

# Image analysis of Festuca seeds

Joan M. Barreto O.

4/20/2020

## Methods

Image analysis was used to explore a potential relationship between seed size and ploidy level in Festuca species. Seeds from each accessions were randomly selected and imaged with an Epson Perfection V6 flatbed scanner at 600 dpi (dots per inch) with size 1987 (width) by 2169 (length) pixels. The TIF images (Tagged Image File Format) were processed with a custom Python script ([https://github.com/joanmanbar/Seed\\_Morphology](https://github.com/joanmanbar/Seed_Morphology)) that transformed the original images to threshold and measure the seed length and width respectively as the length (in pixels) of the major and minor axes of a fitted ellipse, whereas the area is calculated as the number of pixels in the seed. The script returned a dataset containing the image name, the seed number and their dimensional attributes. The data analysis was performed in R (R Core Team, 2014) by fitting liner models with seed area, length, and width as response and the ploidy level as predictor. To remove potential artifacts, we filtered out the seeds that were not within the 45th and 55th quantile. Finally, we used the *ggplot2* package (Wickham, 2009) for plots and figures.

## Preparing the data

```
# Get working directory
mydir <- getwd()

# Read data
mydata1 <- read.csv(paste0(mydir, '/input/Seeds_data_results.csv'))
ploidy <- read.csv(paste0(mydir, '/input/pic_list_with_ploidy.csv'))

# Remove 5X accessions
ploidy <- ploidy[-which(ploidy$Ploidy == "5X"), ]

# Subtract pic number
mydata1$Image_Name <- as.character(mydata1$Image_Name)
mydata1$Image_Name <- substr(mydata1$Image_Name, 1,
                             nchar(mydata1$Image_Name)-4)
mydata1$Image_Name <- as.factor(mydata1$Image_Name)

ploidy$Image_Name <- as.character(ploidy$Image_Name)
ploidy$Image_Name <- substr(ploidy$Image_Name, 1,
                             nchar(ploidy$Image_Name)-4)
ploidy$Image_Name <- as.factor(ploidy$Image_Name)

# Merge files
mydata2 <- merge(ploidy, mydata1)
colnames(mydata2)[1] <- "Accession"

# Keep seed with area values around the median
mydata3 <- group_by(mydata2, Accession) %>%
```

```

  filter(Area > quantile(Area, 0.45)) %>% filter(Area < quantile(Area, 0.55))
mydata3 <- as.data.frame(mydata3)

# Count total number of seeds per accession
# Total_seeds <- mydata2 %>% group_by (Accession) %>% summarise(n = n())
Total_seeds <- mydata3 %>% group_by (Accession) %>% summarise(n = n())

# Visualize coefficient of variation per accession
Accessions <- group_by(mydata3, Ploidy, Accession)

Accessions <- summarise(Accessions,
  sampled_seeds = n(),
  mean_area = mean(Area, na.rm = TRUE),
  sd_area = sd(Area, na.rm = TRUE),
  cv_area = cv(Area))
write.csv(Accessions, './output/Table_1.csv', row.names = FALSE)

# kable(Accessions, caption = "Coefficient of Variation")

```

## Convert pixels to milimeters

Now we must convert seed dimensions in pixels to mm! We used this website (<https://www.pixelcalculator.com/>) to get the conversion factors from px to mm.

```

## Convert px to mm
# Original Width (or length) = 2169 px (92mm)
W_px <- 2169
W_mm <- 92
# Original Height = 1987 px (84mm)
H_px <- 1987
H_mm <- 84
# 600 dpi

px2mm <- 0.5 * ( (W_px/W_mm) + (H_px/H_mm) )    # Average of both dimensions

# Apply conversion factors
mydata3$Area <- mydata3$Area / (W_mm*H_mm)
mydata3$Length <- mydata3$Length / px2mm
mydata3$Width <- mydata3$Width / px2mm

