Pseudomonas Epiphytic Growth and Virulence Analysis

2024 Summer

Contents

Load Libraries

```
pacman::p_load(ggplot2, readxl, RColorBrewer, paletteer, viridis, wesanderson, ggbeeswarm,
    dplyr, tidyverse, devtools, emmeans, cowplot, knitr, survival, here, tibble,
    survminer, lubridate, formatR, gridExtra, ggsurvfit, gtsummary, tidycmprsk, install = FALSE)
```

Load Data

```r
# Define the desired order of the treatments
strains_order <- c("194", "200", "204", "205", "215", "216", "220", "221", "227",
    "228", "B728a", "Cit7", "pisi", "Control")

strains_df <- read_excel("data/strains_R.xlsx")

epi_growth_data <- read_excel("data/pseud_epi_growth_2024summer_R.xlsx")

# Convert the strain variable to a factor with the specified order
epi_growth_data$treatment <- factor(epi_growth_data$treatment, levels = strains_order)

# Join the strains dataframe to epi growth
epi_growth_data <- left_join(epi_growth_data, strains_df, by = "treatment")

# Load virulence data
aphid_virulence_data <- read_csv("data/virulence_new_all.csv")
```

```
## Rows: 4973 Columns: 6
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr (3): date, treatment, replicate
## dbl (3): individual, censored, time
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Convert the treatment variable to a factor with the specified order
aphid_virulence_data$treatment <- factor(aphid_virulence_data$treatment, levels = strains_order)

# Join the strains dataframe to virulence
aphid_virulence_data <- left_join(aphid_virulence_data, strains_df, by = "treatment")
```

Load colors

```r
# Extract the default ggplot colors
default_colors <- (scales::hue_pal())(length(strains_order) - 1)

# Create a named vector for strain colors, ensuring 'Control' is black
strain_colors <- setNames(c(default_colors, "black"), strains_order)

# Load species colors
species_colors <- c(`P. fluorescens` = "dodgerblue2", `P. syringae` = "springgreen3",
    `P. putida` = "sienna4", Control = "black")

species_colors_manual <- c(`194` = "sienna4", `200` = "dodgerblue2", `204` = "dodgerblue2",
    `205` = "dodgerblue2", `215` = "springgreen3", `216` = "dodgerblue2", `220` = "sienna4",
    `221` = "dodgerblue2", `227` = "dodgerblue2", `228` = "dodgerblue2", B728a = "springgreen3",
    Cit7 = "springgreen3", pisi = "springgreen3", Control = "black")
```

Goals

- Create Kaplan-Meier curve for Pseud. virulence data
- Use stats (Wilcox?) to determine statistical significance of each strain
- Compare virulence data with epiphitic growth ability

Virulence Analysis

Note: *In order to help me with this analysis, I am using the following sites - Survival Analysis in R and Hazard Ratio: Interpretation & Definition.*

Caclulcate survival probabilities for each strain and create dataframe

Survival model

```r
# Fit the survival model
km_fit <- survfit(Surv(time, censored) ~ treatment, data = aphid_virulence_data)
```

Extract survival probability

```r
# Fit the survival model
km_fit <- survfit(Surv(time, censored) ~ treatment, data = aphid_virulence_data)

# Extract survival probabilities at specific time points
time_points <- c(72)
km_summary <- summary(km_fit, times = time_points)

# Calculate standard error (assuming fit includes the necessary information)
std_error <- summary(km_fit, times = time_points)$std.err

# Initialize empty lists to store the results
times_list <- list()
treatment_list <- list()
surv_prob_list <- list()
std_error_list <- list()

# Loop over each treatment group and extract survival probabilities at
# specified time points
for (i in 1:length(km_fit$strata)) {
    treatment_name <- names(km_fit$strata)[i]
    for (t in time_points) {
        idx <- which(km_summary$time == t & km_summary$strata == treatment_name)
        if (length(idx) > 0) {
            times_list <- c(times_list, t)
            treatment_list <- c(treatment_list, treatment_name)
            surv_prob_list <- c(surv_prob_list, km_summary$surv[idx])
            std_error_list <- c(std_error_list, km_summary$std.err[idx])
        } else {
```

```r
        times_list <- c(times_list, t)
        treatment_list <- c(treatment_list, treatment_name)
        surv_prob_list <- c(surv_prob_list, NA)
        std_error_list <- c(std_error_list, NA)
      }
    }
}

# Create the data frame
surv_probs <- data.frame(time = unlist(times_list), treatment = unlist(treatment_list),
    surv_prob = unlist(surv_prob_list), std_error_data = unlist(std_error_list))

# Replace 'treatment=' with an empty string
surv_probs$treatment <- gsub("treatment=", "", surv_probs$treatment)

# Convert the treatment variable to a factor with the specified order
surv_probs$treatment <- factor(surv_probs$treatment, levels = strains_order)

# Join the survival dataframe to strains
surv_probs <- left_join(surv_probs, strains_df, by = "treatment")

# Print the data frame
print(surv_probs)
```

```
##    time treatment  surv_prob std_error_data              strain     -80
## 1    72       194 0.12244898    0.019117897     field sample #8     194
## 2    72       200 0.02230483    0.009003784    field sample #14     200
## 3    72       204 0.05281690    0.013272241    field sample #17     204
## 4    72       205 0.07118644    0.014971038  field sample #18-1     205
## 5    72       215 0.22222222    0.024497697  field sample #26-1     215
## 6    72       216 0.15625000    0.021395412  field sample #26-2     216
## 7    72       220 0.17229730    0.021949800  field sample #29-2     220
## 8    72       221 0.09363296    0.017828335  field sample #29-3     221
## 9    72       227 0.16949153    0.021844143     field sample #35     227
## 10   72       228 0.10447761    0.018684542     field sample #36     228
## 11   72     B728a 0.10996564    0.018339411         B728a ΔFlgK   B728a
## 12   72      Cit7 0.17073171    0.022210773                Cit7      14
## 13   72   Control 0.81678082    0.011319228             Control Control
## 14   72      pisi 0.74151436    0.022370650               1704B       6
##    previous_work species_complex          species_match
## 1            new  P. fluorescens          P. paralactis
## 2            new  P. fluorescens         P. fluorescens
## 3            new  P. fluorescens            P. rhodesiae
## 4            new  P. fluorescens           P. orientalis
## 5            new     P. syringae            P. syringae
## 6            new       P. putida              P. putida
## 7            new  P. fluorescens        P. pergaminensis
## 8            new  P. fluorescens         P. fluorescens
## 9            new  P. fluorescens        P. pergaminensis
## 10           new     P. syringae         P. lijiangensis
## 11           old     P. syringae P. syringae pv. Syringae
## 12           old     P. syringae            P. syringae
## 13           old         Control                Control
```

4

```
## 14          old     P. syringae              P. syringae
##                collector              source    ANI
## 1         Kathryn L Herr      Pisum sativum   98.44
## 2         Kathryn L Herr      Pisum sativum    <NA>
## 3         Kathryn L Herr      Pisum sativum   98.92
## 4         Kathryn L Herr      Pisum sativum   95.78
## 5         Kathryn L Herr      Pisum sativum   95.57
## 6         Kathryn L Herr      Pisum sativum   95.42
## 7         Kathryn L Herr      Pisum sativum   98.64
## 8         Kathryn L Herr      Pisum sativum    <NA>
## 9         Kathryn L Herr      Pisum sativum   98.22
## 10        Kathryn L Herr      Pisum sativum    96.6
## 11 Alan Collmer/J. Lindemann?  Phaseolus vulgaris    <NA>
## 12                   <NA>      Pisum sativum    <NA>
## 13               Control            Control Control
## 14         David Baltrus Citrus leaf surface    <NA>
##                       genome_match   Reseq/Genbank
## 1  Pseudomonas paralactis DSM 29164 GCA_001439735.1
## 2                            <NA>            <NA>
## 3   Pseudomonas rhodesiae CIP104664 GCF_024169765.1
## 4  Pseudomonas orientalis DSM 17489 GCF_001439815.1
## 5  Pseudomonas syringae DSM 10604   GCA_000597765.1
## 6        Pseudomonas putida S13.1.2 GCF_000498395.2
## 7   Pseudomonas pergaminensis 1008 GCF_024112395.1
## 8                            <NA>            <NA>
## 9   Pseudomonas pergaminensis 1008 GCF_024112395.1
## 10    Pseudomonas lijiangensis LJ2 GCF_018968705.1
## 11                           <NA>        CP000075
## 12                           <NA>    AEAJ 00000000
## 13                        Control         Control
## 14                           <NA>    AEAI 00000000
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11 https://doi.org/10.1146/annurev.py.28.090190.001103
## 12
## 177. https://doi.org/10.1146/annurev.py.28.090190.001103
## 13
## 14 Baltrus, D. A., Nishimura, M. T., Romanchuk, A., Chang, J. H., Mukhtar, M. S., Cherkis, K., Roach
```

