

Plot Model Predictions

Brooke Anderson

7/15/2021

About the data

Collin took alone and pairwise CFU data for each species at day 21, assuming this was the “equilibrium” point (which would be where population size no longer changes, and $dn/dt == 0$), and used it to fit the Lotka-Volterra competition model to predict cell counts of each species in the complete community.

```
estimates <- rbind(read_csv(here("wrangled_data/model_predictions/coefficient-estimates-pH5.csv")),
                  read_csv(here("wrangled_data/model_predictions/coefficient-estimates-pH7.csv"))) %>%
  mutate(pH = 7)) %>%
  mutate(spec = recode(spec, "X135E" = "135E"), name = recode(name, "X135E" = "135E"))

##
## -- Column specification -----
## cols(
##   spec = col_character(),
##   name = col_character(),
##   est = col_double(),
##   se = col_double(),
##   pH = col_double()
## )

##
## -- Column specification -----
## cols(
##   spec = col_character(),
##   name = col_character(),
##   est = col_double(),
##   se = col_double()
## )

predictions <- rbind(read_csv(here("wrangled_data/model_predictions/predictions-from-pH5.csv"))) %>%
  mutate(pH = 5),
  read_csv(here("wrangled_data/model_predictions/predictions-from-pH7.csv"))) %>%
  mutate(pH = 7)) %>%
  select(spec, pH, replicate, fit) %>%
  mutate(spec = recode(spec, "X135E" = "135E")) %>%
  na.omit()

##
## -- Column specification -----
## cols(
##   fit = col_double(),
##   lwr = col_double(),
##   upr = col_double(),
##   replicate = col_double(),
##   spec = col_character(),
##   fam = col_character()
```

```
## )

##
## -- Column specification -----
## cols(
##   fit = col_double(),
##   lwr = col_double(),
##   upr = col_double(),
##   replicate = col_double(),
##   spec = col_character(),
##   fam = col_character()
## )

data <- read_csv(here("wrangled_data/model_predictions/community-data-long.csv")) %>%
  mutate(spec = recode(spec, "X135E" = "135E"),
         fit = fit - 1)

##
## -- Column specification -----
## cols(
##   rep = col_double(),
##   spec = col_character(),
##   fit = col_double()
## )

colnames(data) <- c("replicate", "spec", "fit")
```

So I have:

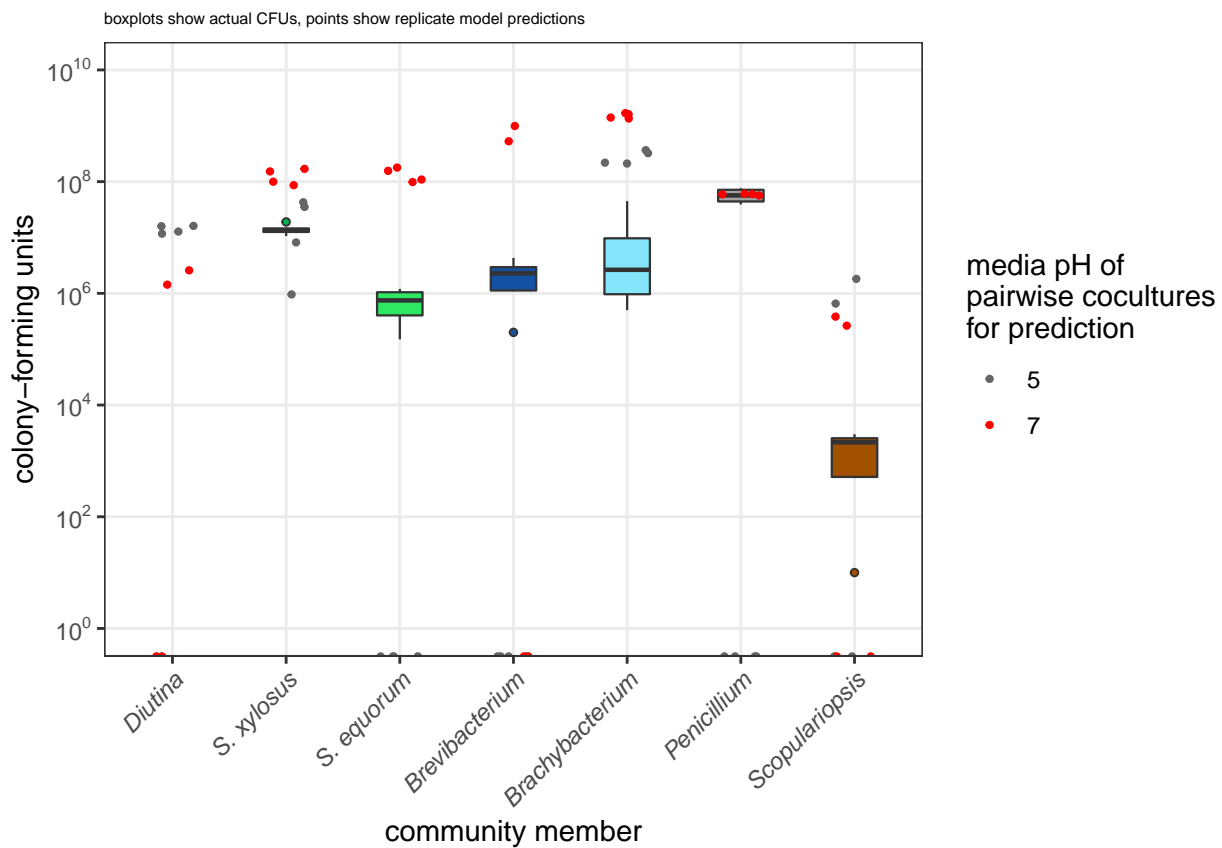
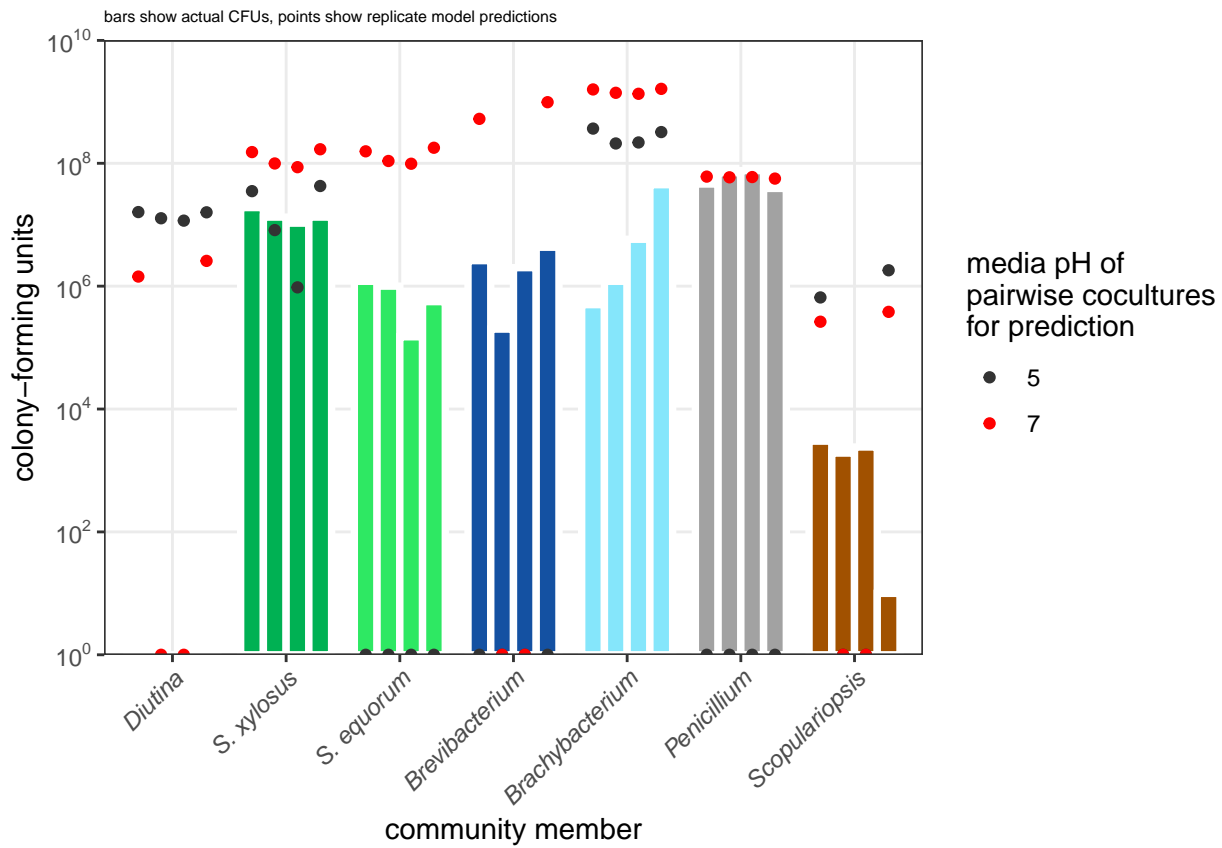
- actual community data,
- predictions of community CFUs from LV modeling of PW/alone growth data, and
- pairwise effect metrics from each pairwise partner species

