Wild Type Investigations

Abigail Seeger

1/18/2021

Contents

The purpose of this document is to include all information for wild type in one place. This is because the log two fold change values seem opposite for many genotypes depending on the experiment - either December, January, or February. One potential reason for this is that the Col0 plants may have started at different sizes.

```
library(knitr)
library(dplyr)
library(ggplot2)
library(viridis)
library(lemon)
library(ggthemes)
library(extrafont)
library(Rfast)
library(formatR)
```

Load in the Data

I will use the data that has not been quantile normalized, but the outliers have been removed.

Here is the data frame that I am using. This has already been cleaned, and I saved the data in a .csv file:

```
depi_data <- read.table("C:/Users/Owner/Documents/Research/Shiu_Lab/Shiu_Lab_R/Data/Clean_DEPI_Data.csv
    sep = ",", header = TRUE)
head(depi_data)
```

```
##   individual_plant_metadata genotype flat_number measurement time_point
## 1     1217_F_DEPI_SC_1_10_6     Col0           1    leafarea          0
## 2     1217_F_DEPI_SC_1_10_6     Col0           1    leafarea          1
## 3     1217_F_DEPI_SC_1_10_6     Col0           1    leafarea          2
## 4     1217_F_DEPI_SC_1_10_6     Col0           1    leafarea          3
## 5     1217_F_DEPI_SC_1_10_6     Col0           1    leafarea          4
## 6     1217_F_DEPI_SC_1_10_6     Col0           1    leafarea          5
##   measured_value border subline full_subline_information month day
## 1            133  FALSE       4                   Col1-4   Dec   1
## 2            135  FALSE       4                   Col1-4   Dec   1
## 3            135  FALSE       4                   Col1-4   Dec   1
## 4            142  FALSE       4                   Col1-4   Dec   1
## 5            143  FALSE       4                   Col1-4   Dec   1
## 6            139  FALSE       4                   Col1-4   Dec   1
##   normalized_value
## 1         162.9944
## 2         164.2500
## 3         164.2500
## 4         169.2500
## 5         170.0000
## 6         167.2500
```

Here are the columns in the data frame:

```
colnames(depi_data)
```

```
##  [1] "individual_plant_metadata" "genotype"
##  [3] "flat_number"               "measurement"
##  [5] "time_point"                "measured_value"
##  [7] "border"                    "subline"
##  [9] "full_subline_information"  "month"
## [11] "day"                       "normalized_value"
```

Note that moving forward, I will use the measured_value column, not the normalized value column.

Initial Information

Here are the number of Col0 plants per experiment:

```
depi_data %>% group_by(month) %>% filter(genotype ==
    "Col0") %>% summarize(Number_Of_Plants = length(unique(individual_plant_metadata)))
```

```
## # A tibble: 3 x 2
##   month Number_Of_Plants
##   <chr>            <int>
## 1 Dec                 13
## 2 Feb                 16
## 3 Jan                 13
```

Here are the summary statistics for Col0 at the start of each experiment:

```
depi_data %>% group_by(month, measurement) %>%
    filter(genotype == "Col0", time_point ==
        "0") %>% summarize(Mean = mean(measured_value),
    Median = median(measured_value), Standard_Deviation = sd(measured_value),
    Max = max(measured_value), Min = min(measured_value)) %>%
    arrange(measurement)
```

```
## # A tibble: 9 x 7
## # Groups:   month [3]
##    month measurement   Mean  Median Standard_Deviation    Max     Min
##    <chr> <chr>        <dbl>   <dbl>              <dbl>  <dbl>   <dbl>
## 1 Dec   leafarea     172.   175                  39.3  239    107
## 2 Feb   leafarea     240.   246.                 46.6  338    167
## 3 Jan   leafarea     208.   223                  41.4  268    148
## 4 Dec   npq           0.489   0.490              0.0439  0.540  0.366
## 5 Feb   npq           0.597   0.595              0.0464  0.699  0.539
## 6 Jan   npq           0.508   0.504              0.0504  0.585  0.413
## 7 Dec   phi2          0.751   0.753              0.0139  0.782  0.731
## 8 Feb   phi2          0.681   0.678              0.0174  0.712  0.651
## 9 Jan   phi2          0.700   0.703              0.0464  0.776  0.596
```