```

## Data visualization and analyses of variance

### Seed area

```

# Get the code for ggplot colors
ggplot_colors <- hue_pal()(4)

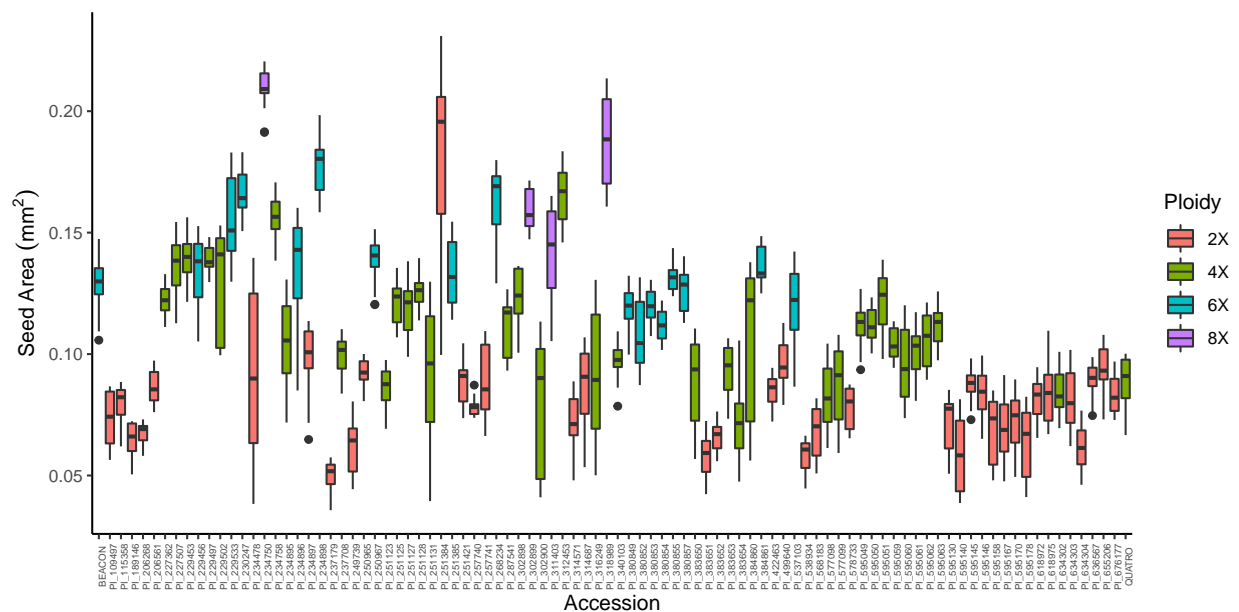
```

```
# Mean area of commercial varieties
BEACON_area_mean <- mean(mydata3[which(mydata3$Accession == "BEACON"), "Area"])
QUATRO_area_mean <- mean(mydata3[which(mydata3$Accession == "QUATRO"), "Area"])
```

```
# Distribution of the seed area across accessions
```

```
bp1 <- ggplot(mydata3, aes(x=Accession, y=Area, fill = Ploidy)) +
  geom_boxplot() +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, size = 5)) +
  labs(y= expression(paste("Seed Area ", (mm2))), x = "Accession")
```

```
bp1
```



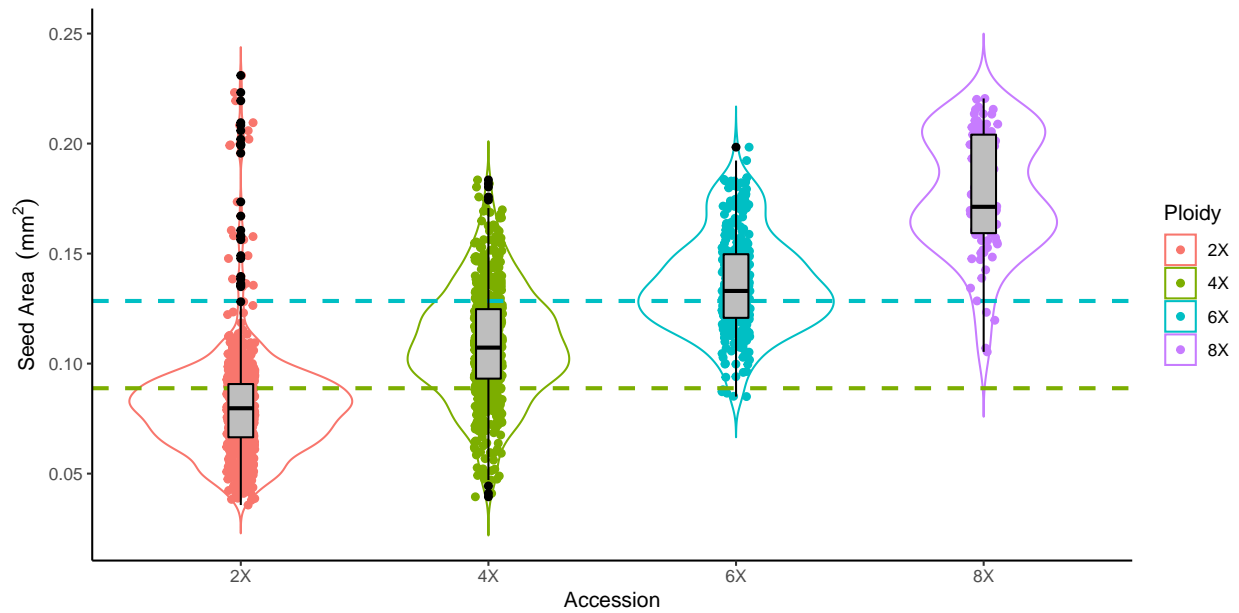
```
ggsave("./output/seed_area_accession.pdf")
```

```
## Saving 9 x 4.5 in image
```

```
# Distribution across ploidy
```

```
vp <- ggplot(mydata3, aes(x=Ploidy, y=Area, color=Ploidy)) +
  geom_violin(trim = FALSE) +
  geom_point(position = position_jitterdodge(seed = 1, dodge.width = 0.9)) +
  labs(y= expression(paste("Seed Area ", (mm2))), x = "Accession") +
  geom_hline(yintercept = QUATRO_area_mean, linetype = 2, color = ggplot_colors[2], size = 1) +
  geom_hline(yintercept = BEACON_area_mean, linetype = 2, color = ggplot_colors[3], size = 1) +
  geom_boxplot(width=0.1, color = 'black', fill = "gray") +
  theme_classic()
```

```
vp
```



```
# vp <- ggplot(Accessions, aes(x=Ploidy, y=mean_area, color=Ploidy)) +
#   geom_violin(trim = FALSE) +
#   geom_point(position = position_jitterdodge(seed = 1, dodge.width = 0.9)) +
#   labs(y= expression(paste("Seed Area ", (mm^2))), x = "Accession") +
#   geom_hline(yintercept = QUATRO_area_mean, linetype = 2, color = ggplot_colors[2], size = 1 ) +
#   geom_hline(yintercept = BEACON_area_mean, linetype = 2, color = ggplot_colors[3], size = 1) +
#   geom_boxplot(width=0.1, color = 'black', fill = "gray") +
#   theme_classic()
#
# vp
ggsave("../output/seed_area_ploidy.pdf")
```