Our main results are that:

- PW growth poorly predicts absolute community composition
- pH 7 PW data better predicts community composition than pH 5 data
- improved predictions from pH 7 pairwise data – coming from Brevi/Brachy pairwise effect metrics?

Let's show these things:

PW growth poorly predicts absolute community composition



Let's focus in on the deviations of predicted population size from actual CFUs when grown in a community:

```

predictions.w <- predictions %>%
  pivot_wider(names_from = "pH", names_prefix = "prediction_pH",
              values_from = "fit")

comparisons <- full_join(data, # %>% mutate(fit = fit+1), # no infinite fold-changes,
                        predictions.w, by = c("spec", "replicate")) %>%
  mutate(diff_ph5_pred = prediction_ph5 - fit,
         diff_ph7_pred = prediction_ph7 - fit,
         perc_ph5_pred = diff_ph5_pred / fit * 100,
         perc_ph7_pred = diff_ph7_pred / fit * 100,
         log_dev_ph5 = ifelse(perc_ph5_pred < 0,
                              -log10(abs(perc_ph5_pred)),
                              log10(perc_ph5_pred)),
         log_dev_ph7 = ifelse(perc_ph7_pred < 0,
                              -log10(abs(perc_ph7_pred)),
                              log10(perc_ph7_pred)))

## Warning in ifelse(perc_ph5_pred < 0, -log10(abs(perc_ph5_pred)),
## log10(perc_ph5_pred)): NaNs produced

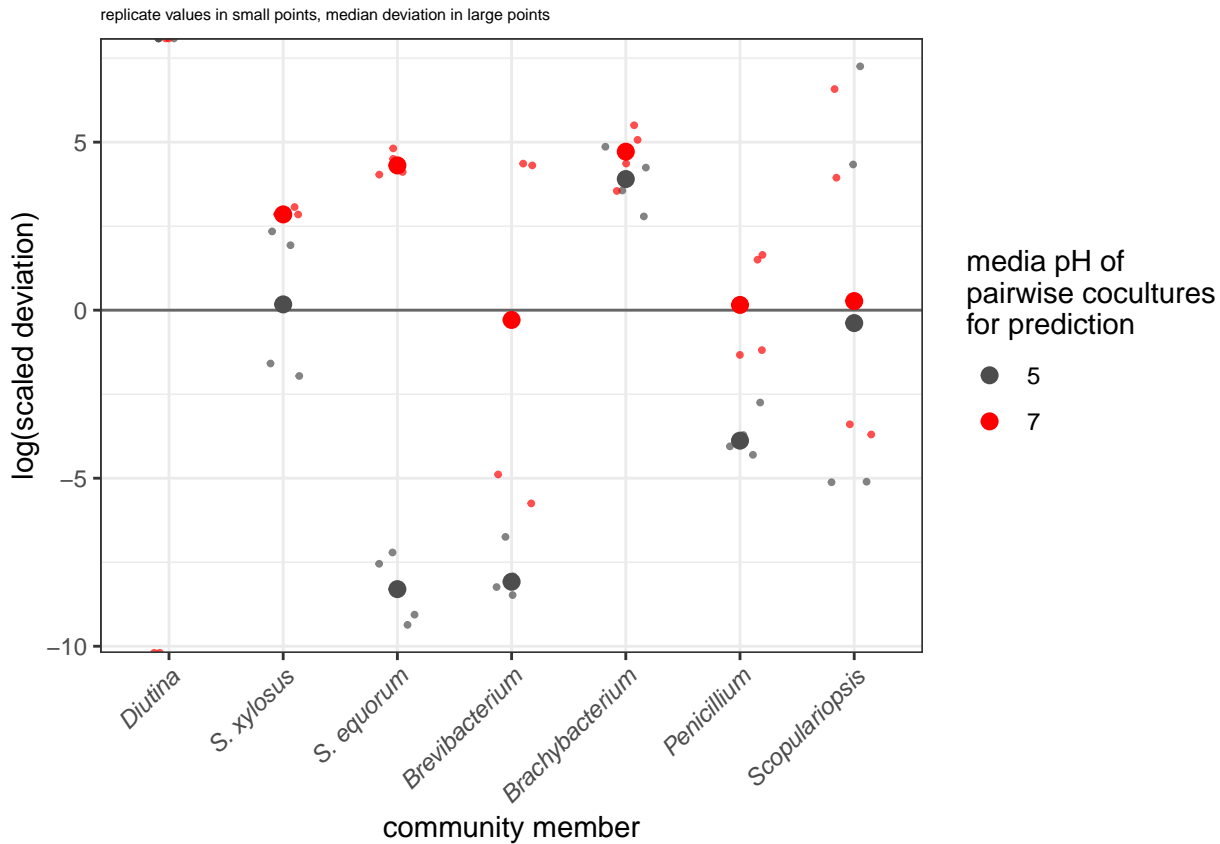
## Warning in ifelse(perc_ph7_pred < 0, -log10(abs(perc_ph7_pred)),
## log10(perc_ph7_pred)): NaNs produced

p.deviations <- ggplot(comparisons %>%
  pivot_longer(cols = c(log_dev_ph5, log_dev_ph7), values_to = "log_deviation",
               names_to = "pH", names_prefix = "log_dev_ph") %>%
  mutate(spec = factor(spec, levels = names(names)[-c(8:9)])),
  aes(x = spec, y = log_deviation, color = pH)) +
  scale_x_discrete(labels = names) +
  geom_jitter(size = .7, alpha = 0.7, width = .2) +
  scale_color_manual(values = c("gray30", "red")) +
  geom_hline(yintercept = 0, color = "gray40") +
  theme_bw() +
  labs(y = "log(scaled deviation)",
       x = "community member",
       color = paste("media pH of\npairwise cocultures\nfor prediction"),
       subtitle = "replicate log(%difference from actual) in small points") +
  theme(axis.text.x = element_text(angle = 45, face = "italic", hjust = 1),
        plot.subtitle = element_text(size = 6))

p.deviations +
  stat_summary(geom = "point", fun = median, size = 2.5) +
  labs(subtitle = "replicate values in small points, median deviation in large points")

## Warning: Removed 10 rows containing non-finite values (stat_summary).
## Warning: Removed 2 rows containing missing values (geom_point).