Here are the summary statistics for all other genotypes:

```
depi_data %>% group_by(month, measurement) %>%
    filter(genotype != "Col0", time_point ==
        "0") %>% summarize(Mean = mean(measured_value),
    Median = median(measured_value), Standard_Deviation = sd(measured_value),
    Max = max(measured_value), Min = min(measured_value)) %>%
    arrange(measurement)
```

```
## # A tibble: 9 x 7
## # Groups:   month [3]
##    month measurement   Mean  Median Standard_Deviation    Max     Min
##    <chr> <chr>        <dbl>   <dbl>              <dbl>  <dbl>   <dbl>
## 1 Dec   leafarea     177.   177                  44.0  302    21
## 2 Feb   leafarea     232.   236                  51.6  408     0
## 3 Jan   leafarea     176.   183                  61.8  314     0
## 4 Dec   npq           0.494   0.491              0.0630  1.46   0.369
## 5 Feb   npq           0.620   0.615              0.0517  0.812  0.296
## 6 Jan   npq           0.555   0.539              0.146   1.54   0.186
## 7 Dec   phi2          0.756   0.758              0.0250  0.819  0.499
## 8 Feb   phi2          0.670   0.676              0.0538  0.724  0
## 9 Jan   phi2          0.692   0.711              0.0960  0.780  0
```

```
plot_data <- depi_data %>% filter(time_point ==
    0)
```

```
plot_data$Classification <- ifelse(plot_data$genotype ==
    "Col0", "Col0", "Other")
```

```
ggplot(data = plot_data, aes(x = genotype,
    y = measured_value, fill = Classification)) +
```

```
geom_boxplot() + facet_grid(measurement ~
month, scales = "free") + labs(x = "Genotype",
y = "Measured Value") + theme(axis.ticks.x = element_blank(),
axis.text.x = element_blank(), axis.title.x = element_blank()) +
scale_fill_viridis_d(option = "D")
```
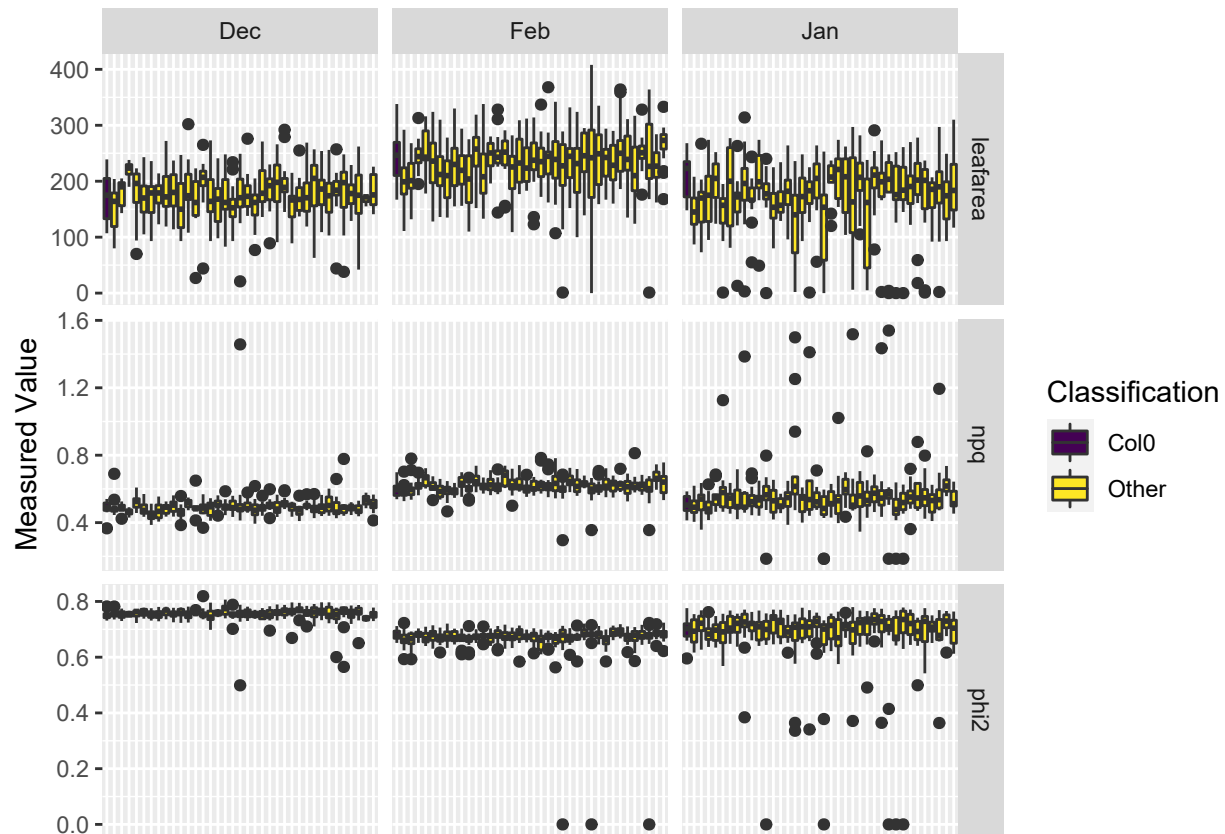

Melissa noted that the behavior of mpk16 is different depending on the month of the experiment.

