Variance Mesocosm Visualisation

2025-08-04

```r
# Load packages
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.3
```

```r
library(nlme)
```

```
## Warning: package 'nlme' was built under R version 4.3.3
```

```r
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.3.3
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.3.3
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
##
##     lmList
```

```r
library(DAAG)
```

```
## Warning: package 'DAAG' was built under R version 4.3.3
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:nlme':
##
##     collapse

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(MuMIn)
```

```
## Warning: package 'MuMIn' was built under R version 4.3.3
```

```r
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.3
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:DAAG':
##
##     vif
```

```r
library(viridis)
```

```
## Warning: package 'viridis' was built under R version 4.3.3
```

```
## Loading required package: viridisLite
```

```r
library(DHARMa)
```

```
## Warning: package 'DHARMa' was built under R version 4.3.3
```

```
## This is DHARMa 0.4.7. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa')
```

```r
library(FSA)
```

```
## Warning: package 'FSA' was built under R version 4.3.3
```

```
## Registered S3 methods overwritten by 'FSA':
##   method      from
##   confint.boot car
##   hist.boot    car
```

```
## ## FSA v0.9.6. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.
```

```
##
## Attaching package: 'FSA'
```

```
## The following object is masked from 'package:car':
##
##     bootCase
```

```r
library(emmeans)
```

```
## Warning: package 'emmeans' was built under R version 4.3.3
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.3.3
```

```r
library(performance)
library(cowplot)
```

```
## Warning: package 'cowplot' was built under R version 4.3.3
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggpubr':
##
##     get_legend
```

Load in the the datasets

```r
#Import the data
data <- read_excel(file.path(
  "C:/Users/Joyalea/Documents/UBCO/Thesis",
  "Final_Files",
  "RQ1_AMFAbundance",
  "Variance_Longform_M24_M25.xlsx"
))

#Clean and format columns
data <- data %>%
  mutate(
    sample_group = gsub("[A-Z]$", "", sample_id),
    mesocosm_id = case_when(
      sample_group %in% c("Br70", "Br71", "Br72", "Ga70", "Ga71",
                          "Ga72", "Fe70", "Fe71", "Fe72", "Ta70", "Ta71",
                          "Ta72") ~ 24,
      sample_group %in% c("Br73", "Br74", "Br75", "Ga73", "Ga74", "Ga75",
                          "Fe73", "Fe74", "Fe75", "Ta73", "Ta74", "Ta75") ~ 25,
      TRUE ~ NA_real_
    ),
    sample_id = as.factor(sample_id),
    species = as.factor(species),
    sample_group = as.factor(sample_group),
    copies_uL = as.numeric(copies_uL))

#Load dilution + elution data
dilution_data <-read_excel(file.path(
  "C:/Users/Joyalea/Documents/UBCO/Thesis",
  "Final_Files",
  "RQ1_AMFAbundance",
  "Variance_DilutionFactors_All.xlsx"
))

View(data)

#Keep relevant columns
dilution_data <- dilution_data %>%
  select(sample_id, dilution_factor, elution_volume)

#Join the dilution data to the main ddPCR longform dataset
data <- data %>%
  left_join(dilution_data, by = "sample_id")

#Define root mass used per extraction
data <- data %>%
  mutate(root_mass = case_when(
    species %in% c("Bromus", "Festuca") ~ 0.1,
    species %in% c("Gaillardia", "Taraxacum") ~ 0.2,
    TRUE ~ NA_real_
  ))

#Calculate total copies and copies per gram (22 uL used per reaction)
data <- data %>%
```

```
  mutate(
    total_copies = copies_uL * 22 * dilution_factor * elution_volume,
    copies_g = total_copies / root_mass
  )


#Calculate total copies in elution volume. Use total copies to calculate
#copies per gram of host root
data <- data %>%
  mutate(
    total_copies = copies_uL * elution_volume * dilution_factor,
    copies_g = total_copies / root_mass
  )
```

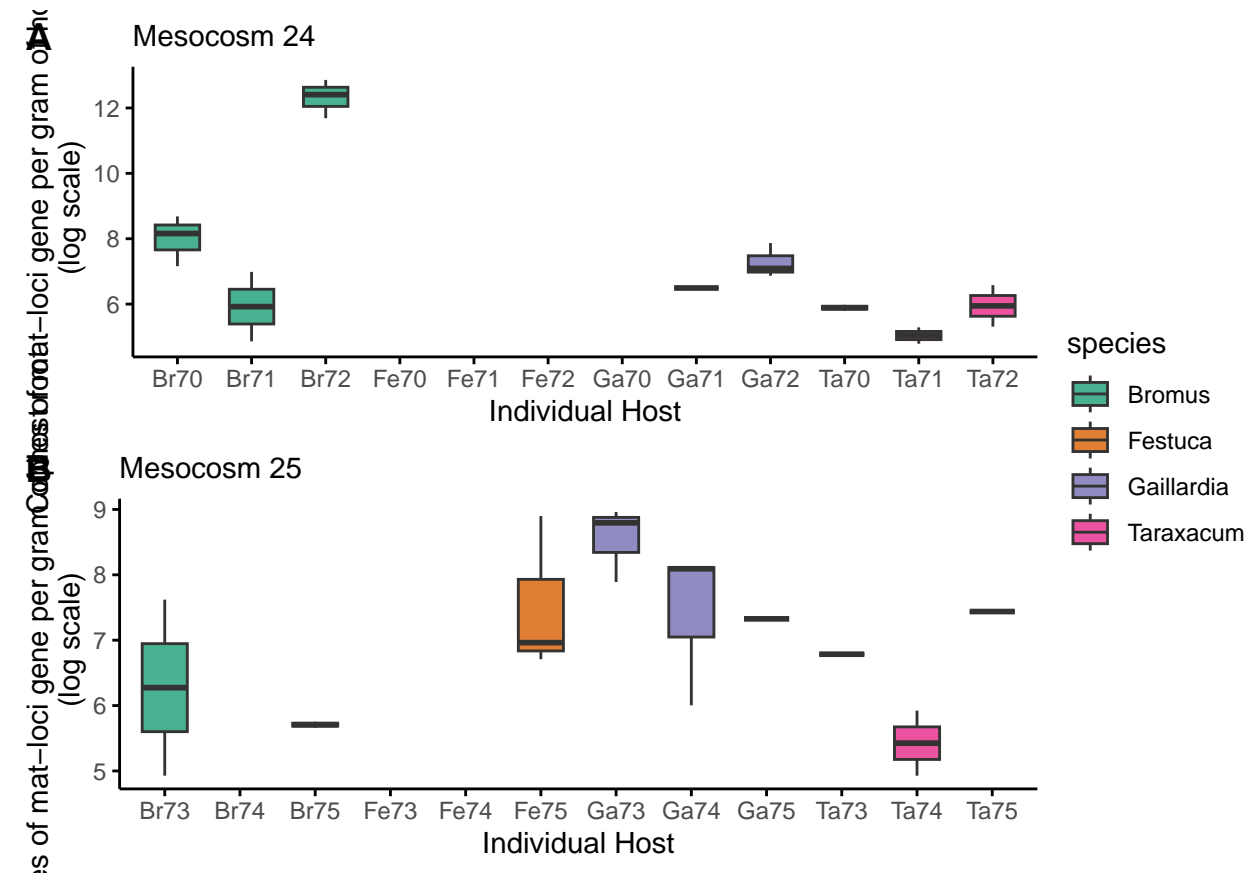## Visualise variation (1) within a host and (2) between hosts

```
## Warning: Removed 18 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning in get_plot_component(plot, "guide-box"): Multiple components found;
## returning the first one. To return all, use 'return_all = TRUE'.

## Warning: Removed 18 rows containing non-finite outside the scale range
## ('stat_boxplot()').
## Removed 18 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```
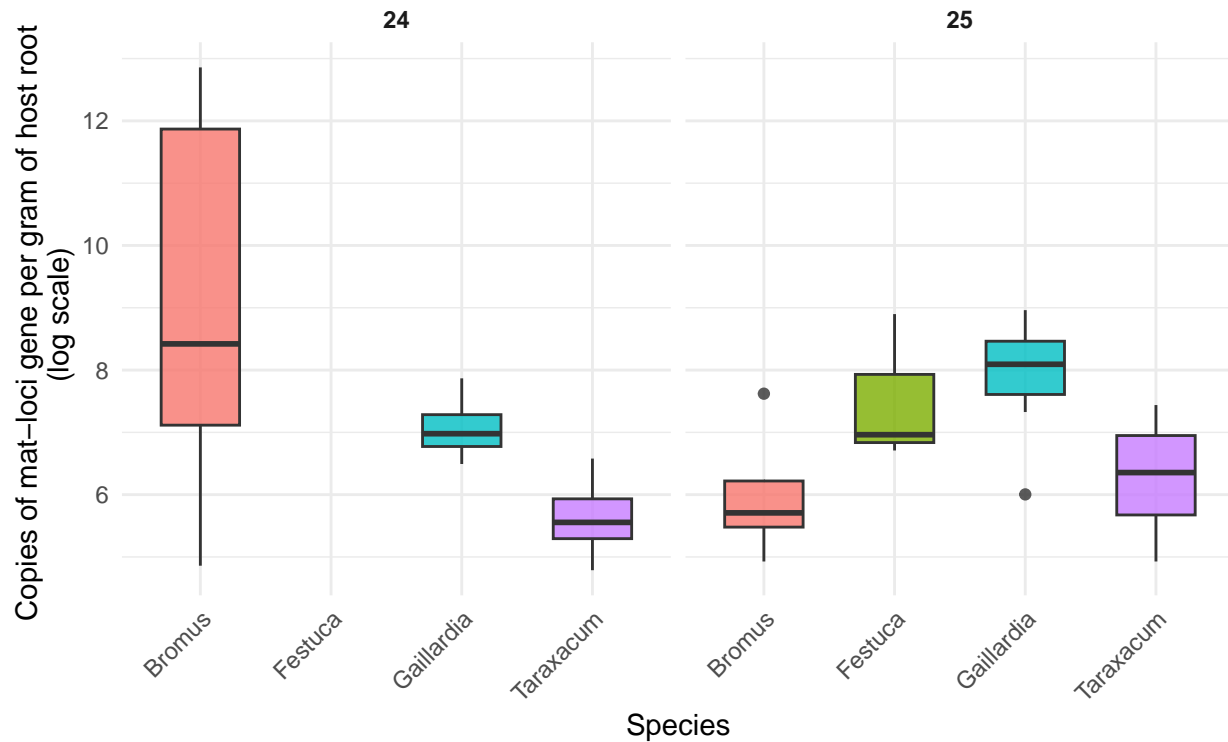
```
## Warning: Removed 36 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

Between–Host Variation in AMF Abundance

Variation across hosts of the same species within each mesocosm

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Calculate the coefficient of variation (1) Within hosts and (2) Between hosts

```r
#Calculate CV within each individual host (within-host)
#Ensure within group comparisons have > 2 observations
valid_within_groups <- data %>%
  filter(copies_g > 0) %>%
  group_by(sample_group, species, mesocosm_id) %>%
  summarise(n_reps = n(), .groups = "drop") %>%
  filter(n_reps >= 2)

cv_within_hosts <- data %>%
  semi_join(valid_within_groups, by = c("sample_group", "species", "mesocosm_id")) %>%
  group_by(sample_group, species, mesocosm_id) %>%
  summarise(
    mean = mean(copies_g, na.rm = TRUE),
    sd = sd(copies_g, na.rm = TRUE),
    cv = (sd / mean) * 100,
    .groups = "drop"
```

```
  ) %>%
  mutate(level = "Within-host")

#Calculate CV between hosts (conspecifics in same mesocosm)
#Have to filter species * mesocosm groups where >2 hosts have non-zero abundance
#Can't claculate CV for 1 observation
valid_between_groups <- data %>%
  filter(copies_g > 0) %>%
  group_by(species, mesocosm_id) %>%
  summarise(n_hosts = n_distinct(sample_group), .groups = "drop") %>%
  filter(n_hosts >= 2)

#calculate CV using groups with >2 observations
cv_between_hosts <- data %>%