Post-hoc test for survival

```r
# Perform pairwise log-rank tests between treatment groups and the control
# group
pairwise_tests_treatment <- pairwise_survdiff(Surv(time, censored) ~ treatment, data = aphid_virulence_d

# Print the pairwise tests
print(pairwise_tests_treatment)
```

```
##
##  Pairwise comparisons using Log-Rank test
##
## data:  aphid_virulence_data and treatment
##
##           194     200     204     205     215     216     220     221     227
## 200     2.3e-12 -       -       -       -       -       -       -       -
## 204     1.6e-08 0.26489 -       -       -       -       -       -       -
## 205     3.7e-06 0.04009 0.38289 -       -       -       -       -       -
## 215     4.6e-05 < 2e-16 < 2e-16 7.8e-16 -       -       -       -       -
## 216     0.20589 3.0e-06 0.00054 0.00915 1.3e-05 -       -       -       -
## 220     1.2e-05 0.34987 0.95899 0.61067 7.5e-10 0.04787 -       -       -
## 221     0.06597 1.7e-06 0.00063 0.01560 3.0e-08 0.57084 0.05899 -       -
## 227     0.00048 0.05275 0.34910 0.75895 1.7e-08 0.14997 0.65606 0.23547 -
## 228     0.02457 2.3e-05 0.00265 0.03730 5.6e-09 0.52684 0.02668 0.75895 0.15756
## B728a   0.30241 3.8e-15 2.0e-10 1.5e-07 0.00193 0.15635 0.00089 0.00913 0.00558
## Cit7    0.97833 6.8e-10 1.0e-06 8.4e-05 0.00044 0.38289 0.00217 0.11439 0.01459
## Control < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16
## pisi    < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16
##           228     B728a   Cit7    Control
## 200     -       -       -       -
## 204     -       -       -       -
## 205     -       -       -       -
## 215     -       -       -       -
## 216     -       -       -       -
## 220     -       -       -       -
## 221     -       -       -       -
## 227     -       -       -       -
## 228     -       -       -       -
## B728a   0.00265 -       -       -
## Cit7    0.05646 0.46912 -       -
## Control < 2e-16 < 2e-16 < 2e-16 -
## pisi    < 2e-16 < 2e-16 < 2e-16 0.00042
##
## P value adjustment method: BH
```

```r
# Perform pairwise log-rank tests between species_complex groups and the
# control group
pairwise_tests_species <- pairwise_survdiff(Surv(time, censored) ~ species_complex,
    data = aphid_virulence_data)

# Print the pairwise tests
print(pairwise_tests_species)
```

```
##
```

```
##  Pairwise comparisons using Log-Rank test
##
## data:  aphid_virulence_data and species_complex
##
##                Control P. fluorescens P. putida
## P. fluorescens < 2e-16 -             -
## P. putida      < 2e-16 0.0084        -
## P. syringae    < 2e-16 < 2e-16       2.6e-10
##
## P value adjustment method: BH
```

Plot Surival Dataframe at different times

```r
surv_plot <- ggplot(data = surv_probs, aes(x = treatment, y = (1-surv_prob), color = species_complex, f
  geom_col() +
  geom_errorbar(aes(ymin = (1-surv_prob) - std_error_data, ymax = (1-surv_prob) + std_error_data),
                width = 0.3,  # Width of error bars
                position = position_dodge(width = 0.9)) +  # Dodge bars slightly
  labs(x = "-80 Strain #",
       y = "Aphid Mortality at 72 Hours (%)") +
  theme_minimal() +
  coord_flip() +
  scale_fill_manual(values = species_colors) +
  scale_color_manual(values = species_colors) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "bottom")

surv_plot
```
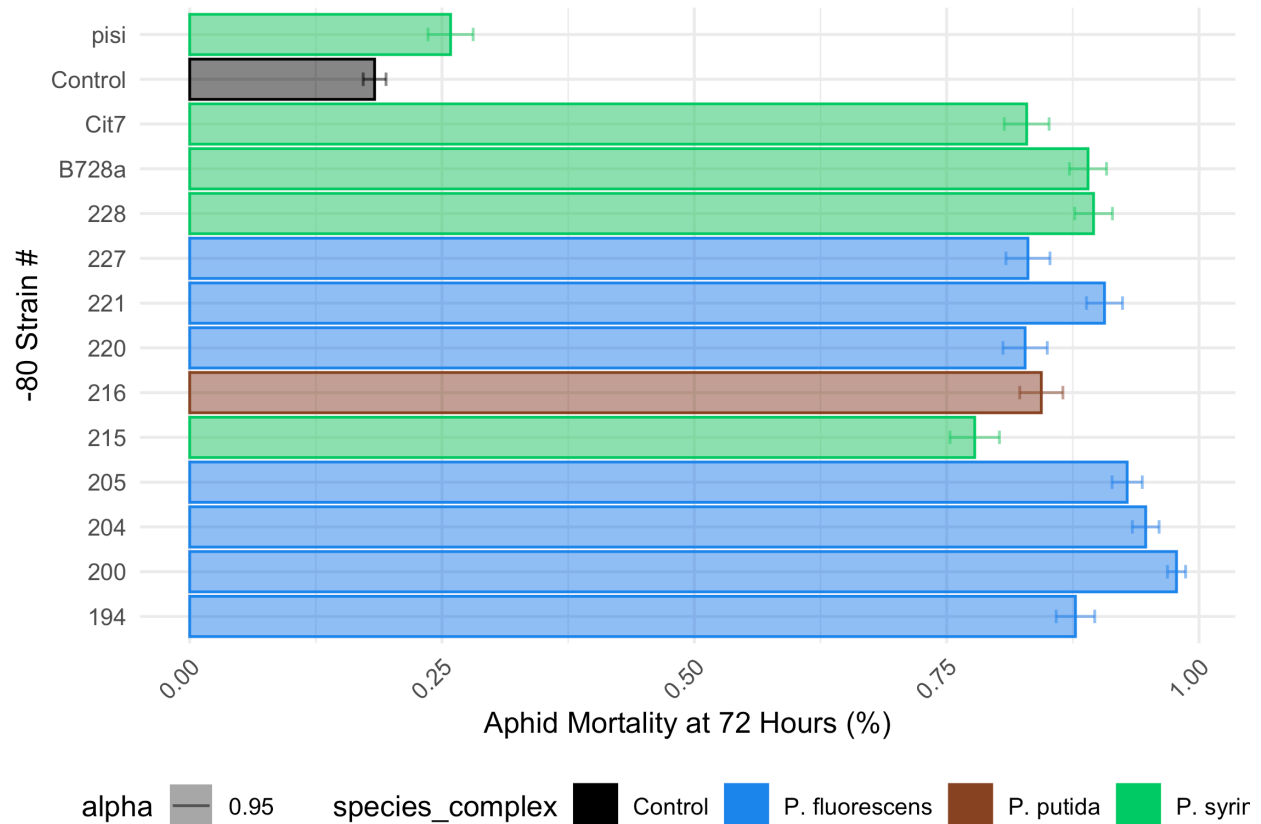
Make Kaplan-Meier Plot