Create plot data:

```
plot_data_npq <- depi_data %>% filter(genotype %in%
    c("Col0", "mpk16")) %>% group_by(time_point,
    month, genotype) %>% filter(measurement ==
    "npq") %>% mutate(mean = mean(measured_value),
    sd = sd(measured_value), month = factor(month,
        levels = c("Dec", "Jan", "Feb"))) %>%
    mutate(upper = mean + sd, lower = mean -
        sd)


plot_data_phi2 <- depi_data %>% filter(genotype %in%
    c("Col0", "mpk16")) %>% group_by(time_point,
    month, genotype) %>% filter(measurement ==
    "phi2") %>% mutate(mean = mean(measured_value),
    sd = sd(measured_value), month = factor(month,
        levels = c("Dec", "Jan", "Feb"))) %>%
```

```
    mutate(upper = mean + sd, lower = mean -
        sd)

plot_data_leafarea <- depi_data %>% filter(genotype %in%
    c("Col0", "mpk16")) %>% group_by(time_point,
    month, genotype) %>% filter(measurement ==
    "leafarea") %>% mutate(mean = mean(measured_value),
    sd = sd(measured_value), month = factor(month,
        levels = c("Dec", "Jan", "Feb"))) %>%
    mutate(upper = mean + sd, lower = mean -
        sd)
```
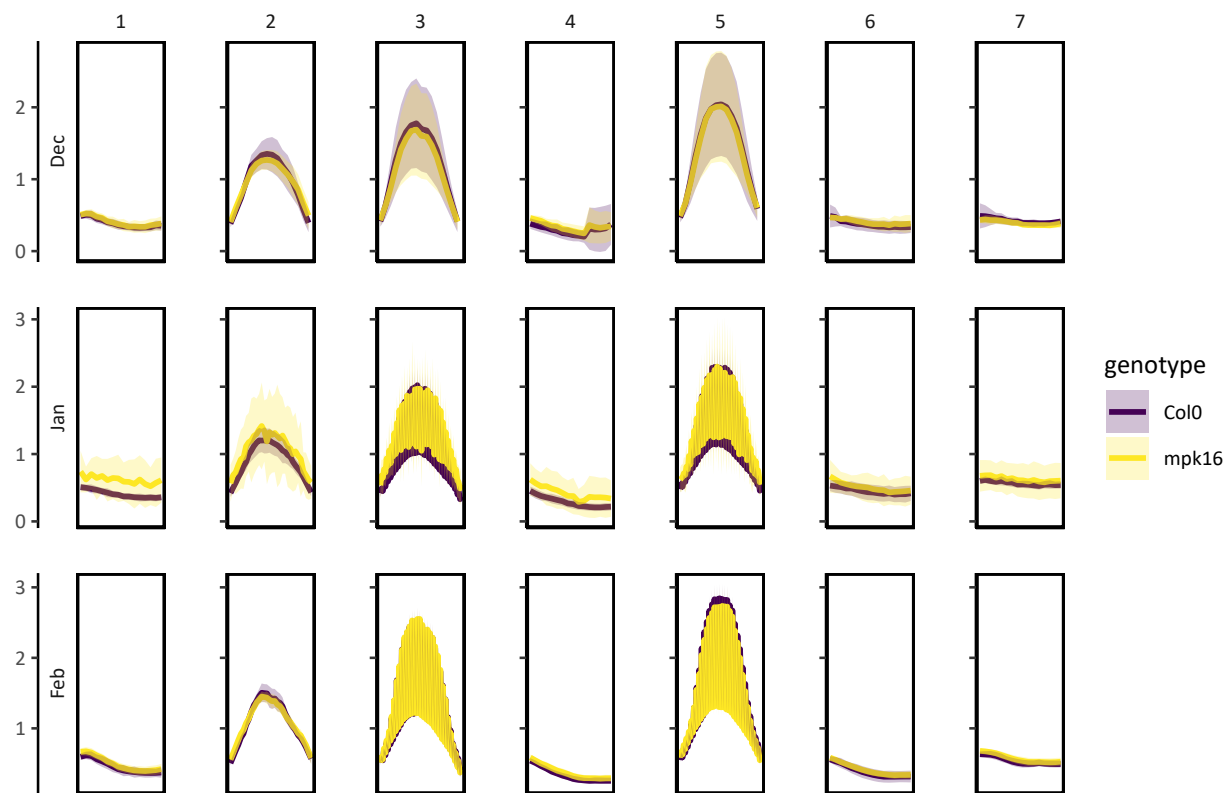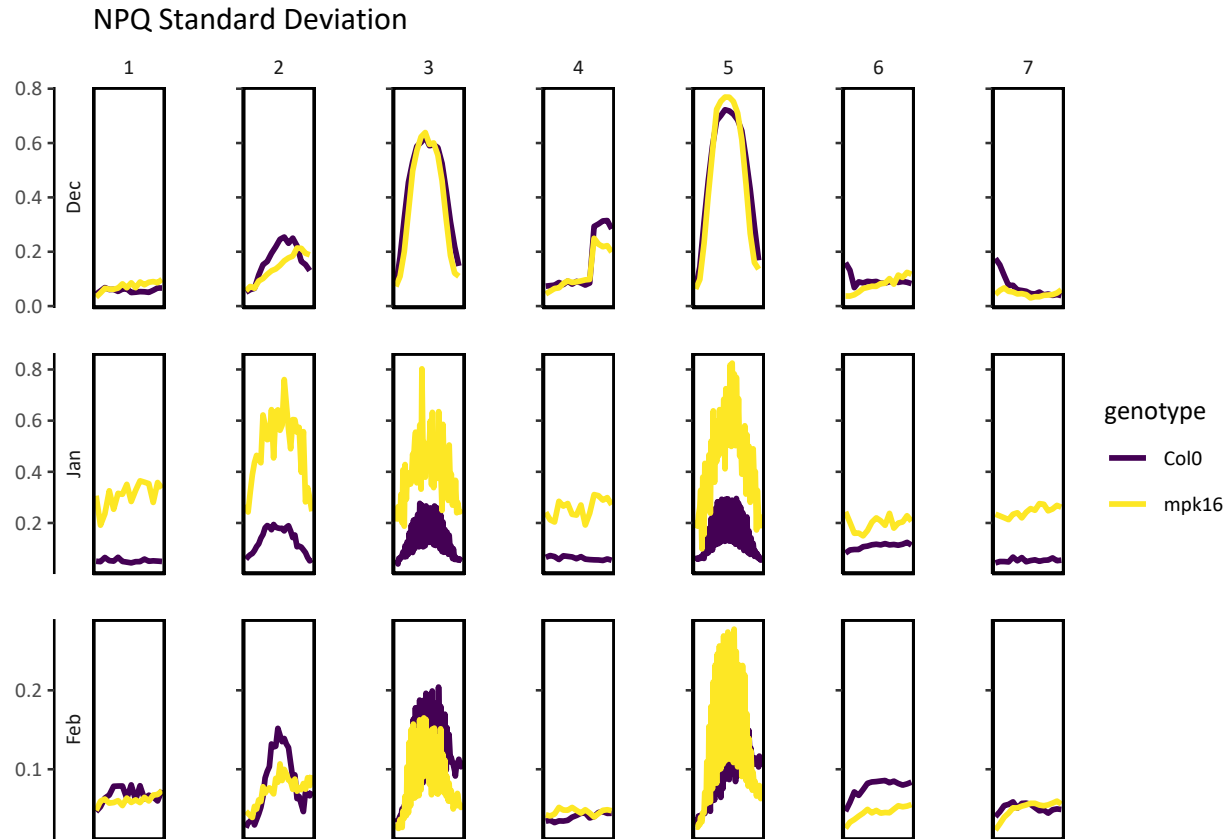
NPQ

```
ggplot(data = plot_data_npq, aes(x = time_point,
    y = mean)) + geom_line(aes(color = genotype),
    size = 1) + facet_rep_grid(month ~ day,
    scales = "free", switch = "y", repeat.tick.labels = FALSE) +
    labs(x = "Hours", y = NULL, title = "Mean NPQ, with Standard Deviation") +
    theme_tufte(base_family = "Calibri",
        base_size = 10) + geom_ribbon(aes(ymin = lower,
    ymax = upper, fill = genotype), alpha = 0.25) +
    # geom_errorbar(aes(ymin = mean - sd,
# ymax = mean + sd), width = .2)+
theme(strip.background.x = element_blank(),
    axis.title.x = element_blank(), axis.text.x = element_blank(),
    axis.ticks.x = element_blank(), panel.border = element_rect(color = "black",
        fill = NA, size = 1), axis.line = element_line(),
    panel.spacing = unit(1, "lines")) + scale_color_viridis_d(begin = 0,
    end = 1, option = "viridis", aesthetics = c("colour",
        "fill"))
```