```



For *Diutina* (in that pH 7 predicts it to die off), *S. equorum*, *Brevibacterium*, and *Penicillium*, pH 7 PW data better predicts community composition than pH 5 data:

Can we look back into estimates and peg interactions that might be contributing extra to strong prediction deviations from actual CFUs?

Estimates should be interpreted as a per-capita replacement measurement: for every 1

```
## <list_of<
##   tbl_df<
##     spec : character
##     name : character
##     est_5: double
##     est_7: double
##     se_5 : double
##     se_7 : double
##   >
## >[7]>
## [[1]]
## # A tibble: 7 x 6
##   spec name      est_5      est_7      se_5      se_7
##   <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 135E K      1.80e+7  8.83e+6 2525975. 3057441.
## 2 135E JB370  1.19e+0 -1.07e+0    1.19    2.83
## 3 135E BC10  -1.46e-1 -3.06e-3    0.0396   0.0157
## 4 135E JBC   1.04e-1  1.63e-1    0.0387   0.0919
## 5 135E BC9   -8.19e-2 -2.04e-2    0.0198   0.0162
## 6 135E JB5    1.21e-3 -5.38e-4    0.00219  0.00130
## 7 135E JB7   -8.47e-5 -9.12e-4    0.00172  0.00111
##
```

```
## [[2]]
## # A tibble: 7 x 6
##   spec name      est_5      est_7      se_5      se_7
##   <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 BC10 K      8.59e+7  2.55e+8 17395694.  34120612.
## 2 BC10 JB370 -2.10e+1  1.74e+2   4.29    97.8
## 3 BC10 JBC     1.12e+0  2.21e+0   0.536    0.951
## 4 BC10 BC9     -6.09e-1  4.79e-1   0.909    1.26
## 5 BC10 135E    -4.50e-1  1.55e+0   0.805    3.45
## 6 BC10 JB5      4.00e-3 -2.77e-3   0.0112   0.0150
## 7 BC10 JB7     -2.40e-3 -1.68e-2   0.0211   0.0213
##
## [[3]]
## # A tibble: 7 x 6
##   spec name      est_5      est_7      se_5      se_7
##   <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 BC9 K      48564127.  2.77e+8 34111561.  38270130.
## 2 BC9 JB7     288927.  -2.83e-2 772340.    0.0257
## 3 BC9 JB5     19682.  4.96e-3 50595.    0.0176
## 4 BC9 JB370    -24.7  1.11e+2   4.15    89.1
## 5 BC9 135E     -4.21 -6.67e-1   1.40    3.03
## 6 BC9 JBC       0.507  2.21e+0   0.868    1.51
## 7 BC9 BC10      0.228  9.41e-1   0.454    0.265
##
## [[4]]
## # A tibble: 7 x 6
##   spec name      est_5      est_7      se_5      se_7
##   <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 JB370 K      6450740.  910919. 1092010. 128776.
## 2 JB370 135E     0.197   -0.112   0.160   0.0241
## 3 JB370 JBC     0.128   0.0133  0.0500  0.00392
## 4 JB370 BC9    -0.0145  0.00142  0.00475  0.000874
## 5 JB370 JB5    -0.00928 -0.000171 0.0153  0.000284
## 6 JB370 JB7    -0.00609 -0.000304 0.00627  0.000164
## 7 JB370 BC10   -0.00183  0.00172  0.00906  0.00128
##
## [[5]]
## # A tibble: 7 x 6
##   spec name      est_5      est_7      se_5      se_7
##   <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 JB5 K      357069137.  3.73e+9 295783144.  440177922.
## 2 JB5 JB7     289212.  8.35e-1 597239.    0.364
## 3 JB5 135E    -92.8  2.28e+1  32.8    58.4
## 4 JB5 JB370    29.0  2.81e+3  57.2    645.
## 5 JB5 BC10   -17.7  6.25e-1  5.77    2.57
## 6 JB5 BC9     10.5  2.12e+0  16.7    2.07
## 7 JB5 JBC      5.39  6.90e+1  8.18    13.7
##
## [[6]]
## # A tibble: 7 x 6
##   spec name      est_5      est_7      se_5      se_7
##   <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 JB7 K     -20627980.  1.88e+9 97255200. 193547405.
## 2 JB7 135E    -148.  -1.91e+2  9.67    22.7
## 3 JB7 JB370   -33.3  2.37e+2  22.0    228.
## 4 JB7 JB5     -28.5  1.13e-2  330.    0.134
```

```
## 5 JB7 BC10 -16.1 -2.48e+0 2.08 0.949
## 6 JB7 BC9 -0.727 -1.27e+0 7.72 0.717
## 7 JB7 JBC -0.144 7.24e+0 1.04 6.50
##
## [[7]]
## # A tibble: 7 x 6
## spec name est_5 est_7 se_5 se_7
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>
## 1 JBC K 6.96e+7 5.64e+7 8528382. 4605177.
## 2 JBC 135E 5.33e+4 NA 36676. NA
## 3 JBC JB5 2.00e+3 1.74e-2 30360. 0.119
## 4 JBC JB370 3.06e+1 -1.29e+3 38.9 5531.
## 5 JBC JB7 -1.56e+1 7.04e-3 2.43 0.00558
## 6 JBC BC10 5.61e-1 -2.81e-2 0.853 0.0721
## 7 JBC BC9 -4.69e-3 5.02e-2 0.828 0.0453
```

Some of these estimates aren't making sense – e.g. at pH 5 for $i = \text{JB5}$, $j = \text{JB7}$, $\alpha_{ij} > 2e5$? But neither of these species affect the other in pairwise coculture??

Let's generate a table that we can import into cytoscape to compare to actual effects: