Survival analysis in an experimental microbial aquatic community: Visualizations

Samuel Pawel

29 Juli 2021

```r
## some libraries
library(ggplot2)
library(dplyr)
library(survival)

## import data set
dd <- readRDS("Dataset/extinction.week.Rdata")

## glimpse at data set
str(dd)
```

```
## 'data.frame':    3400 obs. of  5 variables:
##  $ jar         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ temp        : int  15 20 20 20 15 20 20 20 15 20 ...
##  $ energy      : num  0.275 0.275 0.275 0.825 0.825 0.825 0.275 0.275 0.275 0.825 ...
##  $ species     : chr  "blepharisma" "blepharisma" "blepharisma" "blepharisma" ...
##  $ week.persist: int  8 8 4 3 6 1 6 6 8 2 ...
```

```r
summary(dd)
```

```
##       jar              temp          energy         species
##  Min.   :  1.00   Min.   :15.0   Min.   :0.275   Length:3400
##  1st Qu.: 50.75   1st Qu.:15.0   1st Qu.:0.275   Class :character
##  Median :100.50   Median :17.5   Median :0.550   Mode  :character
##  Mean   :100.50   Mean   :17.5   Mean   :0.550
##  3rd Qu.:150.25   3rd Qu.:20.0   3rd Qu.:0.825
##  Max.   :200.00   Max.   :20.0   Max.   :0.825
##   week.persist
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :1.000
##  Mean   :3.302
##  3rd Qu.:8.000
##  Max.   :8.000
```

```r
## visualize how long species persisted stratified by experimental conditions
dd$condition <- paste0("Temperature: ", dd$temp,
                       "°C x Energy: ", dd$energy)
ggplot(data = dd, aes(x = week.persist, fill = condition)) +
    geom_bar() +
    facet_wrap(~ species, ncol = 4) +
    scale_x_continuous(minor_breaks = NULL) +
```

```
labs(x = "Weeks persisted", fill = "") +
guides(fill = guide_legend(ncol = 2)) +
theme_bw() +
theme(legend.position = "top")
```
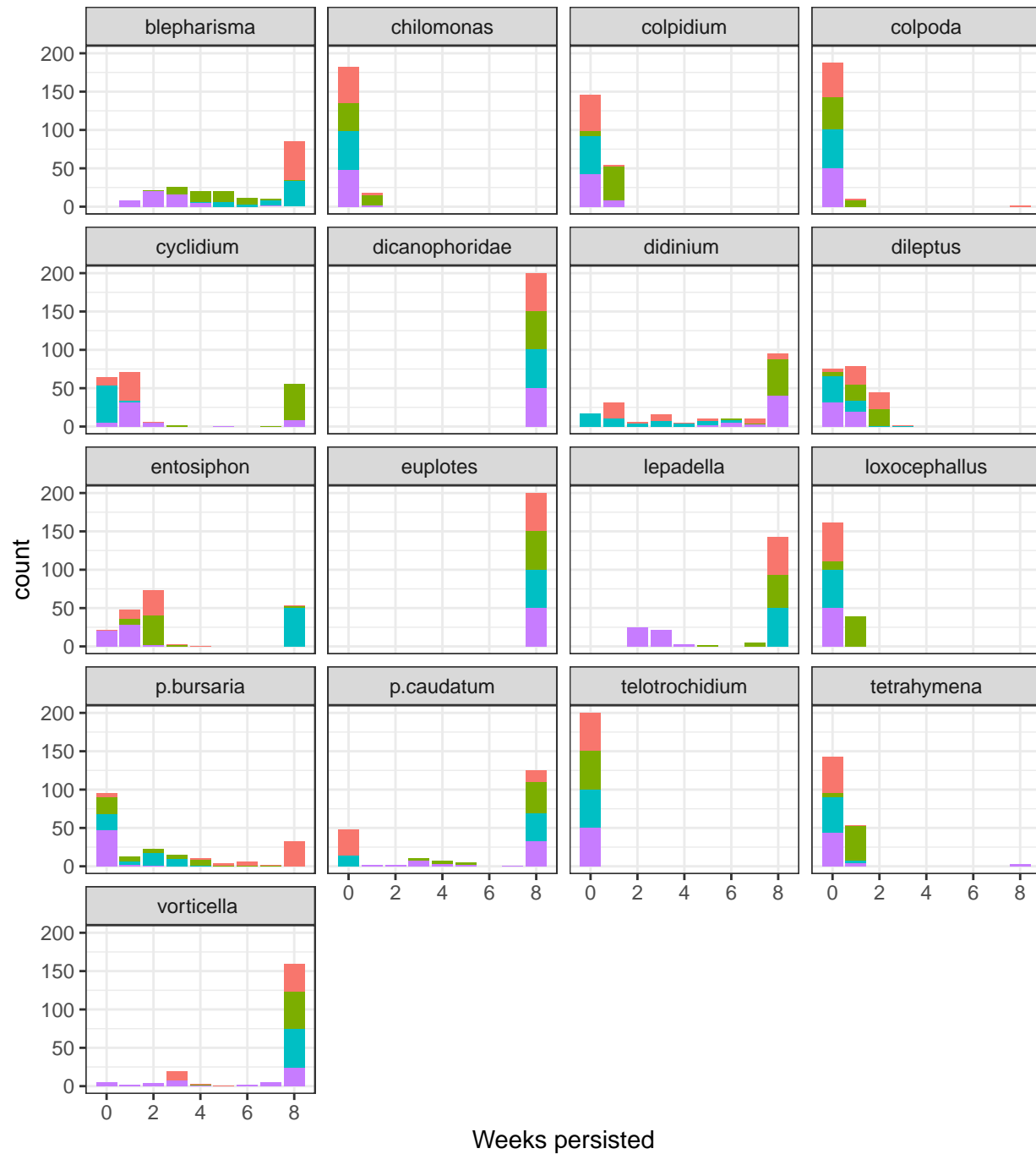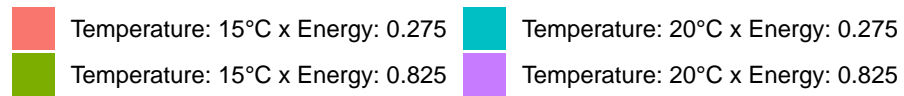
```r
## create survival outcome that takes into account interval censoring
dd$timeSurv <- Surv(time = dd$week.persist,
                    time2 = ifelse(dd$week.persist == 8, Inf, dd$week.persist + 1),
                    type = "interval2")

## compute Kaplan Meier survival curves for each species and for all treatments
applyGrid <- expand.grid(species = unique(dd$species),
                         energy = unique(dd$energy),
                         temp = unique(dd$temp), stringsAsFactors = FALSE)
plotList <- lapply(X = seq(1, nrow(applyGrid)), FUN = function(i) {
    spDat <- filter(dd,
                    temp == applyGrid$temp[i],
                    species == applyGrid$species[i],
                    energy == applyGrid$energy[i])
    kmFit <- survfit(formula = timeSurv ~ 1, data = spDat)
    kmPredictions <- summary(kmFit, times = seq(0, 8, 1), extend = TRUE)
    data.frame(surv = kmPredictions$surv,
               lower = kmPredictions$lower,
               upper = kmPredictions$upper,
               time = kmPredictions$time,
               applyGrid[i,])
})

plotDF <- do.call("rbind", plotList)
plotDF$condition <- paste0("Temperature: ", plotDF$temp,
                           "°C x Energy: ", plotDF$energy)

## plot survival curves
ggplot(data = plotDF, aes(x = time , y = surv, color = condition)) +
    geom_step(alpha = 0.8) +
    facet_wrap(~ species, ncol = 4) +
    scale_x_continuous(minor_breaks = NULL) +
    labs(x = "Time (weeks)", y = "Survival probability", color = "") +
    guides(color = guide_legend(ncol = 2)) +
    theme_bw() +
    theme(legend.position = "top")
```
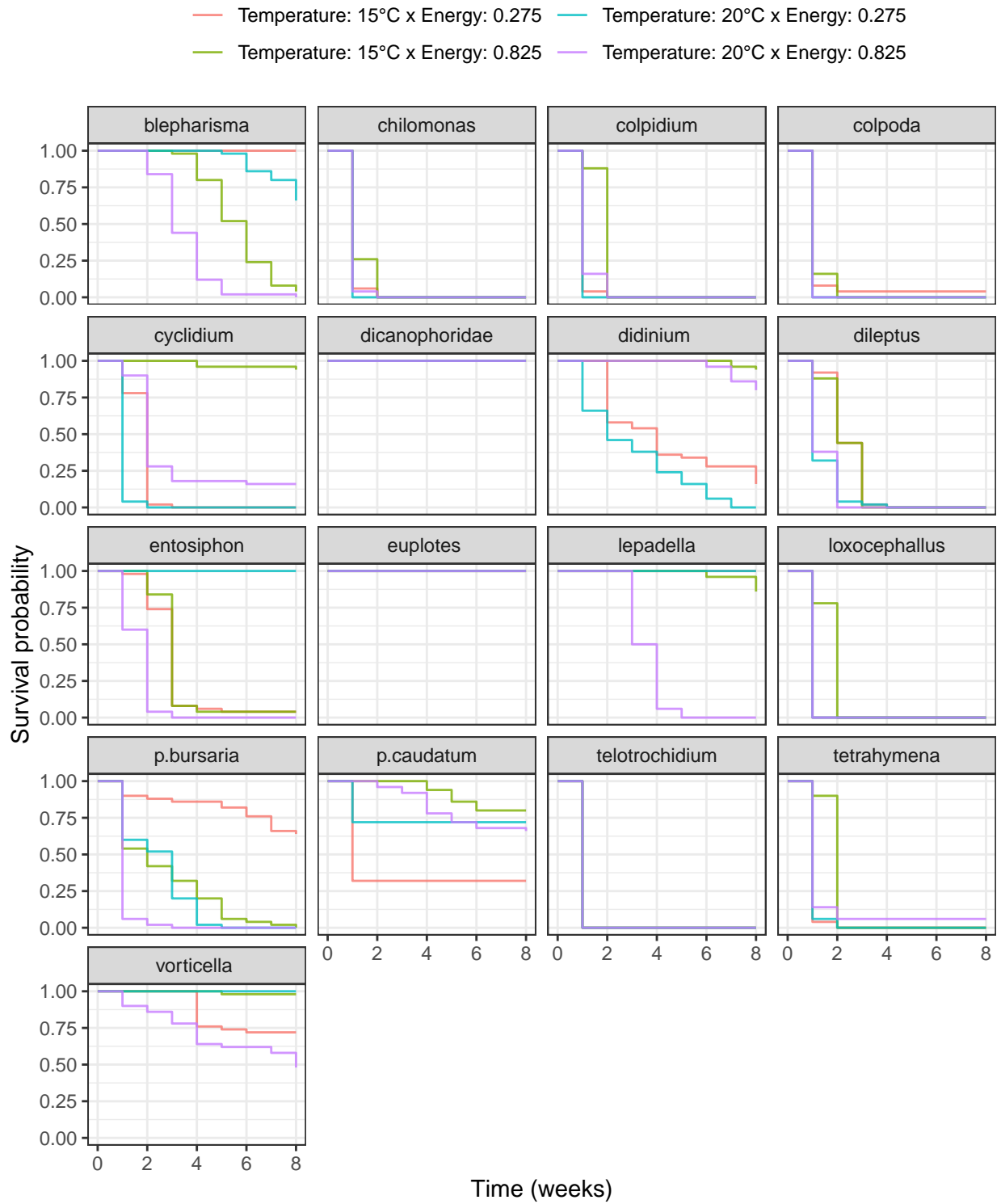
- There are species which went extinct already in the first week or which survived all 8 weeks. The estimated survivor curves look odd in both cases.

- Some species show either increased or decreased survival depending on the experimental conditions.

- The survival of some species seems unaffected by the experimental conditions.