Mean NPQ, with Standard Deviation


```
ggplot(data = plot_data_npq, aes(x = time_point,
    y = sd)) + geom_line(aes(color = genotype),
    size = 1) + facet_rep_grid(month ~ day,
    scales = "free", switch = "y", repeat.tick.labels = FALSE) +
    labs(x = "Hours", y = NULL, title = "NPQ Standard Deviation") +
    theme_tufte(base_family = "Calibri",
        base_size = 10) + theme(strip.background.x = element_blank(),
    axis.title.x = element_blank(), axis.text.x = element_blank(),
    axis.ticks.x = element_blank(), panel.border = element_rect(color = "black",
        fill = NA, size = 1), axis.line = element_line(),
    panel.spacing = unit(1, "lines")) + scale_color_viridis_d(begin = 0,
    end = 1, option = "viridis", aesthetics = c("colour",
        "fill"))
```

NPQ Standard Deviation

There are some patterns here:

- Generally, mpk16 is more variable than Col0 in January, but the genotypes have roughly the same variability in the other experiments

- Flucuating and sinusoidal light days have higher variability

- February has less variability than both December and January

Phi2

```
ggplot(data = plot_data_phi2, aes(x = time_point,
    y = mean)) + geom_line(aes(color = genotype),
    size = 1) + facet_rep_grid(month ~ day,
    scales = "free", switch = "y", repeat.tick.labels = FALSE) +
    labs(x = "Hours", y = NULL, title = "Mean Phi2, with Standard Deviation") +
    theme_tufte(base_family = "Calibri",
        base_size = 10) + geom_ribbon(aes(ymin = lower,
    ymax = upper, fill = genotype), alpha = 0.25) +
    # geom_errorbar(aes(ymin = mean - sd,
# ymax = mean + sd), width = .2)+
theme(strip.background.x = element_blank(),
    axis.title.x = element_blank(), axis.text.x = element_blank(),
    axis.ticks.x = element_blank(), panel.border = element_rect(color = "black",
        fill = NA, size = 1), axis.line = element_line(),
```

```
    panel.spacing = unit(1, "lines")) + scale_color_viridis_d(begin = 0,
    end = 1, option = "viridis", aesthetics = c("colour",
        "fill"))
```
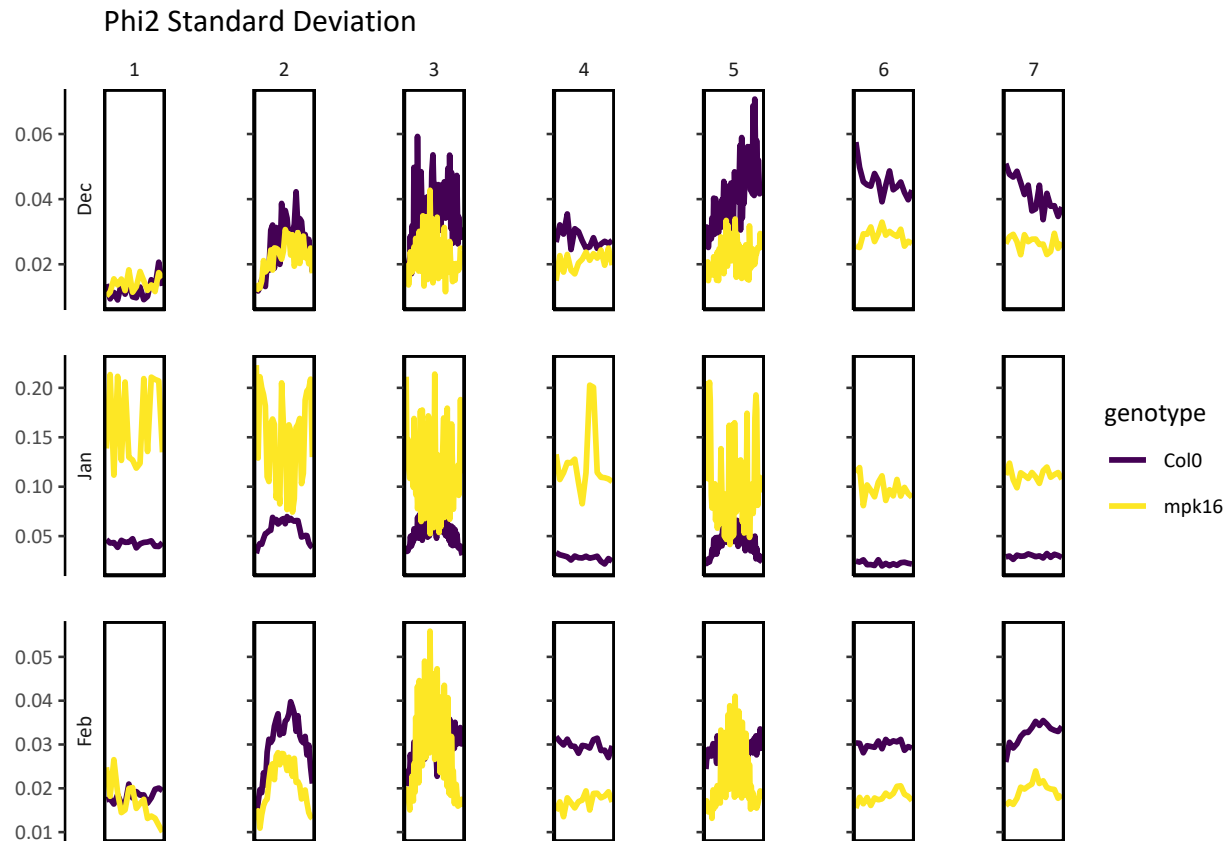