Cohort Survival Curve

```
# Plot the Kaplan-Meier survival curves by treatment
surv_curv_treatment <- survfit2(Surv(time, censored) ~ treatment, data = aphid_virulence_data) %>%
    ggsurvfit(type = "survival") + labs(x = "Hours", y = "Overall survival probability") +
    theme_minimal() + scale_fill_manual(values = species_colors_manual) + scale_color_manual(values = sp
    theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 45,
        hjust = 1), legend.position = "right")

# Plot the Kaplan-Meier survival curves by species complex
surv_curv_species <- survfit2(Surv(time, censored) ~ species_complex, data = aphid_virulence_data) %>%
    ggsurvfit(type = "survival") + labs(x = "Hours", y = "Overall survival probability") +
    theme_minimal() + scale_fill_manual(values = species_colors) + scale_color_manual(values = species_
    theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 45,
        hjust = 1), legend.position = "right")

# show plots
surv_curv_treatment
```
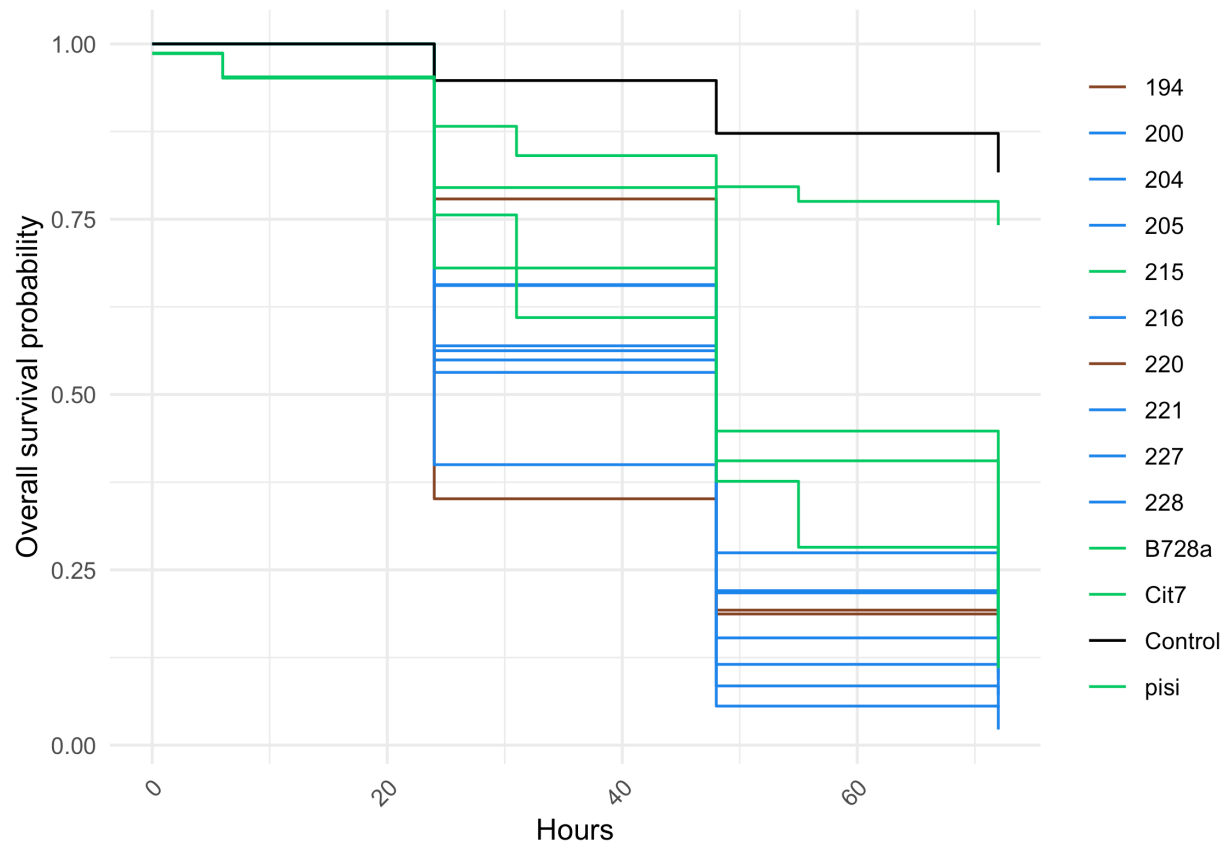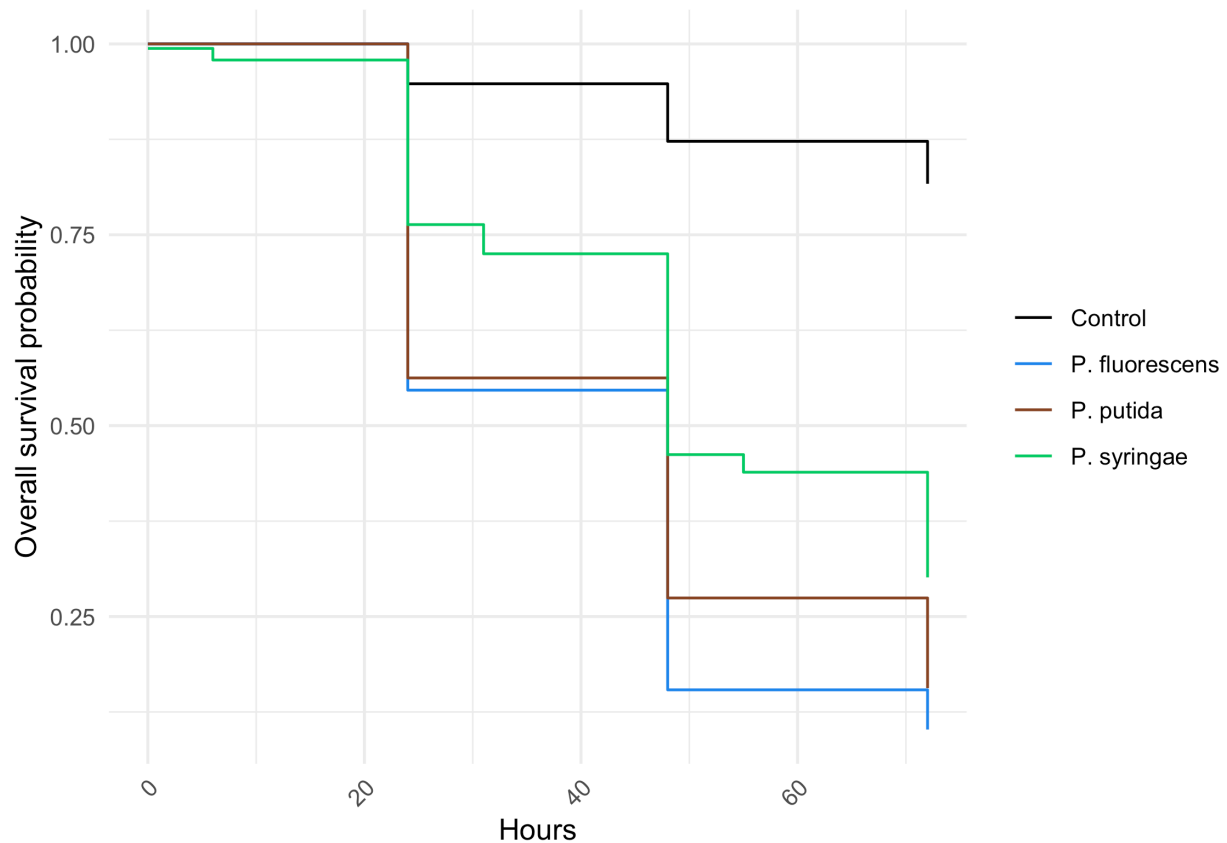
```
## Warning: No shared levels found between `names(values)` of the manual scale and the
## data's fill values.
```

surv_curv_species

```
## Warning: No shared levels found between `names(values)` of the manual scale and the
## data's fill values.
```

