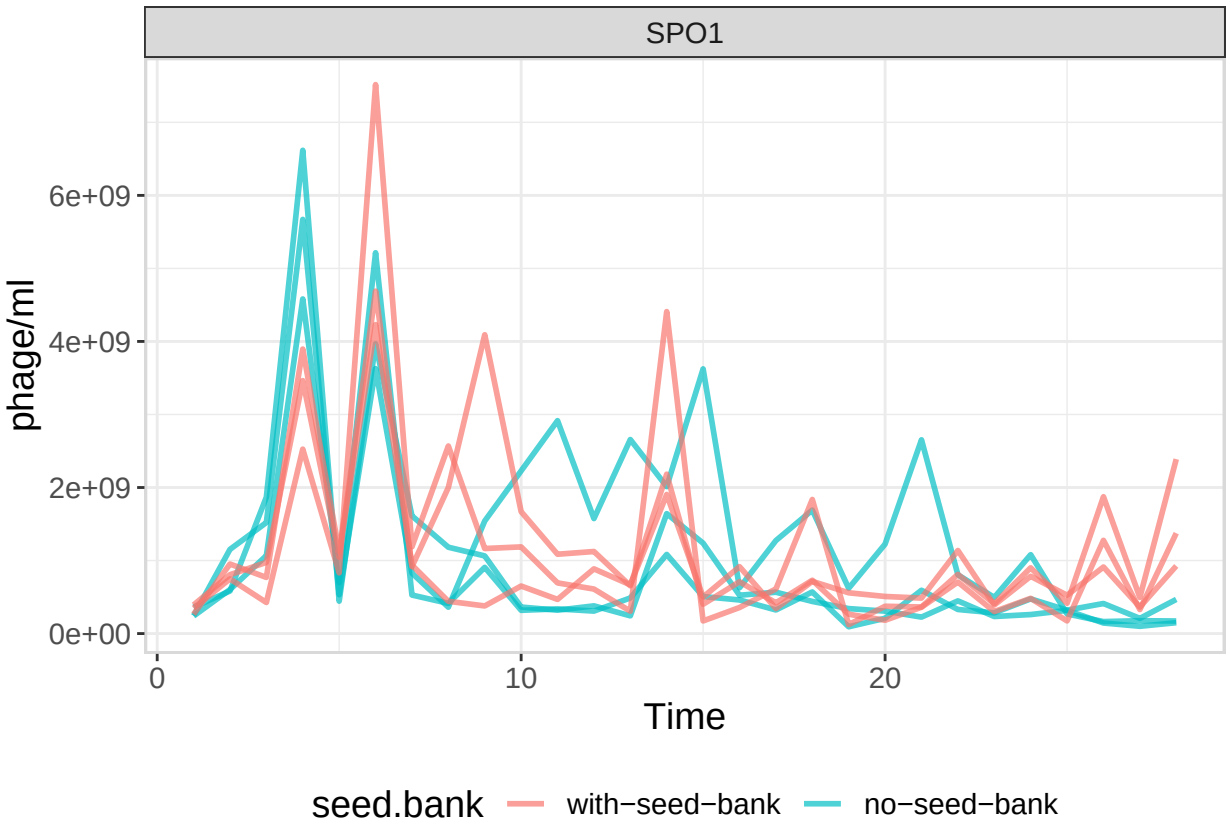
```
d <-
  d%>%
  #make time an integer for correlation structure
  # by converting to days as time unit
  mutate(Time=as.integer(2*Time))%>%
  # for the lme model all fixed effects need to be factors
  mutate(time.fct=as.factor(Time))%>%
  mutate(phage=as.factor(phage))%>%
  mutate(seed.bank=as.factor(seed.bank))%>%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

Select response variable to be analyzed

```
var.response <- "phage/ml"

d <- d %>%
  mutate(response=phage.ml )%>%
  # remove no phage data
  filter(phage=="SPO1")%>%
  # remove time of infection (T=0) the change in that first period is an outlier to the data
  filter(Time>0)
```

Here we analyze **phage/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 ~ seed.bank*line, data = d)
```

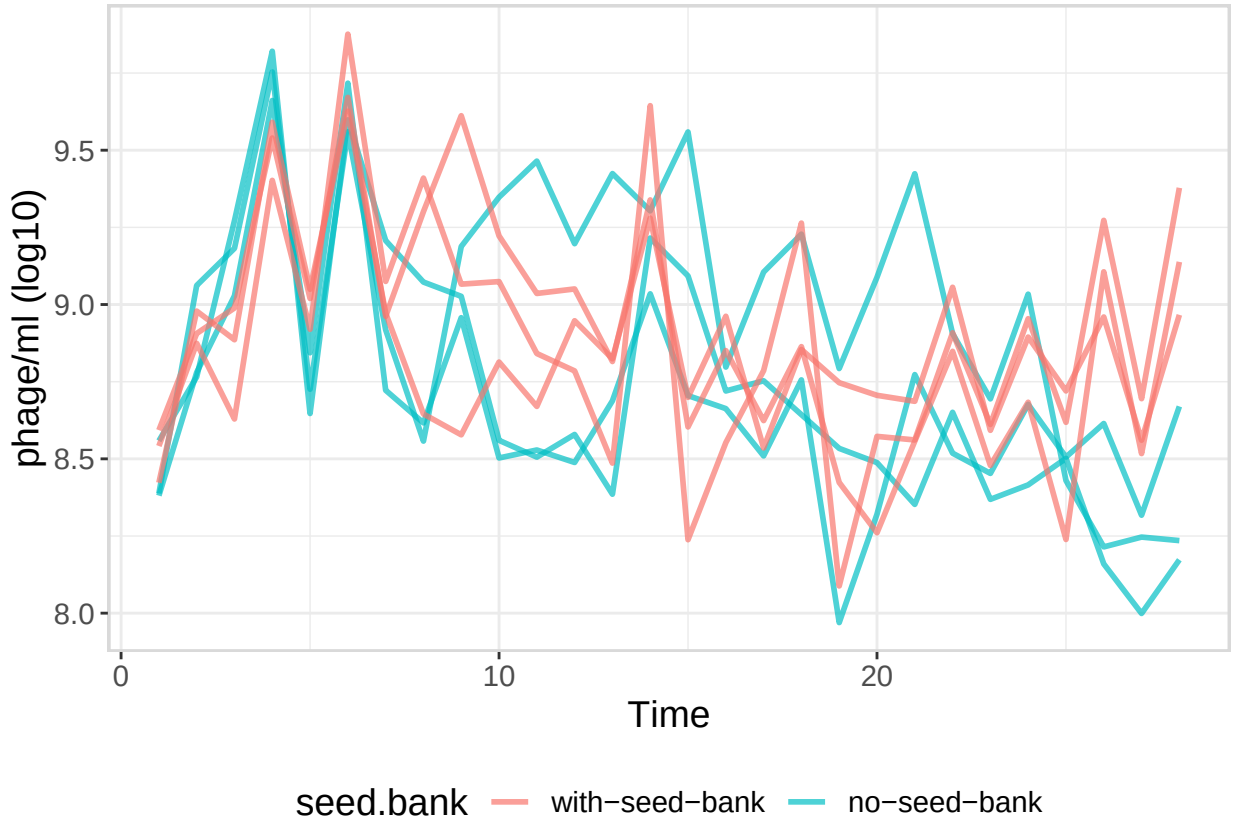
```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   5  0.8425 0.5214
##       162
```

The data fulfills the assumption of equal variance across test groups, with no need for transformation. I will log transform just for plotting:

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 ~ seed.bank*line, data = d)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   5  1.3612 0.2416
##       162
```

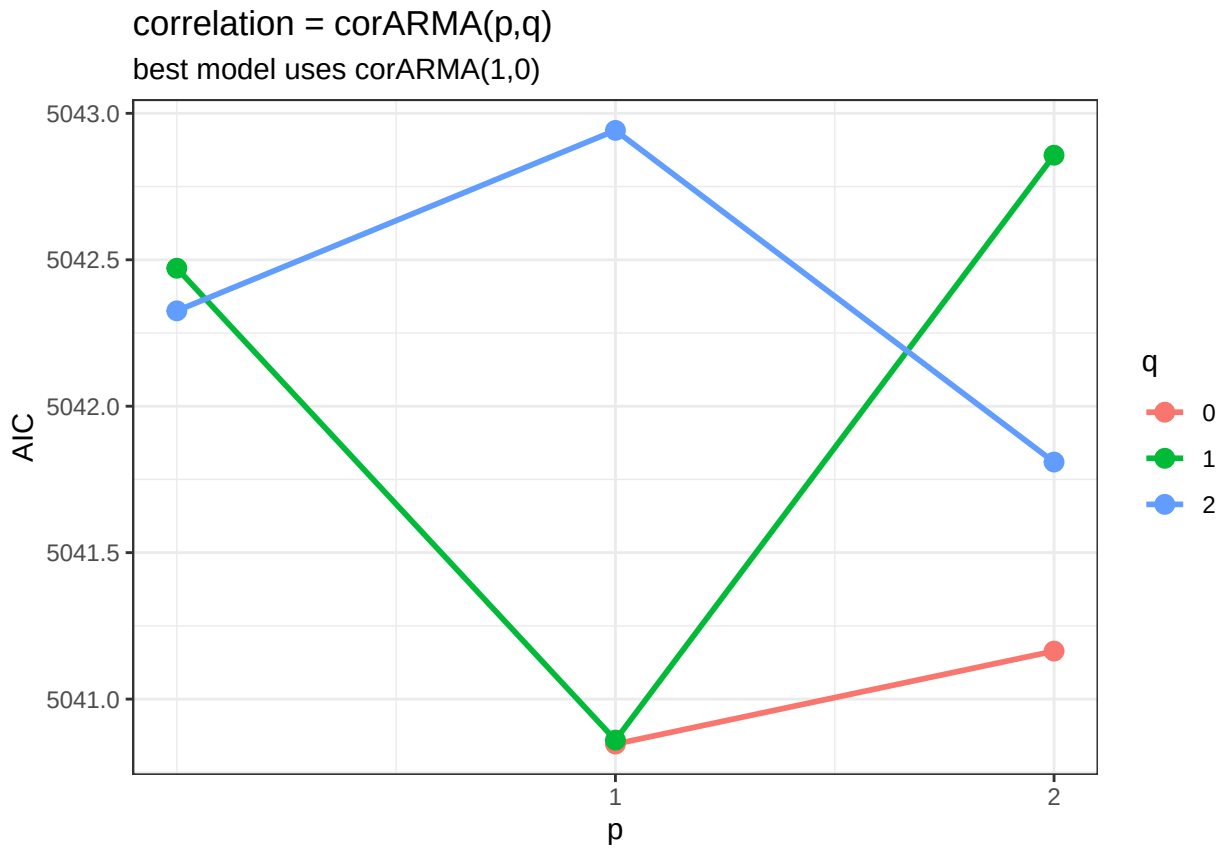
Using un-transformed response data fro analyse below

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 ~  seed.bank * 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    60 5042. 5205. -2461. ""          NA       NA     ARMA_P1Q0
## 2     2    62 5042. 5210. -2459. "1 vs 2"    4.52     0.105 ARMA_P1Q1
## 3     3    60 5041. 5204. -2461. "2 vs 3"    3.35     0.187 ARMA_P2Q0
## 4     4    60 5041. 5204. -2460. ""          NA       NA     ARMA_P2Q2
## 5     5    59 5041. 5201. -2461. "4 vs 5"    1.99     0.159 ARMA_P0Q2
```

best model is corARMA(1,0)

Though it is not significantly better than 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 652798216. -2461. 5041. 5201.     1     0 ARMA_P1Q0
```

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

```
##                     numDF denDF  F-value p-value
## (Intercept)             1   108 88.94591  <.0001
## seed.bank               1     4  0.11708  0.7494
## time.fct               27   108 18.93735  <.0001
## seed.bank:time.fct     27   108  2.57313  0.0003
```

Time is very significant as main effect and also in the interaction seed-bank x time.

Time as main effect - In the plotted data it appears that phage densities are declining over time.

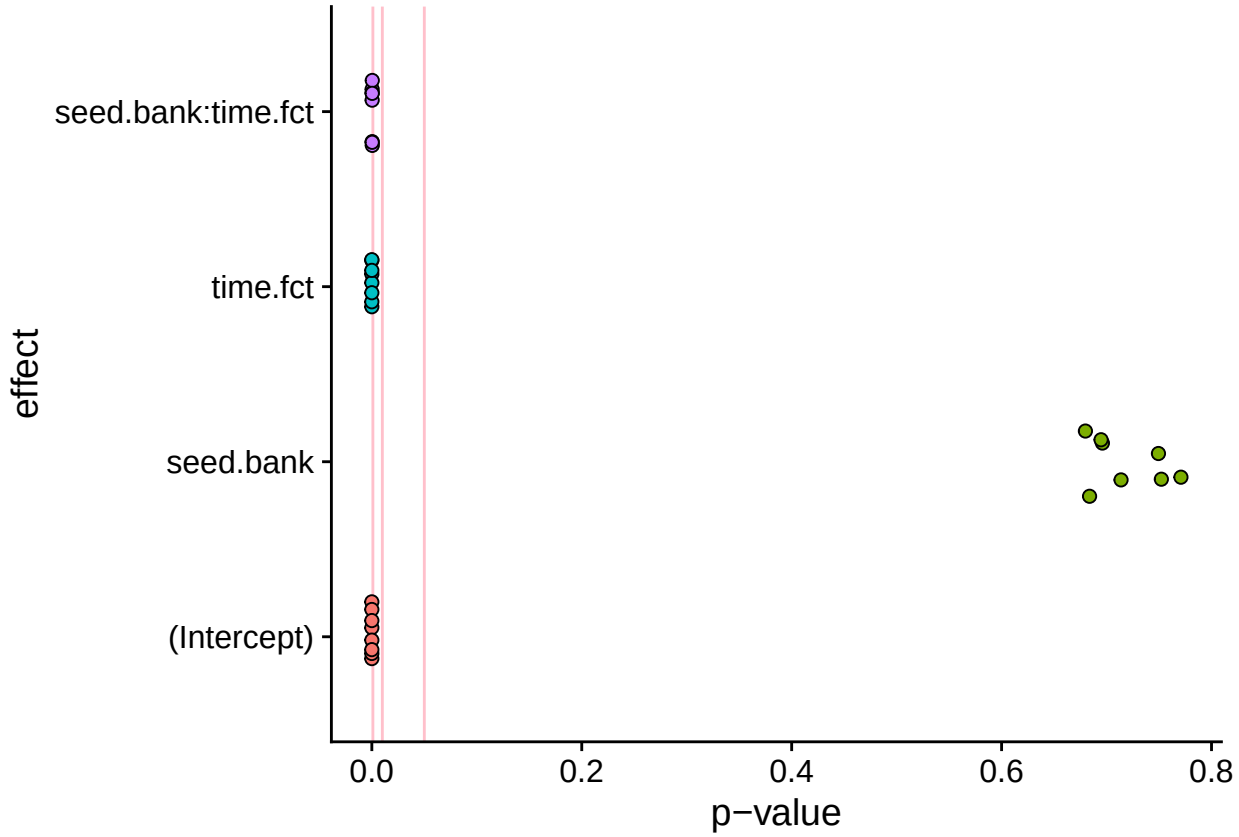
seed-bank x time - There is some separation between seed-bank treatments,at the very end of the experiment. It looks like phage numbers are dropping in the no-seed-bank lines.

Before looking into the potential drivers of this effect 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)


all.models%>%
  mutate(effect=fct_inorder(effect))%>%
  ggplot(aes(x=effect, y=`p-value`))+
  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()
```

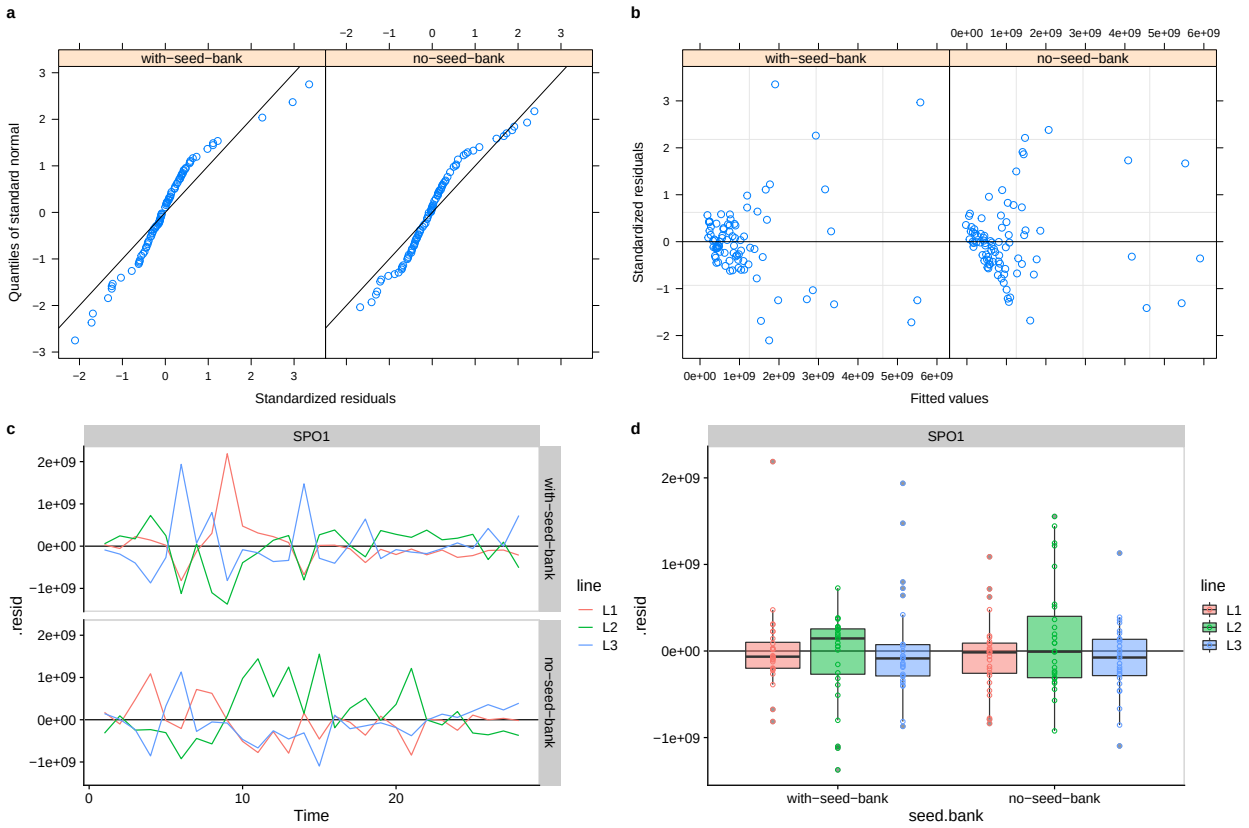


The results are pretty similar across the models. For the **seed.bank** the effect remains non-significant in all models having a P>0.05 for that main effect

Distribution of model residuals For the best model.

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

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

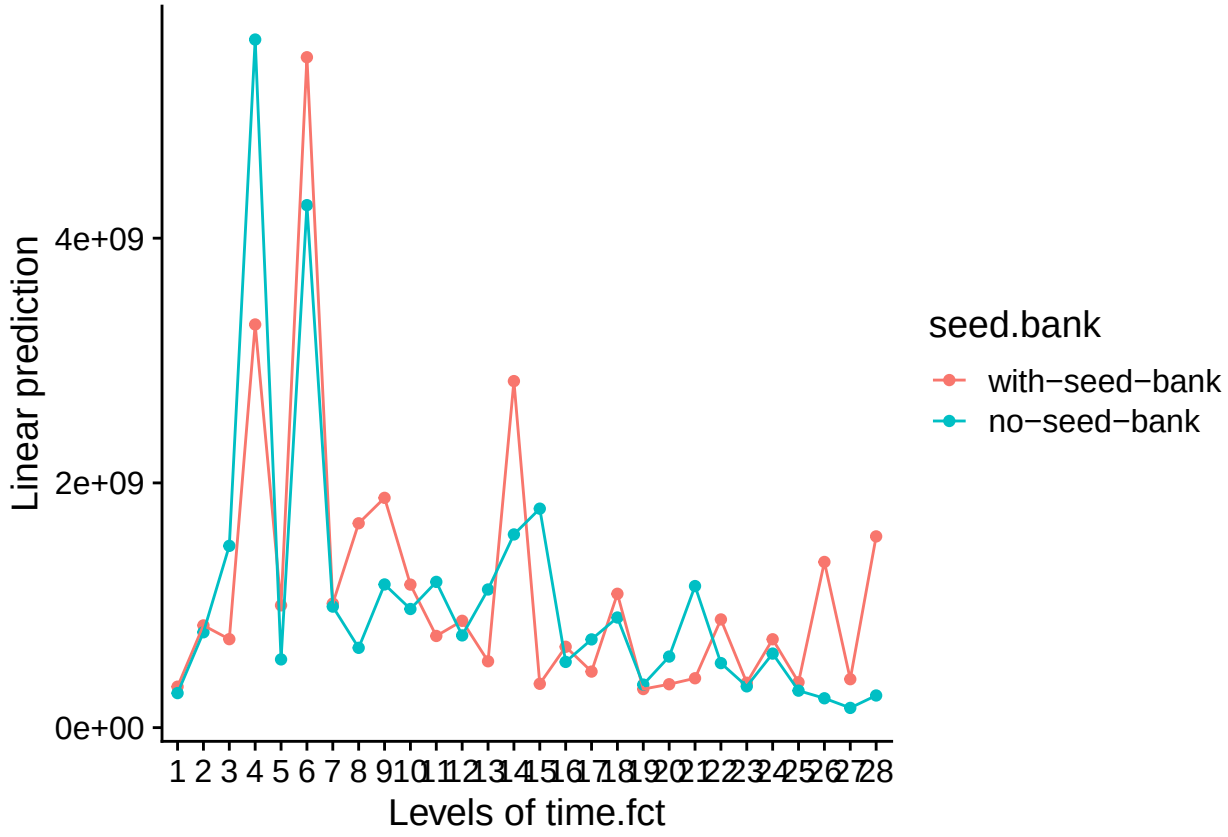
post hoc

What in the seed-bank X time interaction is causing a significant effect on phage population 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, seed.bank~time.fct)+
  theme_cowplot()
```

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



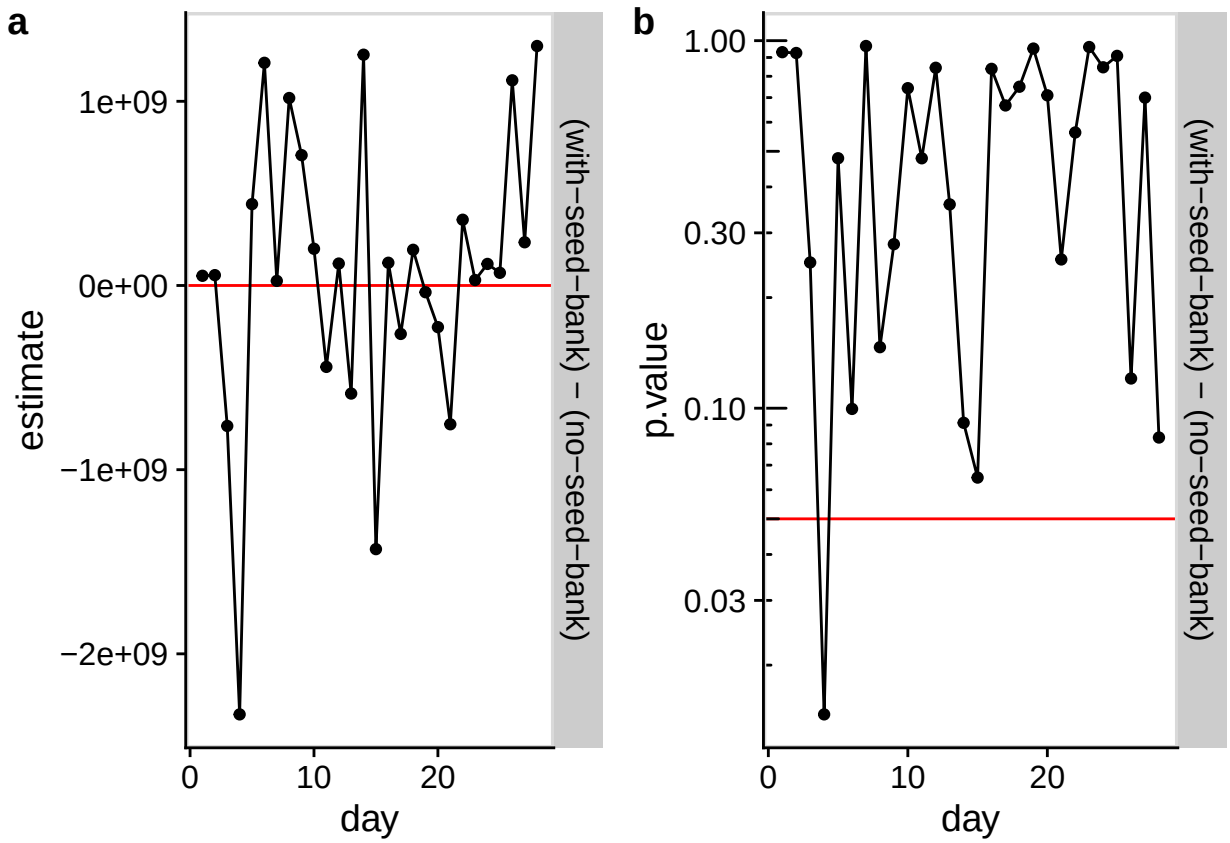
As noted above, it seems that in the later time points the with-seed-bank treatment has higher phage densities. However this is not eough of difference for a seed-bank main effect. How do seed bank treatments differ over time?

```
coevo.emm <- emmeans(m.best, ~ seed.bank * 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="seed.bank")%>%
  tidy

p1 <-
  coevo.pairs%>%
  # 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%>%

  # 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")
```



The difference between the two seed bank treatments **is NOT** significant in the last three samples, as was expected. The interaction's source of significance is in day 4 where we observe an elevtion of phage densities in the absence of seed banks.

Time as a main effect

```
emm_t <- emmeans(m.best, ~ time.fct)
```

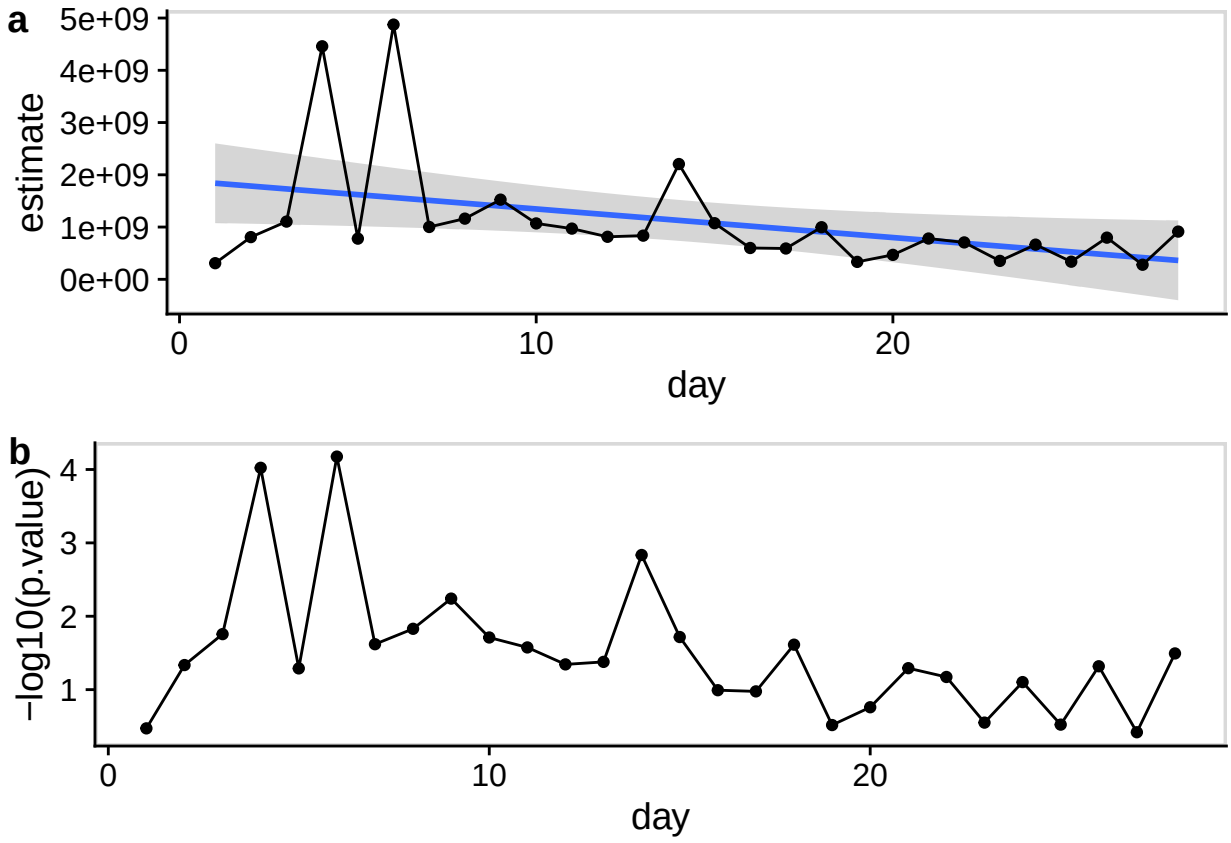
```
## Warning in sweep(X, 1, sqrt(weights), "*"): STATS is longer than the extent of
## 'dim(x)[MARGIN]'
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
# emm_t%>%
#   tidy()

p1 <-
  tidy(emm_t) %>%
  # make time continuous for plotting.
  mutate(day=as.numeric(time.fct))%>%
  #plot
  ggplot(aes(x=day, y=estimate))+
  geom_smooth(method = "lm")+
  geom_point()+
```

```
  geom_line()+
  theme_cowplot()+
  panel_border()

p2 <-
  tidy(emm_t) %>%
  # make time continuous for plotting.
  mutate(day=as.numeric(time.fct))%>%
  #plot
  ggplot(aes(x=day, y=-log10(p.value)))+
  geom_point()+
  geom_line()+
  theme_cowplot()+
  panel_border()



plot_grid(p1,p2,nrow = 2, labels = "auto")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