Mean Phi2, with Standard Deviation
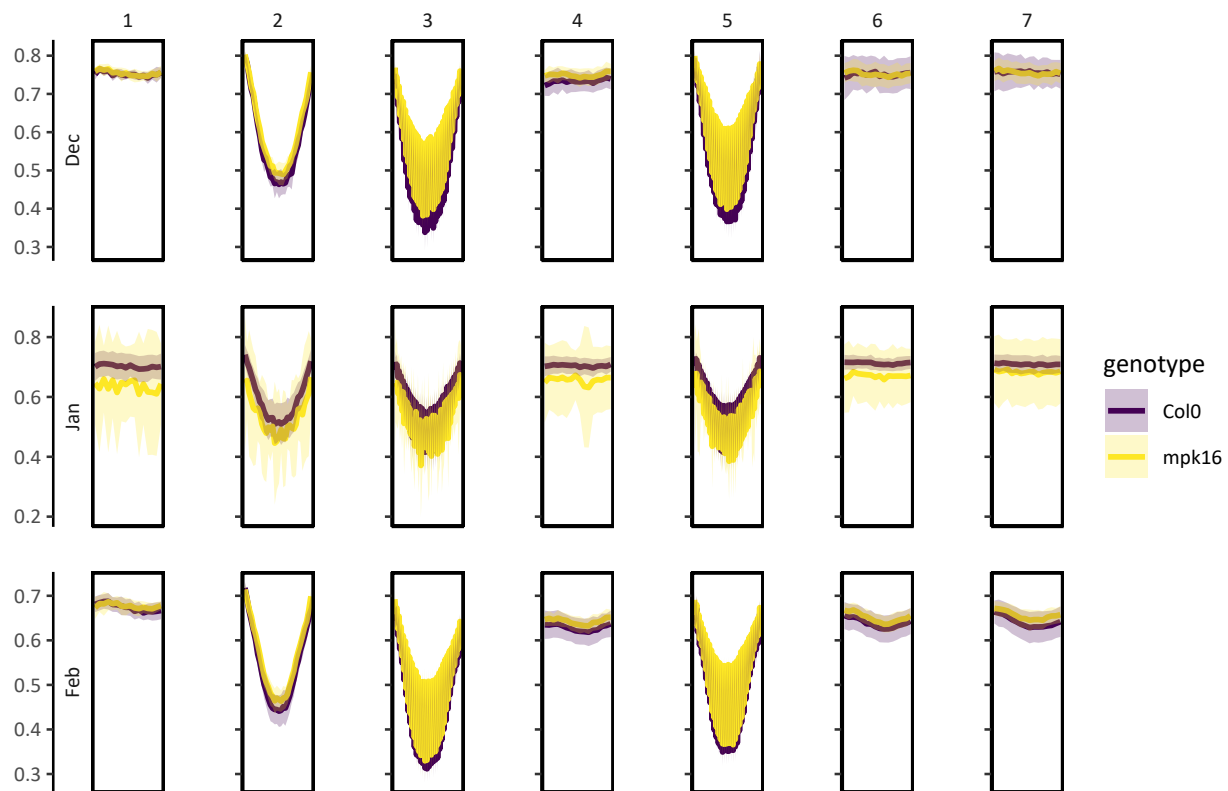
```
ggplot(data = plot_data_phi2, aes(x = time_point,
    y = sd)) + geom_line(aes(color = genotype),
    size = 1) + facet_rep_grid(month ~ day,
    scales = "free", switch = "y", repeat.tick.labels = FALSE) +
    labs(x = "Hours", y = NULL, title = "Phi2 Standard Deviation") +
    theme_tufte(base_family = "Calibri",
        base_size = 10) + theme(strip.background.x = element_blank(),
    axis.title.x = element_blank(), axis.text.x = element_blank(),
    axis.ticks.x = element_blank(), panel.border = element_rect(color = "black",
        fill = NA, size = 1), axis.line = element_line(),
    panel.spacing = unit(1, "lines")) + scale_color_viridis_d(begin = 0,
    end = 1, option = "viridis", aesthetics = c("colour",
        "fill"))
```
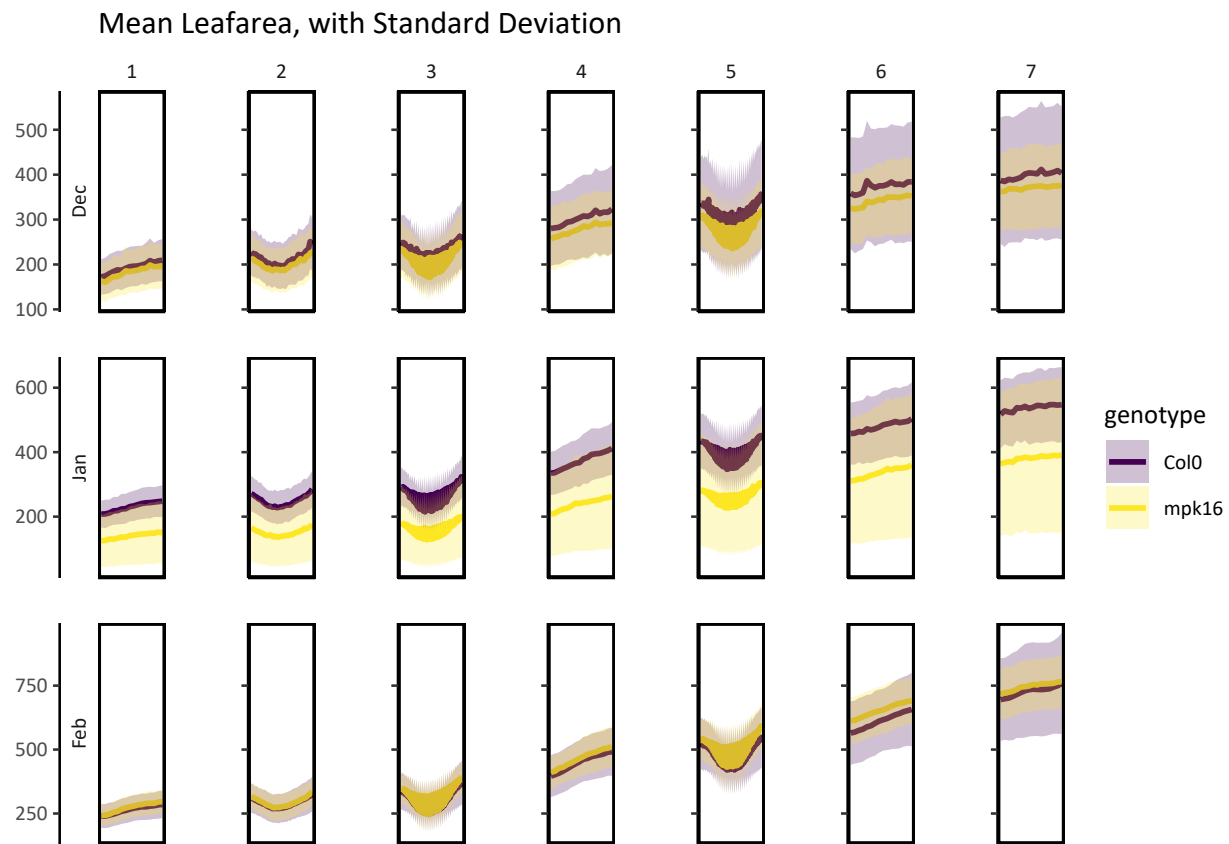
Phi2 Standard Deviation