Epiphytic Growth Analysis

Calculate mean/variance epiphytic growth ability

```r
# Replace NA with a lower value or remove them for visualization Remove rows
# with NA in CFU_per_10_leafdiscs_log_10
epi_growth_clean <- epi_growth_data %>%
    filter(!is.na(CFU_per_10_leafdiscs_log_10))

# Convert CFU_per_10_leafdiscs_log_10 to numeric, handling scientific notation
epi_growth_clean$CFU_per_10_leafdiscs_log_10 <- as.numeric(gsub("<", "", epi_growth_clean$CFU_per_10_lea
```

```
## Warning: NAs introduced by coercion
```

```r
# Subset data to remove unfinished strains and extract the letter part from the
# 'plant_rep' column
epi_growth_clean <- epi_growth_clean %>%
    mutate(plant_rep_letter = substr(plant_rep, 1, 1))

# Calculate the mean and SD CFU per strain for each plant rep letter and retain
# the specified columns
epi_growth_mean_per_rep <- epi_growth_clean %>%
```

```
    group_by(treatment, plant_rep_letter) %>%
    summarise(CFU_per_10_leafdiscs_log_10 = mean(CFU_per_10_leafdiscs_log_10, na.rm = TRUE),
        sd_CFU = sd(CFU_per_10_leafdiscs_log_10, na.rm = TRUE), sample = first(sample),
        person = first(person), species_complex = first(species_complex), block = first(block),
        treatment = first(treatment), plant_rep = first(plant_rep), .groups = "drop")

# Calculate the mean and SD CFU per strain for each plant rep letter and retain
# the specified columns
epi_growth_mean_per_strain <- epi_growth_clean %>%
    group_by(treatment) %>%
    summarise(CFU_per_10_leafdiscs_log_10 = mean(CFU_per_10_leafdiscs_log_10, na.rm = TRUE),
        sd_CFU = sd(CFU_per_10_leafdiscs_log_10, na.rm = TRUE), sample = first(sample),
        person = first(person), species_complex = first(species_complex), block = first(block),
        treatment = first(treatment), plant_rep = first(plant_rep), .groups = "drop")

epi_growth_mean_per_strain
```

```
## # A tibble: 13 x 8
##    treatment CFU_per_10_leafdiscs_l~1 sd_CFU sample person species_complex block
##    <chr>                        <dbl>  <dbl> <chr>  <chr>  <chr>           <chr>
##  1 194                           5.07     NA 194-A1 Havi   P. fluorescens  2-1
##  2 200                           5.26     NA 200-A1 Havi   P. fluorescens  2-6
##  3 204                           4.93     NA 204-A1 Havi   P. fluorescens  2-3
##  4 205                           4.87     NA 205-A1 Sara   P. fluorescens  3
##  5 215                           3        NA 215-A1 Sara   P. syringae     3
##  6 216                           3.59     NA 216-A1 Sara   P. putida       3
##  7 220                           4.82     NA 220-A1 Sara   P. fluorescens  5
##  8 221                           5.25     NA 221-A1 Havi   P. fluorescens  2-7
##  9 227                           4.16     NA 227-A1 Sara   P. fluorescens  5
## 10 228                           3.83     NA 228-A1 Sara   P. syringae     5
## 11 B728a                         3.72     NA B728a~ Sara   P. syringae     4
## 12 Cit7                          4.28     NA Cit7-~ Havi   P. syringae     2-3
## 13 pisi                          3.32     NA pisi-~ Havi   P. syringae     2-4
## # i abbreviated name: 1: CFU_per_10_leafdiscs_log_10
## # i 1 more variable: plant_rep <chr>
```

Plot Epiphytic Growth

```
# plot by treatment
epi_growth_treatment <- ggplot(data = epi_growth_mean_per_rep, aes(x = treatment,
    y = CFU_per_10_leafdiscs_log_10, color = species_complex, fill = species_complex)) +
    geom_boxplot(outlier.shape = NA, alpha = 0.3) + geom_beeswarm(stroke = 0.5, size = 0.8,
    alpha = 0.8) + theme_minimal() + coord_flip() + scale_fill_manual(values = species_colors) +
    scale_color_manual(values = species_colors) + scale_y_continuous(limits = c(2.5,
    5.5)) + theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 45,
    hjust = 1), legend.position = "bottom")

# plot by species complex
epi_growth_species <- ggplot(data = epi_growth_mean_per_rep, aes(x = species_complex,
    y = CFU_per_10_leafdiscs_log_10, color = species_complex, fill = species_complex)) +
    geom_boxplot(outlier.shape = NA, alpha = 0.3) + geom_beeswarm(stroke = 0.5, size = 0.8,
```

```
    alpha = 0.8) + theme_minimal() + scale_fill_manual(values = species_colors) +
    scale_color_manual(values = species_colors) + scale_y_continuous(limits = c(2.5,
    5.5)) + theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 45,
    hjust = 1), legend.position = "bottom")


# show plots
epi_growth_treatment
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```
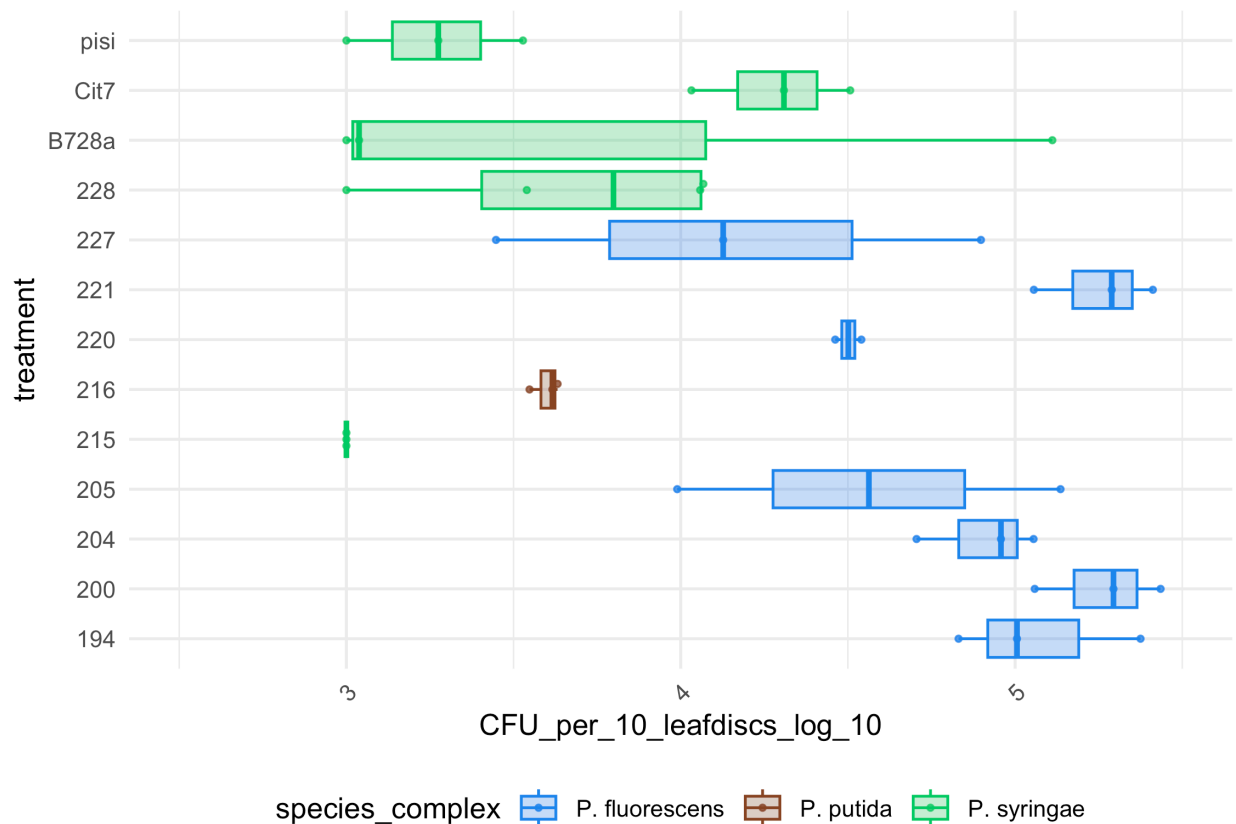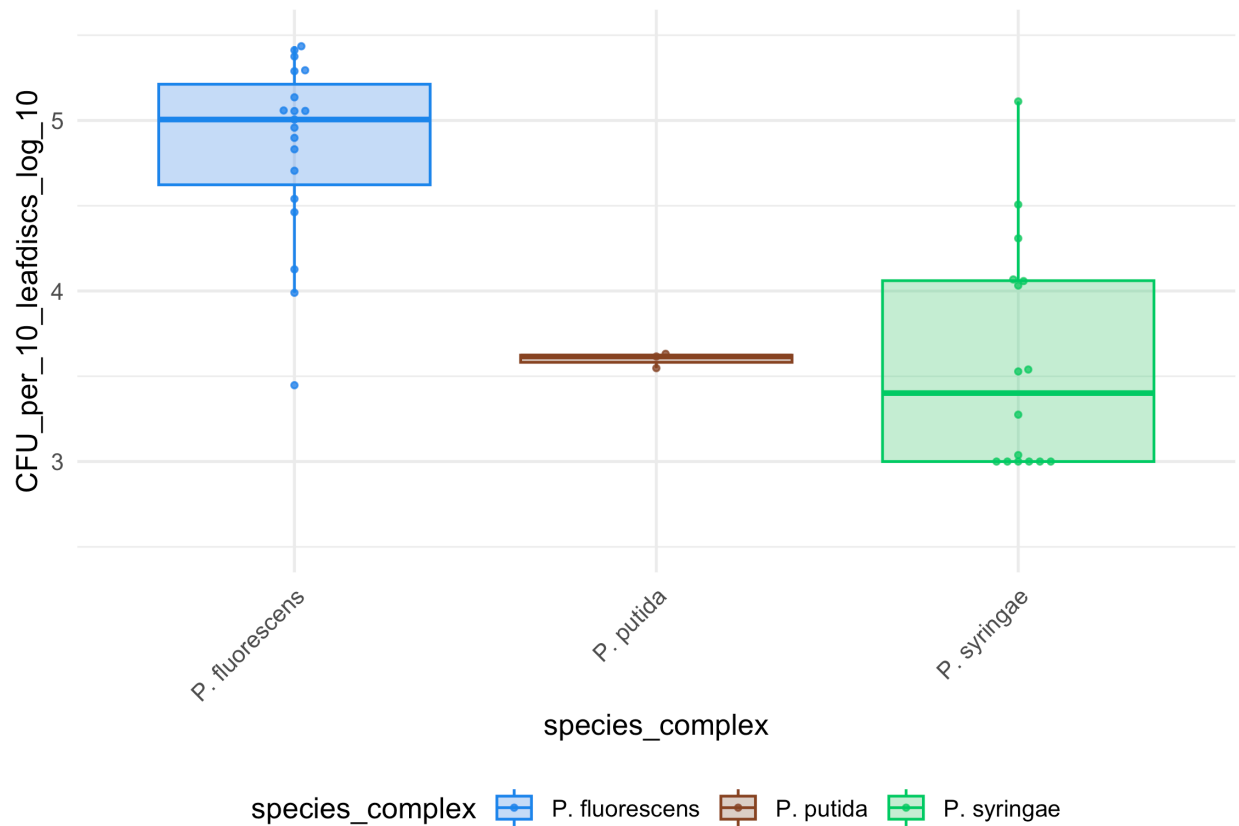


```
epi_growth_species
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
## Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Epiphytic growth stats

```
# stats per treatment
epi_growth_mod_treatment = lm(CFU_per_10_leafdiscs_log_10 ~ treatment, data = epi_growth_mean_per_rep)

emmeans(epi_growth_mod_treatment, pairwise ~ treatment)
```

```
## $emmeans
##  treatment emmean    SE df lower.CL upper.CL
##  194         5.07 0.305 27     4.45     5.70
##  200         5.26 0.305 27     4.64     5.89
##  204         4.91 0.305 27     4.28     5.53
##  205         4.90 0.305 27     4.28     5.53
##  215         3.00 0.305 27     2.37     3.63
##  216         3.60 0.305 27     2.97     4.22
##  220         4.87 0.305 27     4.24     5.49
##  221         5.25 0.305 27     4.63     5.88
##  227         4.16 0.305 27     3.53     4.78
##  228         3.67 0.264 27     3.13     4.21
##  B728a       3.72 0.305 27     3.09     4.34
##  Cit7        4.28 0.305 27     3.66     4.91
##  pisi        3.27 0.305 27     2.64     3.89
##
```

```
## Confidence level used: 0.95
##
## $contrasts
##  contrast     estimate     SE df t.ratio p.value
##  194 - 200    -0.19212 0.431 27  -0.446  1.0000
##  194 - 204     0.16452 0.431 27   0.382  1.0000
##  194 - 205     0.16752 0.431 27   0.389  1.0000
##  194 - 215     2.07066 0.431 27   4.807  0.0028
##  194 - 216     1.47217 0.431 27   3.417  0.0782
##  194 - 220     0.20476 0.431 27   0.475  1.0000
##  194 - 221    -0.18180 0.431 27  -0.422  1.0000
##  194 - 227     0.91321 0.431 27   2.120  0.6504
##  194 - 228     1.40429 0.403 27   3.485  0.0676
##  194 - B728a   1.35417 0.431 27   3.143  0.1373
##  194 - Cit7    0.78812 0.431 27   1.829  0.8188
##  194 - pisi    1.80298 0.431 27   4.185  0.0134
##  200 - 204     0.35664 0.431 27   0.828  0.9996
##  200 - 205     0.35964 0.431 27   0.835  0.9996
##  200 - 215     2.26278 0.431 27   5.253  0.0009
##  200 - 216     1.66429 0.431 27   3.863  0.0288
##  200 - 220     0.39688 0.431 27   0.921  0.9990
##  200 - 221     0.01032 0.431 27   0.024  1.0000
##  200 - 227     1.10533 0.431 27   2.566  0.3752
##  200 - 228     1.59641 0.403 27   3.962  0.0228
##  200 - B728a   1.54629 0.431 27   3.589  0.0537
##  200 - Cit7    0.98024 0.431 27   2.275  0.5509
##  200 - pisi    1.99510 0.431 27   4.631  0.0044
##  204 - 205     0.00299 0.431 27   0.007  1.0000
##  204 - 215     1.90614 0.431 27   4.425  0.0074
##  204 - 216     1.30765 0.431 27   3.035  0.1691
##  204 - 220     0.04024 0.431 27   0.093  1.0000
##  204 - 221    -0.34632 0.431 27  -0.804  0.9997
##  204 - 227     0.74869 0.431 27   1.738  0.8620
##  204 - 228     1.23976 0.403 27   3.077  0.1563
##  204 - B728a   1.18965 0.431 27   2.762  0.2756
##  204 - Cit7    0.62360 0.431 27   1.448  0.9559
##  204 - pisi    1.63845 0.431 27   3.803  0.0331
##  205 - 215     1.90314 0.431 27   4.418  0.0075
##  205 - 216     1.30465 0.431 27   3.029  0.1713
##  205 - 220     0.03724 0.431 27   0.086  1.0000
##  205 - 221    -0.34931 0.431 27  -0.811  0.9997
##  205 - 227     0.74569 0.431 27   1.731  0.8650
##  205 - 228     1.23677 0.403 27   3.069  0.1586
##  205 - B728a   1.18665 0.431 27   2.755  0.2788
##  205 - Cit7    0.62060 0.431 27   1.441  0.9574
##  205 - pisi    1.63546 0.431 27   3.796  0.0336
##  215 - 216    -0.59849 0.431 27  -1.389  0.9671
##  215 - 220    -1.86590 0.431 27  -4.331  0.0093
##  215 - 221    -2.25246 0.431 27  -5.229  0.0010
##  215 - 227    -1.15745 0.431 27  -2.687  0.3114
##  215 - 228    -0.66637 0.403 27  -1.654  0.8961
##  215 - B728a  -0.71649 0.431 27  -1.663  0.8925
##  215 - Cit7   -1.28254 0.431 27  -2.977  0.1886
##  215 - pisi   -0.26768 0.431 27  -0.621  1.0000
```

```
##  216 - 220    -1.26741 0.431 27   -2.942   0.2011
##  216 - 221    -1.65397 0.431 27   -3.839   0.0304
##  216 - 227    -0.55896 0.431 27   -1.298   0.9802
##  216 - 228    -0.06788 0.403 27   -0.168   1.0000
##  216 - B728a  -0.11800 0.431 27   -0.274   1.0000
##  216 - Cit7   -0.68405 0.431 27   -1.588   0.9188
##  216 - pisi    0.33081 0.431 27    0.768   0.9998
##  220 - 221    -0.38656 0.431 27   -0.897   0.9992
##  220 - 227     0.70845 0.431 27    1.645   0.8995
##  220 - 228     1.19953 0.403 27    2.977   0.1887
##  220 - B728a   1.14941 0.431 27    2.668   0.3208
##  220 - Cit7    0.58336 0.431 27    1.354   0.9727
##  220 - pisi    1.59822 0.431 27    3.710   0.0410
##  221 - 227     1.09501 0.431 27    2.542   0.3886
##  221 - 228     1.58608 0.403 27    3.936   0.0243
##  221 - B728a   1.53597 0.431 27    3.565   0.0567
##  221 - Cit7    0.96992 0.431 27    2.251   0.5662
##  221 - pisi    1.98477 0.431 27    4.607   0.0047
##  227 - 228     0.49108 0.403 27    1.219   0.9880
##  227 - B728a   0.44096 0.431 27    1.024   0.9973
##  227 - Cit7   -0.12509 0.431 27   -0.290   1.0000
##  227 - pisi    0.88977 0.431 27    2.065   0.6844
##  228 - B728a  -0.05012 0.403 27   -0.124   1.0000
##  228 - Cit7   -0.61617 0.403 27   -1.529   0.9362
##  228 - pisi    0.39869 0.403 27    0.989   0.9980
##  B728a - Cit7 -0.56605 0.431 27   -1.314   0.9782
##  B728a - pisi  0.44881 0.431 27    1.042   0.9969
##  Cit7 - pisi   1.01486 0.431 27    2.356   0.5000
##
## P value adjustment: tukey method for comparing a family of 13 estimates
```

```r
hist(resid(epi_growth_mod_treatment))
```

Histogram of resid(epi_growth_mod_treatment)
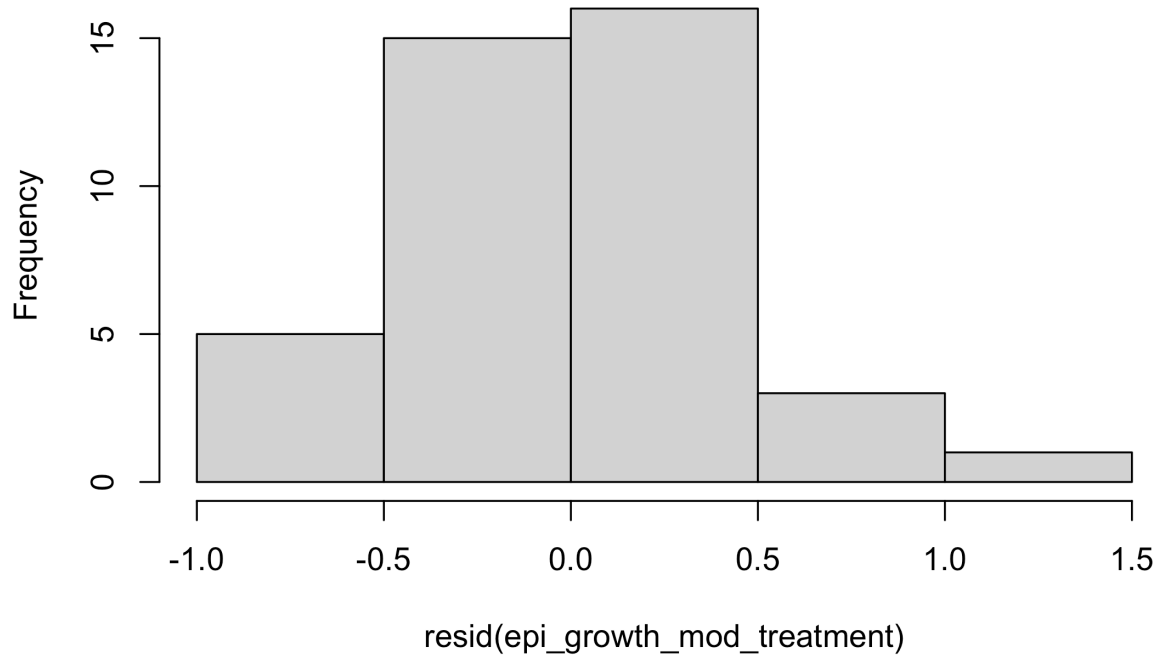

```r
# stats per species
epi_growth_mod_species = lm(CFU_per_10_leafdiscs_log_10 ~ species_complex, data = epi_growth_mean_per_re

emmeans(epi_growth_mod_species, pairwise ~ species_complex)
```

```
## $emmeans
##  species_complex emmean    SE df lower.CL upper.CL
##  P. fluorescens    4.92 0.129 37     4.66     5.18
##  P. putida         3.60 0.341 37     2.91     4.29
##  P. syringae       3.59 0.148 37     3.29     3.89
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast                    estimate    SE df t.ratio p.value
##  P. fluorescens - P. putida   1.31844 0.364 37   3.619  0.0025
##  P. fluorescens - P. syringae 1.32533 0.196 37   6.767  <.0001
##  P. putida - P. syringae      0.00689 0.371 37   0.019  0.9998
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```r
hist(resid(epi_growth_mod_species))
```

Histogram of resid(epi_growth_mod_species)

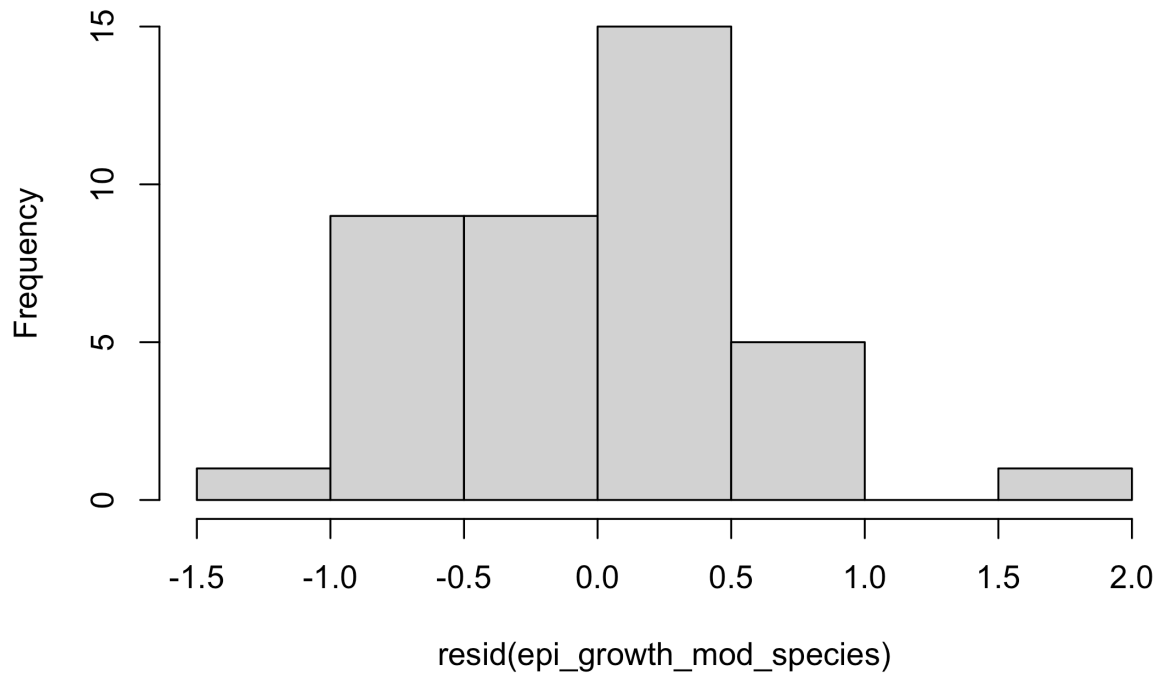## Combine epiphytic and virulence data

```
# Combine strain_stats and surv_probs using left_join
epi_virulence_data <- left_join(epi_growth_mean_per_strain, surv_probs, by = "treatment")

# Print the combined data
print(epi_virulence_data)
```

```
## # A tibble: 13 x 22
##    treatment CFU_per_10_leafdiscs~1 sd_CFU sample person species_complex.x block
##    <chr>                      <dbl>  <dbl> <chr>  <chr>  <chr>             <chr>
## 1  194                         5.07     NA 194-A1 Havi   P. fluorescens    2-1
## 2  200                         5.26     NA 200-A1 Havi   P. fluorescens    2-6
## 3  204                         4.93     NA 204-A1 Havi   P. fluorescens    2-3
## 4  205                         4.87     NA 205-A1 Sara   P. fluorescens    3
## 5  215                         3        NA 215-A1 Sara   P. syringae       3
## 6  216                         3.59     NA 216-A1 Sara   P. putida         3
## 7  220                         4.82     NA 220-A1 Sara   P. fluorescens    5
## 8  221                         5.25     NA 221-A1 Havi   P. fluorescens    2-7
## 9  227                         4.16     NA 227-A1 Sara   P. fluorescens    5
## 10 228                         3.83     NA 228-A1 Sara   P. syringae       5
## 11 B728a                       3.72     NA B728a~ Sara   P. syringae       4
## 12 Cit7                        4.28     NA Cit7-~ Havi   P. syringae       2-3
```

```
## 13 pisi                              3.32     NA pisi-~ Havi   P. syringae       2-4
## # i abbreviated name: 1: CFU_per_10_leafdiscs_log_10
## # i 15 more variables: plant_rep <chr>, time <dbl>, surv_prob <dbl>,
## #   std_error_data <dbl>, strain <chr>, `-80` <chr>, previous_work <chr>,
## #   species_complex.y <chr>, species_match <chr>, collector <chr>,
## #   source <chr>, ANI <chr>, genome_match <chr>, `Reseq/Genbank` <chr>,
## #   Paper <chr>
```

```r
# Calculate correlation coefficient between survival probability and epiphytic
# growth ability
correlation <- -1 * cor(epi_virulence_data$surv_prob, epi_virulence_data$CFU_per_10_leafdiscs_log_10,
    method = "pearson")
```

Plot epi virulence plot

```r
# Create scatter plot for 72 hours
epi_virulence_plot_strains <- ggplot(epi_virulence_data, aes(x = CFU_per_10_leafdiscs_log_10,
    y = 1 - surv_prob)) + geom_point(aes(color = species_complex.y)) + geom_smooth(method = "lm",
    se = TRUE, color = "black", fill = "black", alpha = 0.15, linewidth = 0.5) +
    labs(x = "Mean CFU per 10 Leaf Discs (Log Transformed)", y = "Aphid Mortality at 72 Hours (%)",
        caption = paste("Pearson's Correlation Coefficient:", round(correlation,
            2))) + scale_x_continuous(limits = c(2.5, 5.5)) + scale_fill_manual(values = species_colors)
    scale_color_manual(values = species_colors) + theme_minimal()

epi_virulence_plot_strains
```

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

```
## Warning: No shared levels found between `names(values)` of the manual scale and the
## data's fill values.
```

Pearson's Correlation Coefficient: 0.56
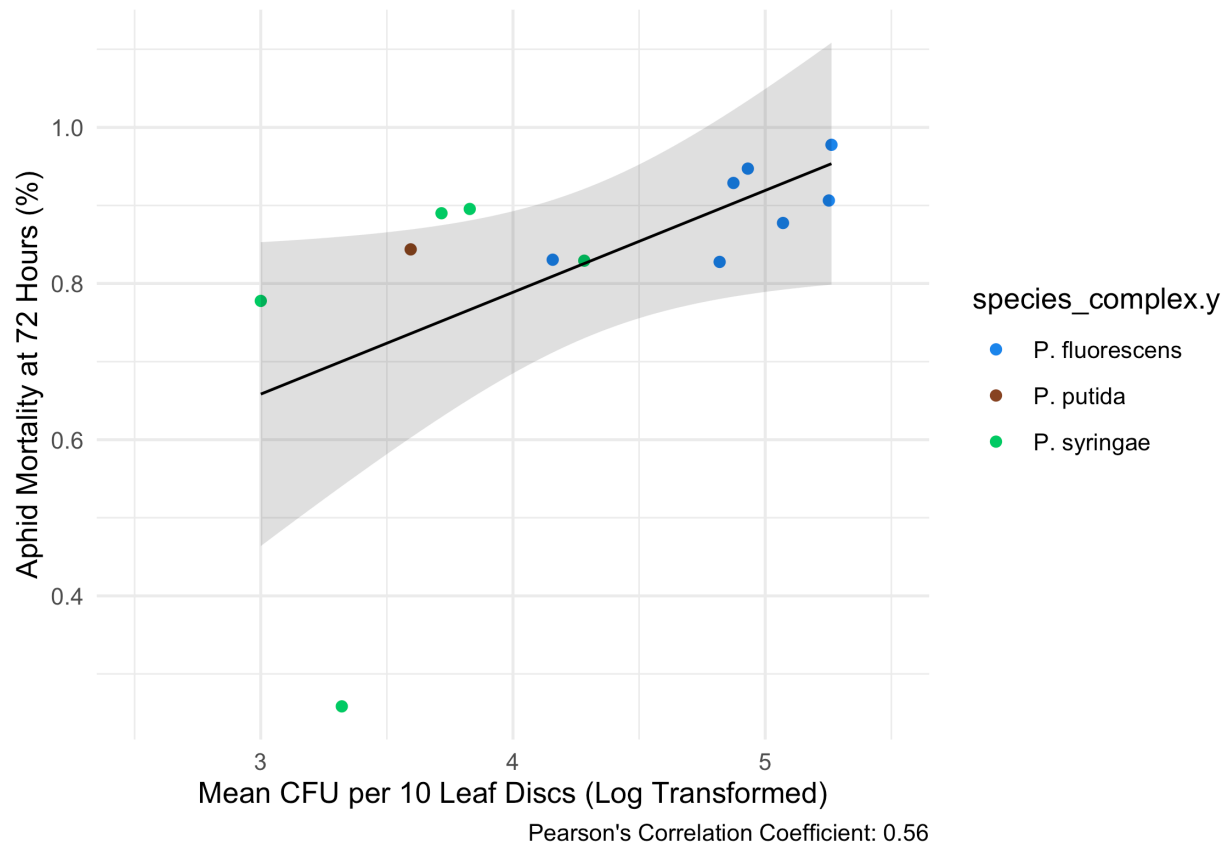
Session Information

```
devtools::session_info()
```

```
## ─ Session info ──────────────────────────────────────────────────────────
##  setting  value
##  version  R version 4.4.0 (2024-04-24)
##  os       macOS Ventura 13.4
##  system   x86_64, darwin20
##  ui       X11
##  language (EN)
##  collate  en_US.UTF-8
##  ctype    en_US.UTF-8
##  tz       America/New_York
##  date     2024-07-02
##  pandoc   3.1.11 @ /Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/x86_64/ (via rma
##
## ─ Packages ──────────────────────────────────────────────────────────────
##  ! package      * version date (UTC) lib source
##  P abind          1.4-5   2016-07-21 [?] RSPM
```

```
## P backports         1.4.1   2021-12-13 [?] CRAN (R 4.4.0)
## P beeswarm          0.4.0   2021-06-01 [?] CRAN (R 4.4.0)
## P bit               4.0.5   2022-11-15 [?] CRAN (R 4.4.0)
## P bit64             4.0.5   2020-08-30 [?] CRAN (R 4.4.0)
## P broom             1.0.6   2024-05-17 [?] CRAN (R 4.4.0)
## P broom.helpers     1.15.0  2024-04-05 [?] CRAN (R 4.4.0)
## P cachem            1.0.8   2023-05-01 [?] CRAN (R 4.4.0)
## P car               3.1-2   2023-03-30 [?] RSPM
## P carData           3.0-5   2022-01-06 [?] RSPM
## P cellranger        1.1.0   2016-07-27 [?] CRAN (R 4.4.0)
## P cli               3.6.2   2023-12-11 [?] CRAN (R 4.4.0)
## P colorspace        2.1-0   2023-01-23 [?] CRAN (R 4.4.0)
## P cowplot         * 1.1.3   2024-01-22 [?] CRAN (R 4.4.0)
## P crayon            1.5.2   2022-09-29 [?] CRAN (R 4.4.0)
## P data.table        1.15.4  2024-03-30 [?] CRAN (R 4.4.0)
## P devtools        * 2.4.5   2022-10-11 [?] RSPM
## P digest            0.6.35  2024-03-11 [?] CRAN (R 4.4.0)
## P dplyr           * 1.1.4   2023-11-17 [?] CRAN (R 4.4.0)
## P ellipsis          0.3.2   2021-04-29 [?] RSPM
## P emmeans         * 1.10.2  2024-05-20 [?] RSPM
## P estimability      1.5.1   2024-05-12 [?] RSPM
## P evaluate          0.23    2023-11-01 [?] CRAN (R 4.4.0)
## P fansi             1.0.6   2023-12-08 [?] CRAN (R 4.4.0)
## P farver            2.1.2   2024-05-13 [?] CRAN (R 4.4.0)
## P fastmap           1.1.1   2023-02-24 [?] CRAN (R 4.4.0)
## P forcats         * 1.0.0   2023-01-29 [?] CRAN (R 4.4.0)
## P formatR         * 1.14    2023-01-17 [?] RSPM
## P fs                1.6.4   2024-04-25 [?] CRAN (R 4.4.0)
## P generics          0.1.3   2022-07-05 [?] CRAN (R 4.4.0)
## P ggbeeswarm      * 0.7.2   2023-04-29 [?] CRAN (R 4.4.0)
## P ggplot2         * 3.5.1   2024-04-23 [?] CRAN (R 4.4.0)
## P ggpubr          * 0.6.0   2023-02-10 [?] RSPM
## P ggsignif          0.6.4   2022-10-13 [?] RSPM
## P ggsurvfit       * 1.1.0   2024-05-08 [?] CRAN (R 4.4.0)
## P glue              1.7.0   2024-01-09 [?] CRAN (R 4.4.0)
## P gridExtra       * 2.3     2017-09-09 [?] RSPM
## P gt                0.10.1  2024-01-17 [?] CRAN (R 4.4.0)
## P gtable            0.3.5   2024-04-22 [?] CRAN (R 4.4.0)
## P gtsummary       * 1.7.2   2023-07-15 [?] CRAN (R 4.4.0)
## P here            * 1.0.1   2020-12-13 [?] CRAN (R 4.4.0)
## P hms               1.1.3   2023-03-21 [?] CRAN (R 4.4.0)
## P htmltools         0.5.8.1 2024-04-04 [?] CRAN (R 4.4.0)
## P htmlwidgets       1.6.4   2023-12-06 [?] CRAN (R 4.4.0)
## P httpuv            1.6.15  2024-03-26 [?] RSPM
## P km.ci             0.5-6   2022-04-06 [?] RSPM
## P KMsurv            0.1-5   2012-12-03 [?] RSPM
## P knitr           * 1.46    2024-04-06 [?] CRAN (R 4.4.0)
## P labeling          0.4.3   2023-08-29 [?] CRAN (R 4.4.0)
## P later             1.3.2   2023-12-06 [?] RSPM
## P lattice           0.22-6  2024-03-20 [?] CRAN (R 4.4.0)
## P lifecycle         1.0.4   2023-11-07 [?] CRAN (R 4.4.0)
## P lubridate       * 1.9.3   2023-09-27 [?] CRAN (R 4.4.0)
## P magrittr          2.0.3   2022-03-30 [?] CRAN (R 4.4.0)
## P Matrix            1.7-0   2024-03-22 [?] CRAN (R 4.4.0)
```

```
##  P memoise        2.0.1   2021-11-26 [?] CRAN (R 4.4.0)
##  P mgcv           1.9-1   2023-12-21 [?] CRAN (R 4.4.0)
##  P mime           0.12    2021-09-28 [?] CRAN (R 4.4.0)
##  P miniUI         0.1.1.1 2018-05-18 [?] RSPM
##  P munsell        0.5.1   2024-04-01 [?] CRAN (R 4.4.0)
##  P mvtnorm        1.2-5   2024-05-21 [?] RSPM
##  P nlme           3.1-164 2023-11-27 [?] CRAN (R 4.4.0)
##  P pacman         0.5.1   2019-03-11 [?] CRAN (R 4.4.0)
##  P paletteer    * 1.6.0   2024-01-21 [?] RSPM
##  P pillar         1.9.0   2023-03-22 [?] CRAN (R 4.4.0)
##  P pkgbuild       1.4.4   2024-03-17 [?] RSPM
##  P pkgconfig      2.0.3   2019-09-22 [?] CRAN (R 4.4.0)
##  P pkgload        1.3.4   2024-01-16 [?] RSPM
##  P profvis        0.3.8   2023-05-02 [?] RSPM
##  P promises       1.3.0   2024-04-05 [?] RSPM
##  P purrr        * 1.0.2   2023-08-10 [?] CRAN (R 4.4.0)
##  P R6             2.5.1   2021-08-19 [?] CRAN (R 4.4.0)
##  P RColorBrewer * 1.1-3   2022-04-03 [?] CRAN (R 4.4.0)
##  P Rcpp           1.0.12  2024-01-09 [?] CRAN (R 4.4.0)
##  P readr        * 2.1.5   2024-01-10 [?] CRAN (R 4.4.0)
##  P readxl       * 1.4.3   2023-07-06 [?] CRAN (R 4.4.0)
##  P rematch2       2.1.2   2020-05-01 [?] CRAN (R 4.4.0)
##  P remotes        2.5.0   2024-03-17 [?] CRAN (R 4.4.0)
##    renv           1.0.7   2024-04-11 [1] CRAN (R 4.4.0)
##  P rlang          1.1.3   2024-01-10 [?] CRAN (R 4.4.0)
##  P rmarkdown      2.26    2024-03-05 [?] CRAN (R 4.4.0)
##  P rprojroot      2.0.4   2023-11-05 [?] CRAN (R 4.4.0)
##  P rstatix        0.7.2   2023-02-01 [?] RSPM
##  P rstudioapi     0.16.0  2024-03-24 [?] CRAN (R 4.4.0)
##  P scales         1.3.0   2023-11-28 [?] CRAN (R 4.4.0)
##  P sessioninfo    1.2.2   2021-12-06 [?] RSPM
##  P shiny          1.8.1.1 2024-04-02 [?] RSPM
##  P stringi        1.8.3   2023-12-11 [?] CRAN (R 4.4.0)
##  P stringr      * 1.5.1   2023-11-14 [?] CRAN (R 4.4.0)
##  P survival     * 3.6-4   2024-04-24 [?] CRAN (R 4.4.0)
##  P survminer    * 0.4.9   2021-03-09 [?] RSPM
##  P survMisc       0.5.6   2022-04-07 [?] RSPM
##  P tibble       * 3.2.1   2023-03-20 [?] CRAN (R 4.4.0)
##  P tidycmprsk   * 1.0.0   2023-10-30 [?] CRAN (R 4.4.0)
##  P tidyr        * 1.3.1   2024-01-24 [?] CRAN (R 4.4.0)
##  P tidyselect     1.2.1   2024-03-11 [?] CRAN (R 4.4.0)
##  P tidyverse    * 2.0.0   2023-02-22 [?] CRAN (R 4.4.0)
##  P timechange     0.3.0   2024-01-18 [?] CRAN (R 4.4.0)
##  P tzdb           0.4.0   2023-05-12 [?] CRAN (R 4.4.0)
##  P urlchecker     1.0.1   2021-11-30 [?] RSPM
##  P usethis      * 2.2.3   2024-02-19 [?] RSPM
##  P utf8           1.2.4   2023-10-22 [?] CRAN (R 4.4.0)
##  P vctrs          0.6.5   2023-12-01 [?] CRAN (R 4.4.0)
##  P vipor          0.4.7   2023-12-18 [?] CRAN (R 4.4.0)
##  P viridis      * 0.6.5   2024-01-29 [?] RSPM
##  P viridisLite  * 0.4.2   2023-05-02 [?] CRAN (R 4.4.0)
##  P vroom          1.6.5   2023-12-05 [?] CRAN (R 4.4.0)
##  P wesanderson  * 0.3.7   2023-10-31 [?] RSPM
##  P withr          3.0.0   2024-01-16 [?] CRAN (R 4.4.0)
```

```
##  P xfun              0.43   2024-03-25 [?] CRAN (R 4.4.0)
##  P xml2              1.3.6  2023-12-04 [?] CRAN (R 4.4.0)
##  P xtable            1.8-4  2019-04-21 [?] RSPM
##  P yaml              2.3.8  2023-12-11 [?] CRAN (R 4.4.0)
##  P zoo               1.8-12 2023-04-13 [?] RSPM
##
##  [1] /Users/zahavahrojer/Desktop/Cornell/Hendry Lab/pseud-epi-growth/renv/library/macos/R-4.4/x86_64-
##  [2] /Users/zahavahrojer/Library/Caches/org.R-project.R/R/renv/sandbox/macos/R-4.4/x86_64-apple-darw:
##
##  P -- Loaded and on-disk path mismatch.
##
##  -----------------------------------------------------------------------------
```