RM-ANOVA for coevolution with a seed bank

Total cell 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 <- "cell/ml"

d <- d %>%
  mutate(response=cell.ml )
```

Here we analyze **cell/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.

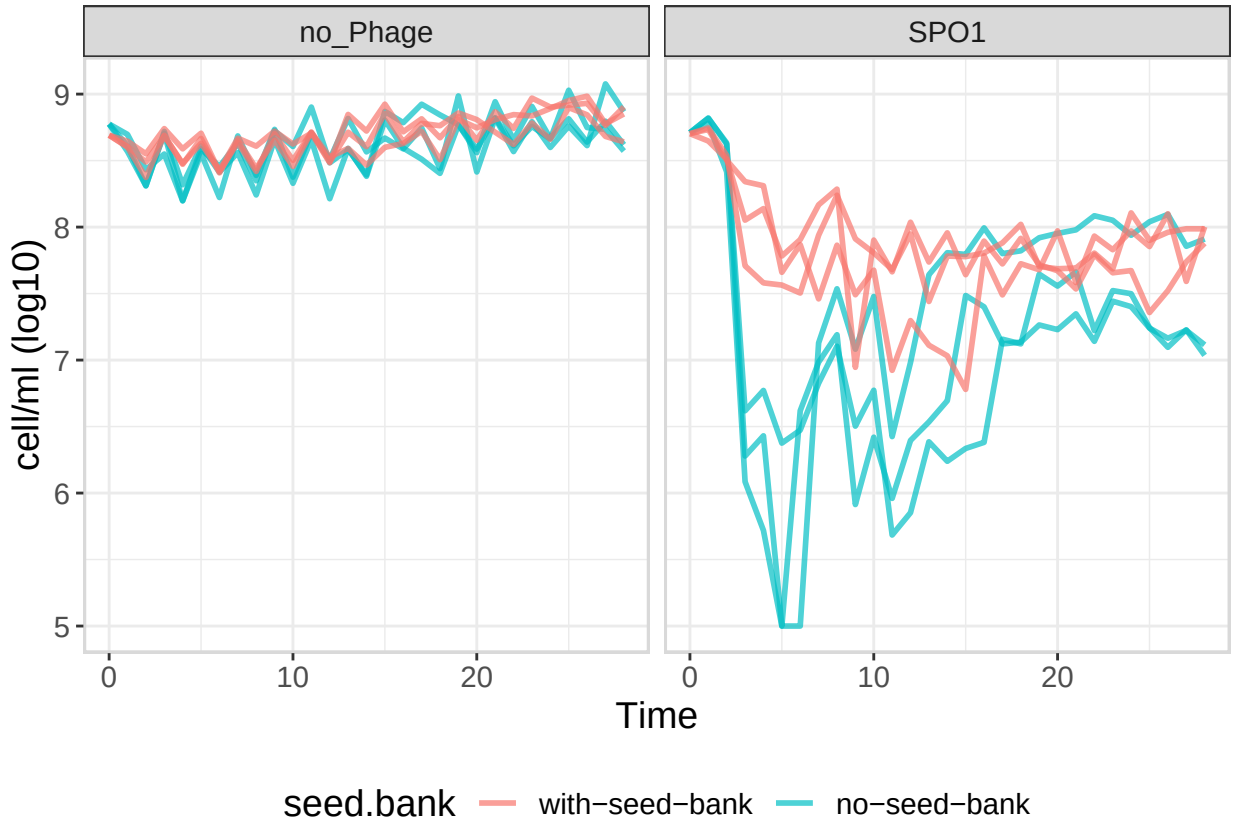
```
# Levene's test with multiple independent variables
car::leveneTest(response ~ phage*seed.bank*line, data = d)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
## group  11  3.4101 0.0001672 ***
##       336
## ---
## 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

```r
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

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

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value   Pr(>F)
## group  11   8.939 5.44e-14 ***
```

```
##       336
## ---
## 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`)

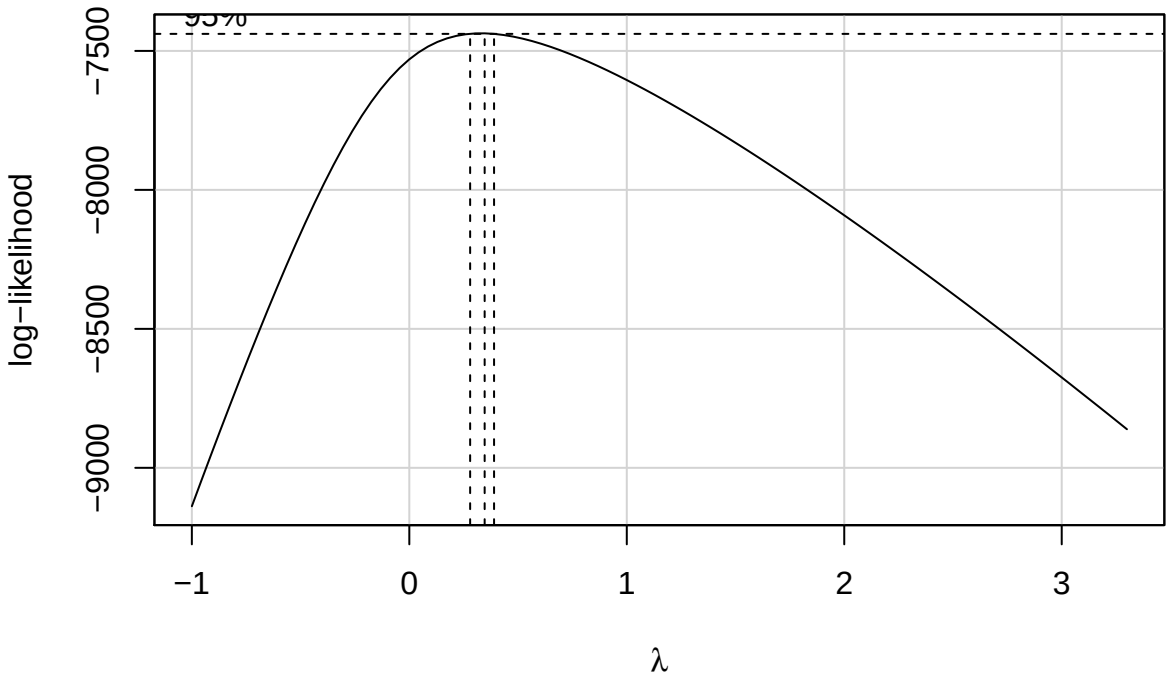
```
# Multivariate transformation to normality within levels of treatments
bx.cx <- powerTransform(response ~ phage*seed.bank*line, d)
summary(bx.cx)
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.3309        0.33       0.2757       0.3861
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df       pval
## LR test, lambda = (0) 187.3946  1 < 2.22e-16
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 336.2832  1 < 2.22e-16
```

Tranformation is required, but not a simple log transformation

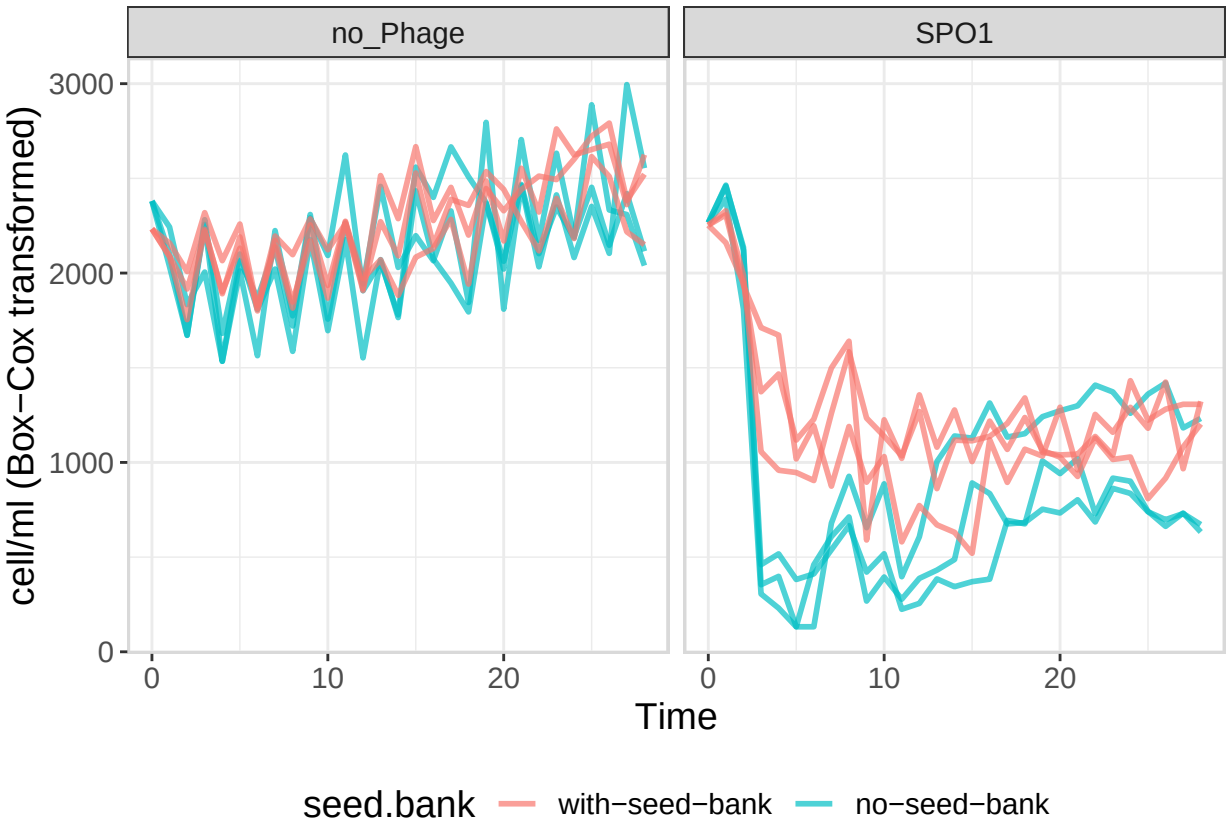
Plot the profile log-likelihood for Box-Cox transformations.

```
  with(d, boxCox(cell.ml ~ phage*seed.bank*line,
                 lambda = seq(-1, bx.cx$roundlam+3, by = 0.1),
                 family="bcPower"))
```



Transform using Box-Cox λ (rounded).

```
d <- d%>%
  mutate(bxcx.response=bcPower(cell.ml, bx.cx$roundlam))

d%>%
ggplot(aes(x=Time, y=bxcx.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,"(Box-Cox transformed)"))
```



Test transformed data for homogeneity of variances

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

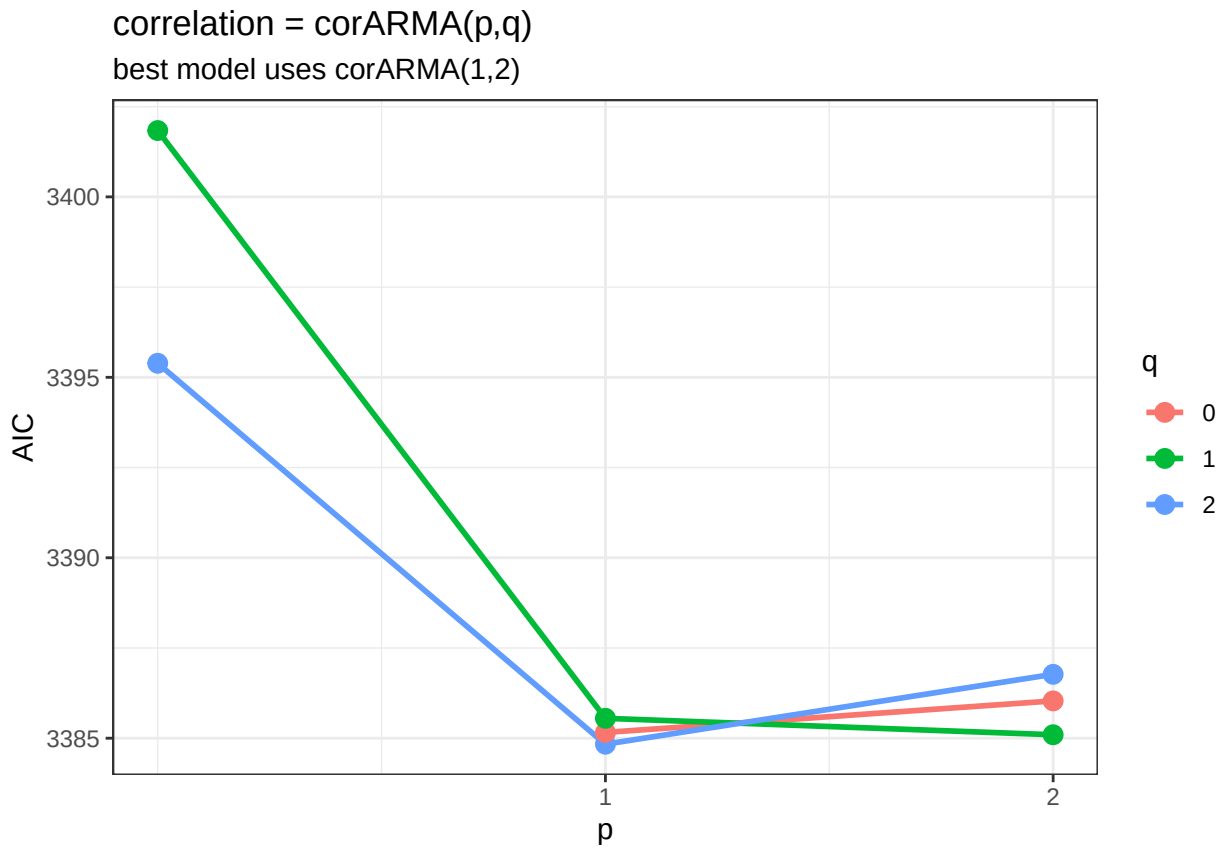
```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group  11  1.5243 0.1209
##       336
```

The data now fulfills the assumption of equal variance across test groups.

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.

```r
# 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(bxcx.response ~ phage * 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   120 3386. 3800. -1573. ""           NA        NA     ARMA_P1Q2
## 2     2   120 3386. 3799. -1573. ""           NA        NA     ARMA_P2Q1
## 3     3   119 3385. 3795. -1574. "2 vs 3"    1.61     0.205 ARMA_P1Q0
## 4     4   121 3385. 3802. -1572. "3 vs 4"    4.06     0.131 ARMA_P1Q1
## 5     5   121 3385. 3802. -1571. ""           NA        NA     ARMA_P2Q0
```

best model is corARMA(1,2)

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  207. -1571. 3385. 3802.     1     2 ARMA_P1Q2
```

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

```
##                            numDF denDF  F-value p-value
## (Intercept)                    1   224 3235.479  <.0001
## phage                          1     8  296.080  <.0001
## seed.bank                      1     8   11.556  0.0094
## time.fct                      28   224   15.165  <.0001
## phage:seed.bank                1     8    3.510  0.0979
## phage:time.fct                28   224   24.501  <.0001
## seed.bank:time.fct            28   224    4.553  <.0001
## phage:seed.bank:time.fct      28   224    2.193  0.0009
```

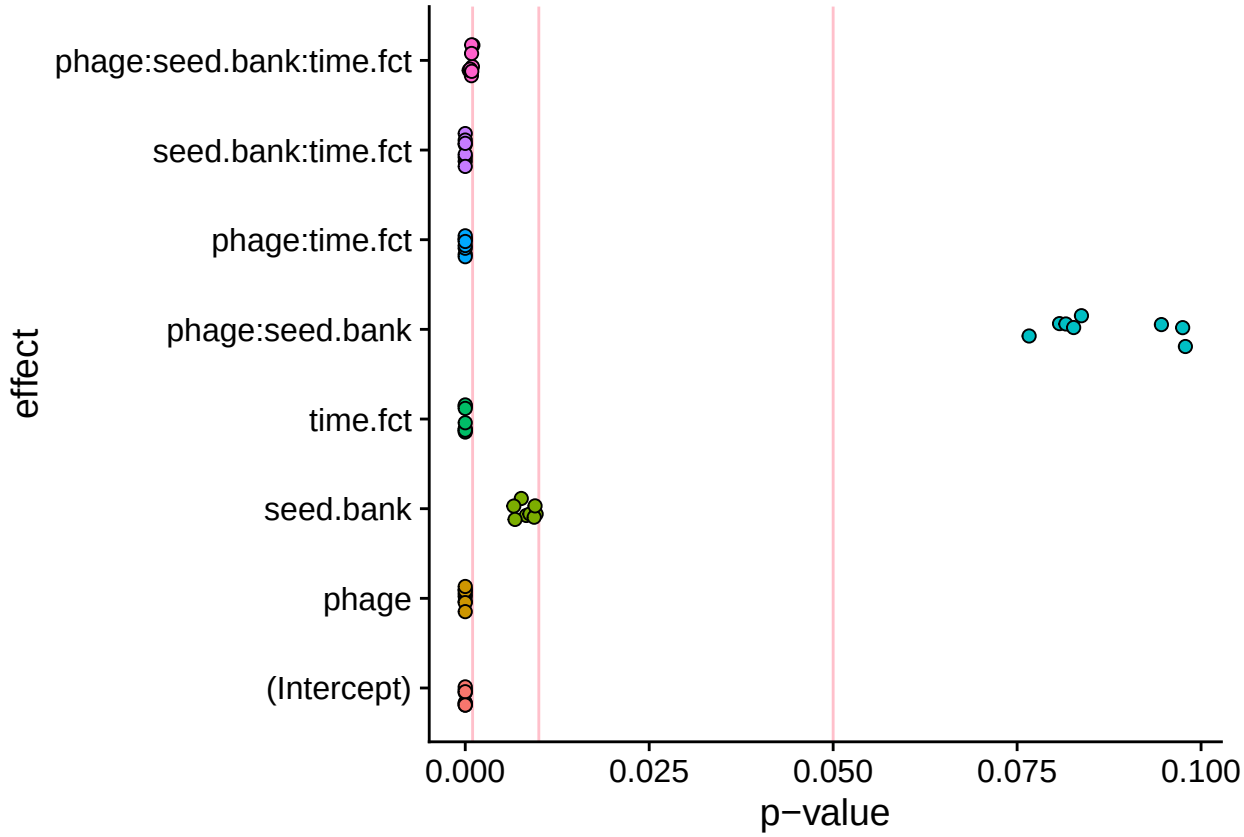
As expected time is very significant at all levels. This includes also a significant 3-way interaction phage x seed-bank x time is detected.
Phage too is strongly significant as a main effect and in interaction with time. Phage X seed-bank is on border of significance.
Seed-bank is significant as main effect

There is evidence to suggest that the seed bank treatment is influencing population dynamics in a phage dependent manner. 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()
```

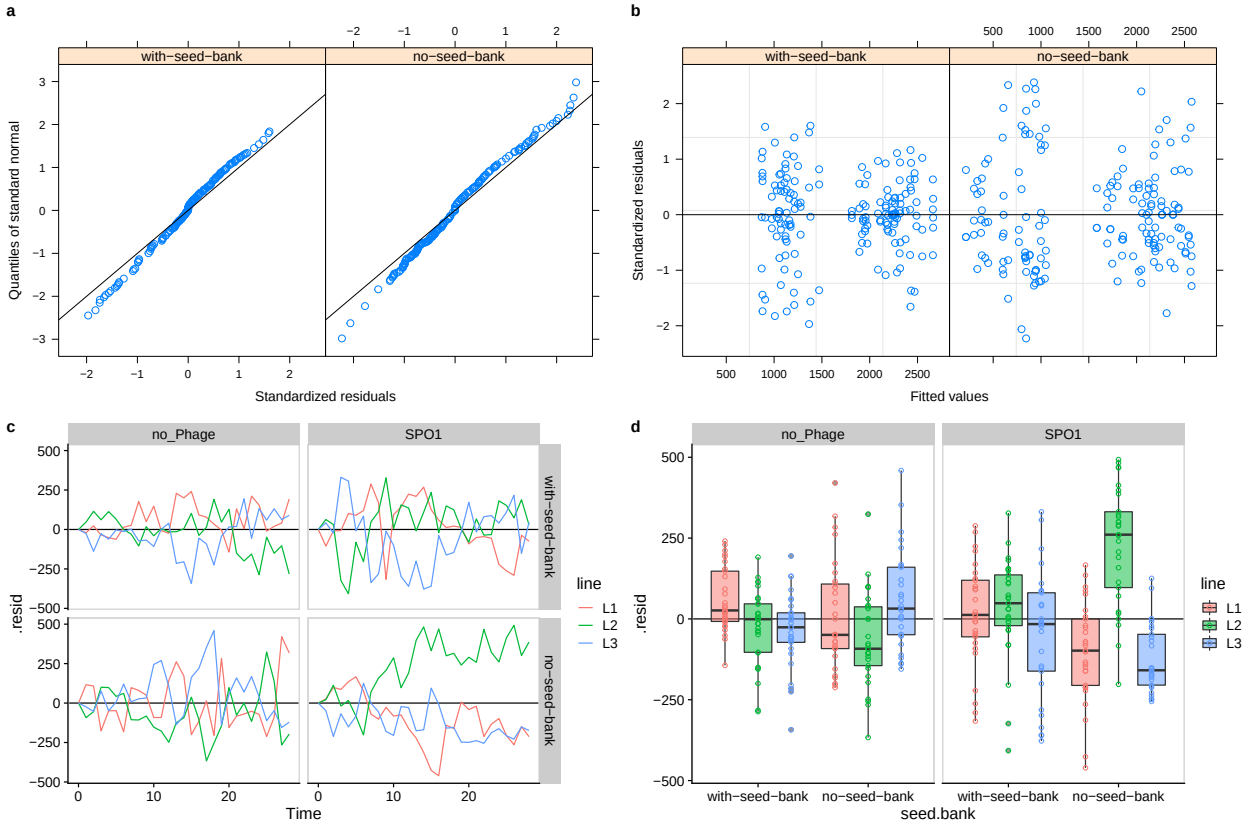


The results are pretty similar across the models. For the **phageXseed.bank** interaction the effect remains non-significant in all models having a P>0.05 for that interaction.

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 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).

A single clear deviation is the no-seed-bank L2 flask.

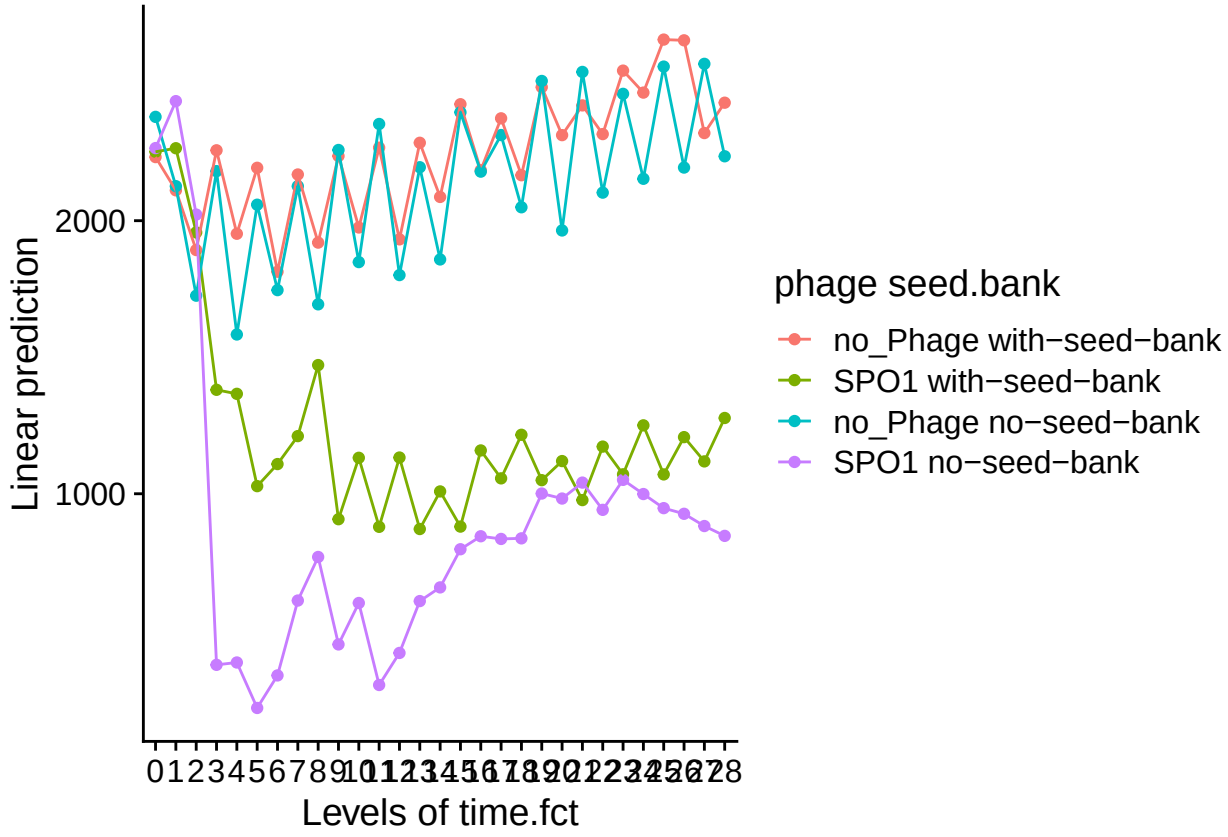
post hoc

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

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



Plot shows that the seed-bank treatment in the presence of phage is likely the treatment interaction standing out. We compare seed banks treatments separately for phage treatments.

```
emm_sb.ph <-
  stats::update(ref_grid(m.best), tran = make.tran("boxcox", param =  bx.cx$roundlam)) %>%
  emmeans(pairwise ~ seed.bank | phage)
```

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

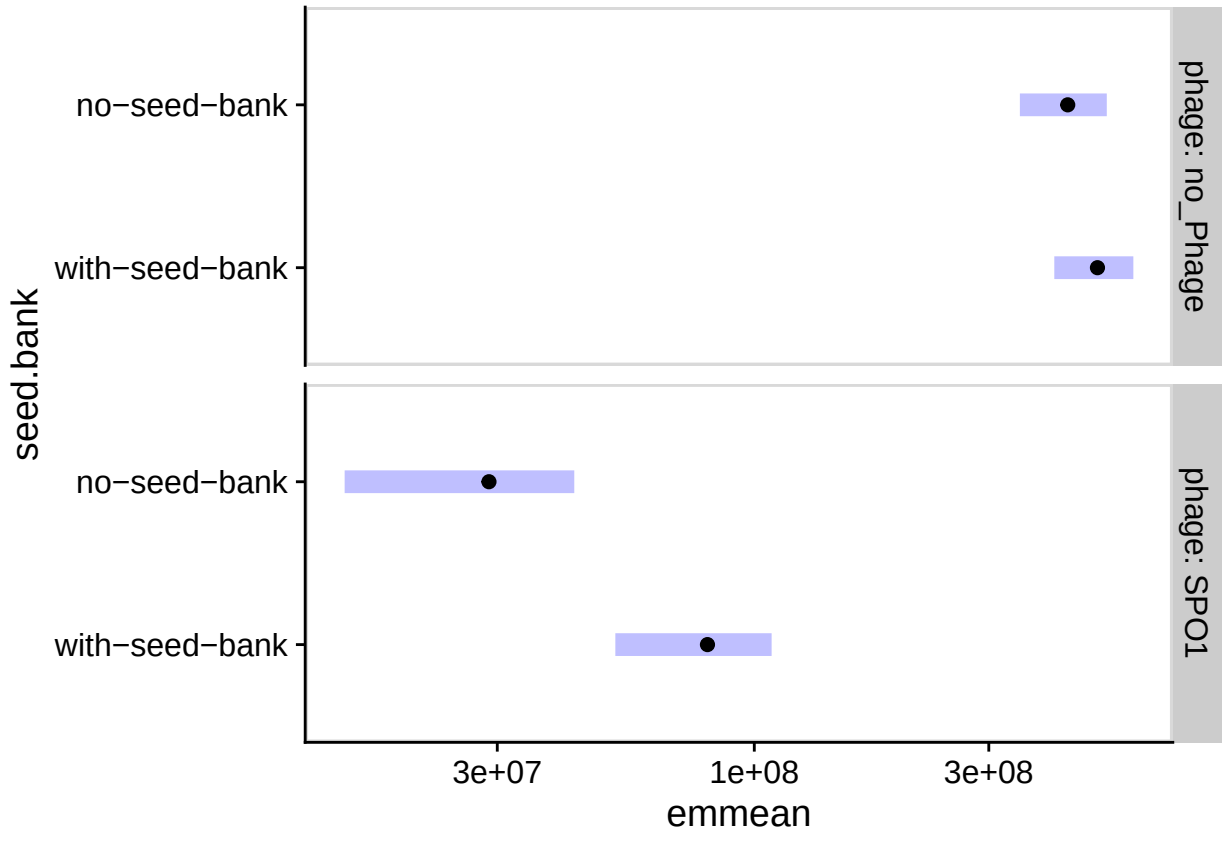
```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
contrast(regrid(emm_sb.ph$emmeans), method = "pairwise")
```

```
## phage = no_Phage:
##  contrast                          estimate       SE df t.ratio p.value
##  (with-seed-bank) - (no-seed-bank) 64876630 56156152  8 1.155   0.2813
##
## phage = SPO1:
##  contrast                          estimate       SE df t.ratio p.value
##  (with-seed-bank) - (no-seed-bank) 51454927 13668395  8 3.765   0.0055
##
## Results are averaged over the levels of: time.fct
## Degrees-of-freedom method: inherited from containment when re-gridding
```

```
plot(regrid(emm_sb.ph$emmeans))+
  theme_cowplot()+
  panel_border()+
  scale_x_log10()
```



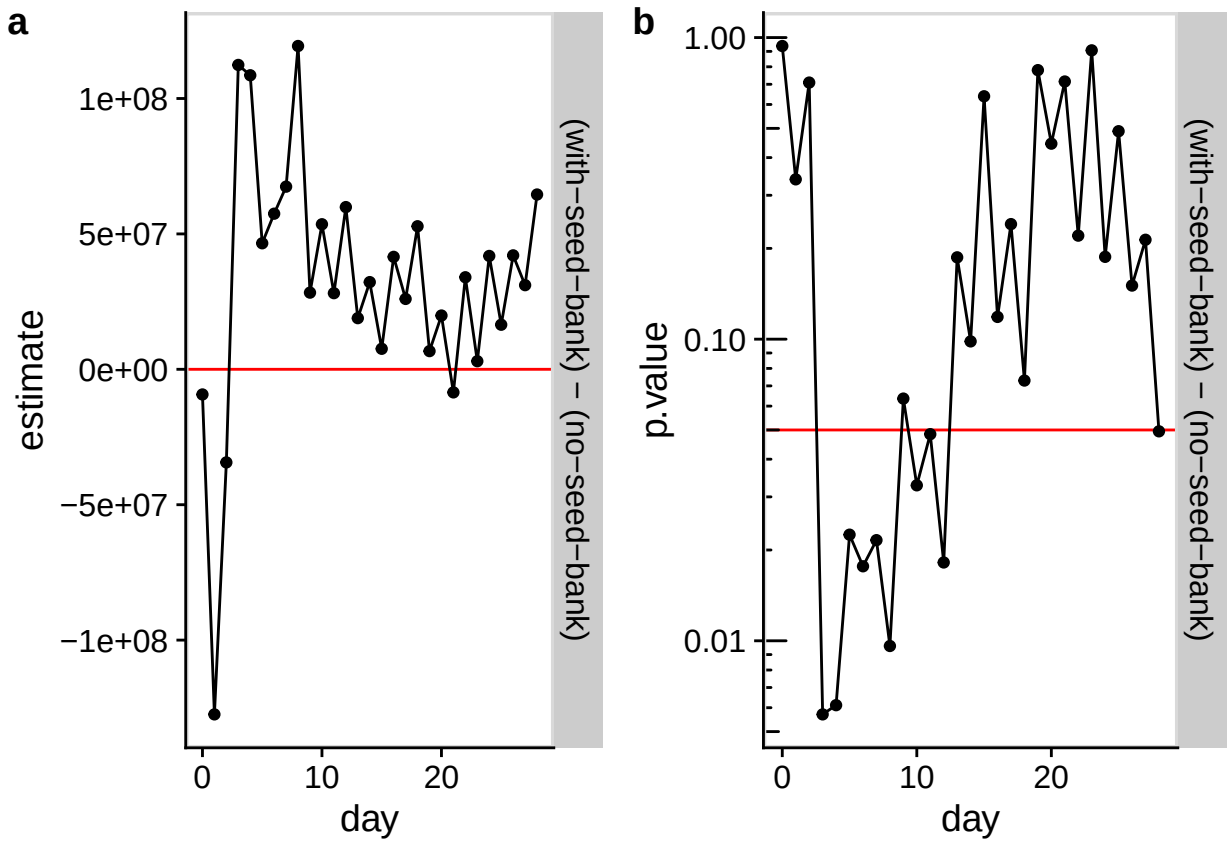
Indeed in the phage infected treatments we see that the populations differ between no-seed-bank and with-seed-bank treatments. The trend is that the absence of a seed-bank results in lower cell densities.

We also saw an interaction with time. We next compare the seed-bank treatments across time, separating the phage treatments, and focusing on the phage treated samples.

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

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

```
coevo.emm.bc <- stats::update(coevo.emm, tran = make.tran("boxcox", param =  bx.cx$roundlam))

coevo.pairs <- pairs(regrid(coevo.emm.bc), simple="seed.bank")%>%
  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")
```



The "absence of seed bank" effect on population density is seen in days 4-12 of the experiment. In that period a significant reduction in cell density is seen. However after that time the effect of no-seed-bank is lost, until the last samples where as a trend, the population density is becoming lower in absence of seed bank and that hits the significance threshold on the last sample day.

Summary

The seed bank treatment alters population dynamics (cell density over time) in phage infected cultures only. Populations coevolving with phages in the absence of seed bank show a significant reduction in cell densities, most notably in the short to medium time frame. This is to say that the presence of a seed bank acts as a buffer that moderates the reduction in host population density caused by phage infection. The results suggest however that other mechanisms (presumably host evolution) can eventually enable the host to achieve the same resilience against phage-induced decimation of its population , but these may be less stable in the long run.