Some observations

- January has much, much higher standard deviation than the other two months

- For December and Februrary, mpk16 has less variability than Col0

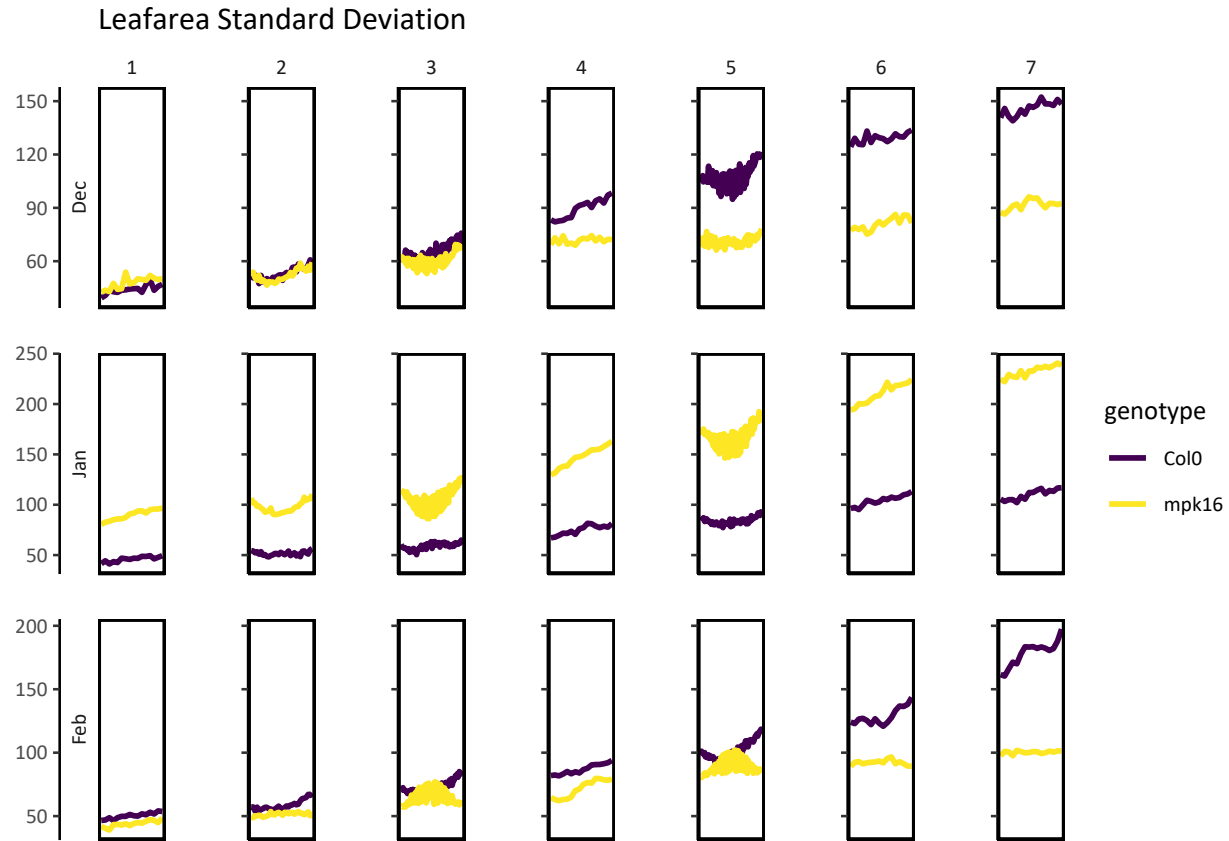- For January, mpk16 has more variability than Col0