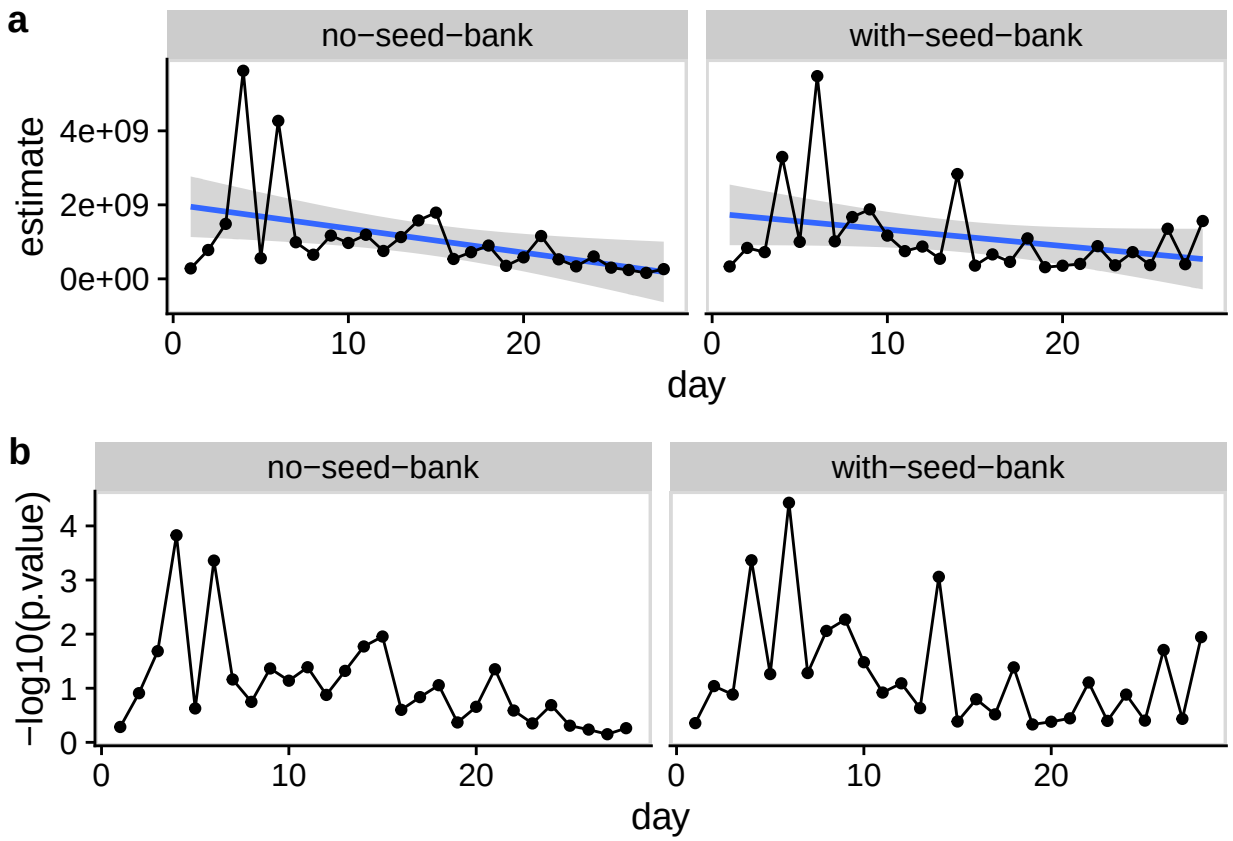
Confirming that there is a trend of decline in phage density over time.

```
phage <- tidy(emm_t) %>% pull(estimate)
Time <-  tidy(emm_t) %>% pull(time.fct) %>% as.numeric()
cor.test(Time, phage)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Time and phage
## t = -2.3236, df = 26, p-value = 0.02823
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.68220981 -0.04921071
## sample estimates:
##        cor
## -0.4146745
```

```
p1 <-
  tidy(coevo.emm) %>%
  # make time continuous for plotting.
  mutate(day=as.numeric(time.fct))%>%
  #plot
  ggplot(aes(x=day, y=estimate))+
  geom_smooth(method = "lm")+
  geom_point()+
  geom_line()+
  facet_grid(~seed.bank)+
  theme_cowplot()+
  panel_border()

p2 <-
  tidy(coevo.emm) %>%
  # make time continuous for plotting.
  mutate(day=as.numeric(time.fct))%>%
  #plot
  ggplot(aes(x=day, y=-log10(p.value)))+
  geom_point()+
  geom_line()+
  facet_grid(~seed.bank)+
  theme_cowplot()+
  panel_border()



plot_grid(p1,p2,nrow = 2, labels = "auto")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Summary

The effect of seed bank treatment on phage population dynamics is very small, with a significant differences between seed-bank treatments observed only at the a single time point, on day 4. This is the same day where the difference in host total cells starts to take effect.

Overall, phage densities decline with time.