  semi_join(valid_between_groups, by = c("species", "mesocosm_id")) %>%
  group_by(species, mesocosm_id) %>%
  summarise(
    mean = mean(copies_g, na.rm = TRUE),
    sd = sd(copies_g, na.rm = TRUE),
    cv = (sd / mean) * 100,
    .groups = "drop"
  ) %>%
  mutate(level = "Between-host")

#Combine into one dataframe
cv_comparison <- bind_rows(cv_within_hosts, cv_between_hosts)
#View CV within and between hosts
print(cv_comparison, n =Inf)
```

```
## # A tibble: 19 x 7
##    sample_group species    mesocosm_id    mean      sd    cv level
##    <fct>        <fct>            <dbl>   <dbl>   <dbl> <dbl> <chr>
##  1 Br70         Bromus              24   3559.   2308.  64.8 Within-host
##  2 Br71         Bromus              24    403     590.  146.  Within-host
##  3 Br72         Bromus              24 249350 132434.   53.1 Within-host
##  4 Br73         Bromus              25    726    1140.  157.  Within-host
##  5 Br75         Bromus              25    201     175.   86.9 Within-host
##  6 Fe75         Festuca             25   3065.   3687.  120.  Within-host
##  7 Ga72         Gaillardia          24   1590     891.   56.1 Within-host
##  8 Ga73         Gaillardia          25   5690    2683.   47.2 Within-host
##  9 Ga74         Gaillardia          25   2358.   1693.   71.8 Within-host
## 10 Ta70         Taraxacum           24    241.    212.   87.6 Within-host
## 11 Ta71         Taraxacum           24    159      55.2  34.7 Within-host
## 12 Ta72         Taraxacum           24    308.    371.  121.  Within-host
## 13 Ta74         Taraxacum           25    171.    189.  111.  Within-host
## 14 <NA>         Bromus              24  84437. 140306.  166.  Between-host
## 15 <NA>         Bromus              25    309     662.  214.  Between-host
## 16 <NA>         Gaillardia          24    603.    890.  147.  Between-host
## 17 <NA>         Gaillardia          25   2852.   2808.   98.5 Between-host
## 18 <NA>         Taraxacum           24    246.    237.   96.7 Between-host
## 19 <NA>         Taraxacum           25    344.    588.  171.  Between-host
```

```r
#Create levels (within and between) for the plot
cv_comparison$level <- factor(cv_comparison$level,
                              levels = c("Within-host", "Between-host"))

#Plot
cv_species_facet <- ggplot(cv_comparison, aes(x = level, y = cv, fill = level)) +
  geom_boxplot(alpha = 0.8, width = 0.6, outlier.shape = NA) +
 # geom_jitter(width = 0.15, shape = 21, size = 2, alpha = 0.7, color = "black") +
  facet_wrap(~ species) +
  scale_y_continuous(limits = c(0, NA), expand = c(0, 0)) +
  labs(
    title = "Coefficient of Variation (CV) within and between hosts",
    subtitle = "Faceted by species",
    x = "Variation Level",
    y = "CV (%)"
  ) +
  theme_minimal() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.line = element_line(colour = "black"),
    axis.ticks = element_line(colour = "black"),
    strip.text = element_text(face = "italic")
  )
print(cv_species_facet)
```

Coefficient of Variation (CV) within and between hosts
Faceted by species
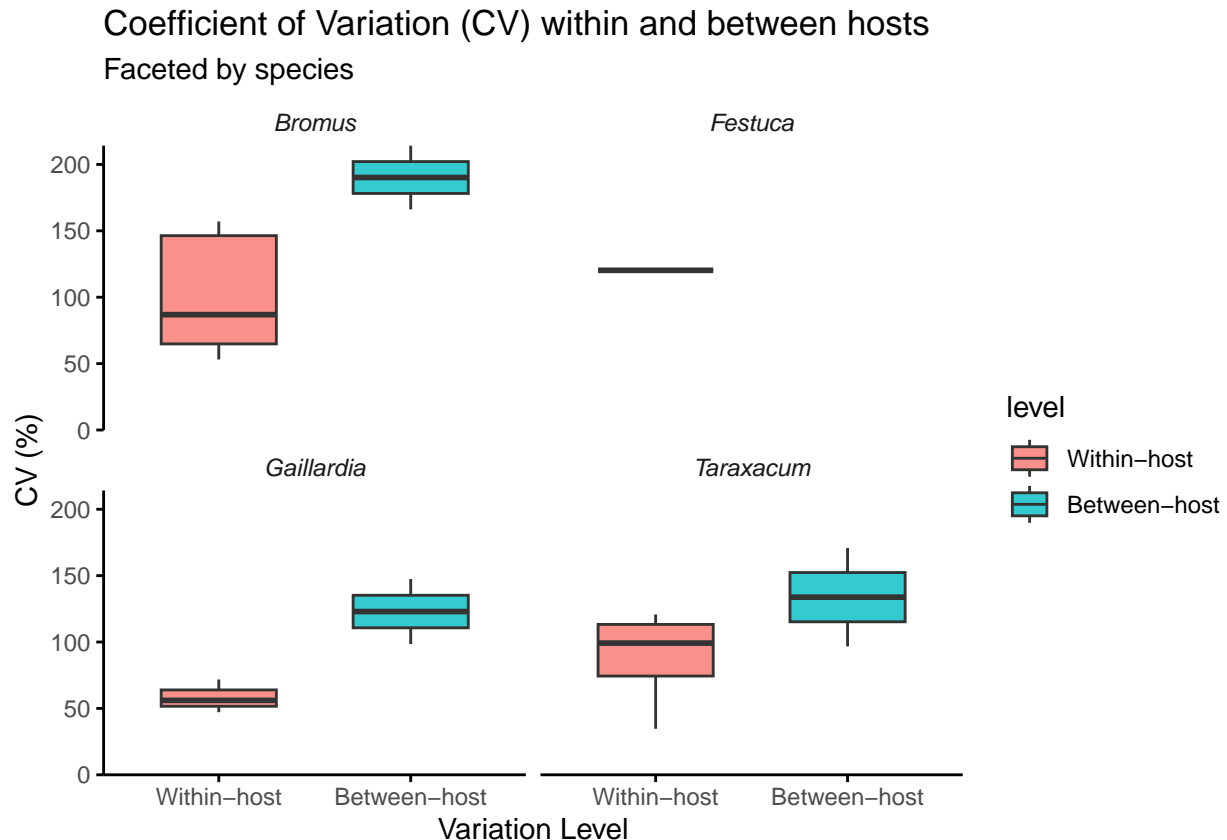
```r
#Perform a wilcoxon-rank summed test to see if there is a significant difference
#in within vs between host CV
wilcox.test(cv ~ level, data = cv_comparison) #significant
```

```
## 
##  Wilcoxon rank sum exact test
## 
## data:  cv by level
## W = 11, p-value = 0.01246
## alternative hypothesis: true location shift is not equal to 0
```