leafarea

```
ggplot(data = plot_data_leafarea, aes(x = time_point,
    y = mean)) + geom_line(aes(color = genotype),
    size = 1) + facet_rep_grid(month ~ day,
    scales = "free", switch = "y", repeat.tick.labels = FALSE) +
    labs(x = "Hours", y = NULL, title = "Mean Leafarea, with Standard Deviation") +
    theme_tufte(base_family = "Calibri",
        base_size = 10) + geom_ribbon(aes(ymin = lower,
    ymax = upper, fill = genotype), alpha = 0.25) +
    # geom_errorbar(aes(ymin = mean - sd,
# ymax = mean + sd), width = .2)+
theme(strip.background.x = element_blank(),
    axis.title.x = element_blank(), axis.text.x = element_blank(),
    axis.ticks.x = element_blank(), panel.border = element_rect(color = "black",
        fill = NA, size = 1), axis.line = element_line(),
    panel.spacing = unit(1, "lines")) + scale_color_viridis_d(begin = 0,
```

```
end = 1, option = "viridis", aesthetics = c("colour",
    "fill"))
```

Mean Leafarea, with Standard Deviation


```
ggplot(data = plot_data_leafarea, aes(x = time_point,
    y = sd)) + geom_line(aes(color = genotype),
    size = 1) + facet_rep_grid(month ~ day,
    scales = "free", switch = "y", repeat.tick.labels = FALSE) +
    labs(x = "Hours", y = NULL, title = "Leafarea Standard Deviation") +
    theme_tufte(base_family = "Calibri",
        base_size = 10) + theme(strip.background.x = element_blank(),
    axis.title.x = element_blank(), axis.text.x = element_blank(),
    axis.ticks.x = element_blank(), panel.border = element_rect(color = "black",
        fill = NA, size = 1), axis.line = element_line(),
    panel.spacing = unit(1, "lines")) + scale_color_viridis_d(begin = 0,
    end = 1, option = "viridis", aesthetics = c("colour",
        "fill"))
```

Leafarea Standard Deviation

Some observations:

- But, for day 7, December and February have more variability in Col0, whereas for January, mpk16 has much more variability than Col0

- The plants are much bigger in January

Minimum leafarea sizes for January

```
depi_data %>% filter(month == "Jan", measurement ==
    "leafarea", time_point == "0") %>% group_by(genotype) %>%
    summarize(Min = min(measured_value),
        Second_Lowest = nth(measured_value,
            2, descending = FALSE), Third_Lowest = nth(measured_value,
            3)) %>% print(n = 38)
```

```
## # A tibble: 38 x 4
##    genotype   Min Second_Lowest Third_Lowest
##    <chr>    <dbl>         <dbl>        <dbl>
## 1 Col0       148           149          161
## 2 mpk1        87           105          120
## 3 mpk1-13     73            90          118
## 4 mpk1-14     95           124          125
```

```
##  5 mpk1-16     124         161         168
##  6 mpk1-17       1          93         104
##  7 mpk1-18      81          98         105
##  8 mpk1-3       13         121         143
##  9 mpk13         3         123         148
## 10 mpk13-20     55         126         150
## 11 mpk14        49         166         171
## 12 mpk14-16      0         153         165
## 13 mpk14-17     80         111         125
## 14 mpk14-18    129         135         140
## 15 mpk14-20     96         121         152
## 16 mpk16         2           3          22
## 17 mpk17        93          94         135
## 18 mpk17-20      1         127         147
## 19 mpk18        56         148         154
## 20 mpk18-20      0           2          55
## 21 mpk20       120         142         198
## 22 mpk3        154         159         168
## 23 mpk3-16     104         119         119
## 24 mpk5          6          35          51
## 25 mpk5-16     105         120         137
## 26 mpk5-17       5          11          44
## 27 mpk5-6       78         139         145
## 28 mpk6          2         166         170
## 29 mpk6-18       0           4         179
## 30 mpk6-20       0         126         137
## 31 mpk6-8        0         145         159
## 32 mpk6-9      122         129         145
## 33 mpk8         18          59         161
## 34 mpk8-17       1           5         126
## 35 mpk8-20      92         117         142
## 36 mpk9          2          92         152
## 37 mpk9-16      93         108         117
## 38 mpk9-18     117         121         123
```

For January, many genotypes have very small plants. Why were these not removed when we looked for outliers? Maybe our outlier threshold is too low?