## Saving 9 x 4.5 in image

```
# Model1 <- lm(Area ~ Ploidy + Accession, data = mydata3)
Model1 <- lm(Area ~ Ploidy, data = mydata3)
# summary(Model1)
# kable(anova(Model1))
kable(HSD.test(Model1, "Ploidy")$groups)
```

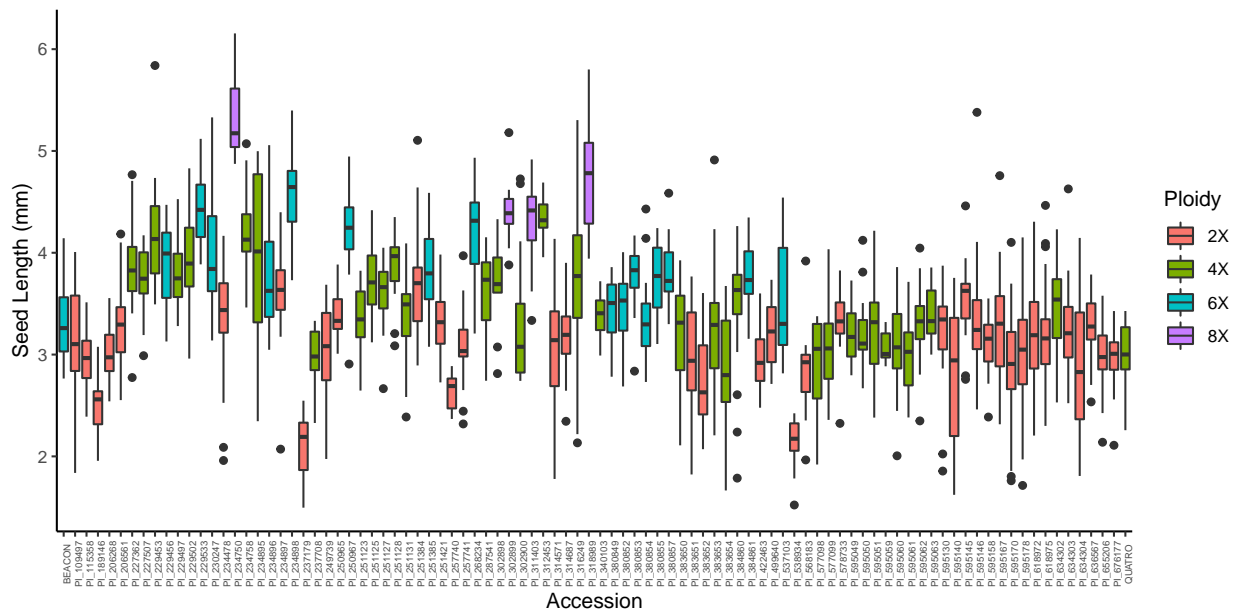
	Area	groups
8X	0.1772859	a
6X	0.1360331	b
4X	0.1082374	c
2X	0.0807685	d
\		
\		

## Seed length

```
# Mean length of commercial varieties
BEACON_length_mean <- mean(mydata3[which(mydata3$Accession == "BEACON"), "Length"])
QUATRO_length_mean <- mean(mydata3[which(mydata3$Accession == "QUATRO"), "Length"])
```

```
# Distribution of the seed length across accessions
bp2 <- ggplot(mydata3, aes(x=Accession, y=Length, fill = Ploidy)) +
  geom_boxplot() +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, size = 5)) +
  labs(y = "Seed Length (mm)", x = "Accession")
```

bp2

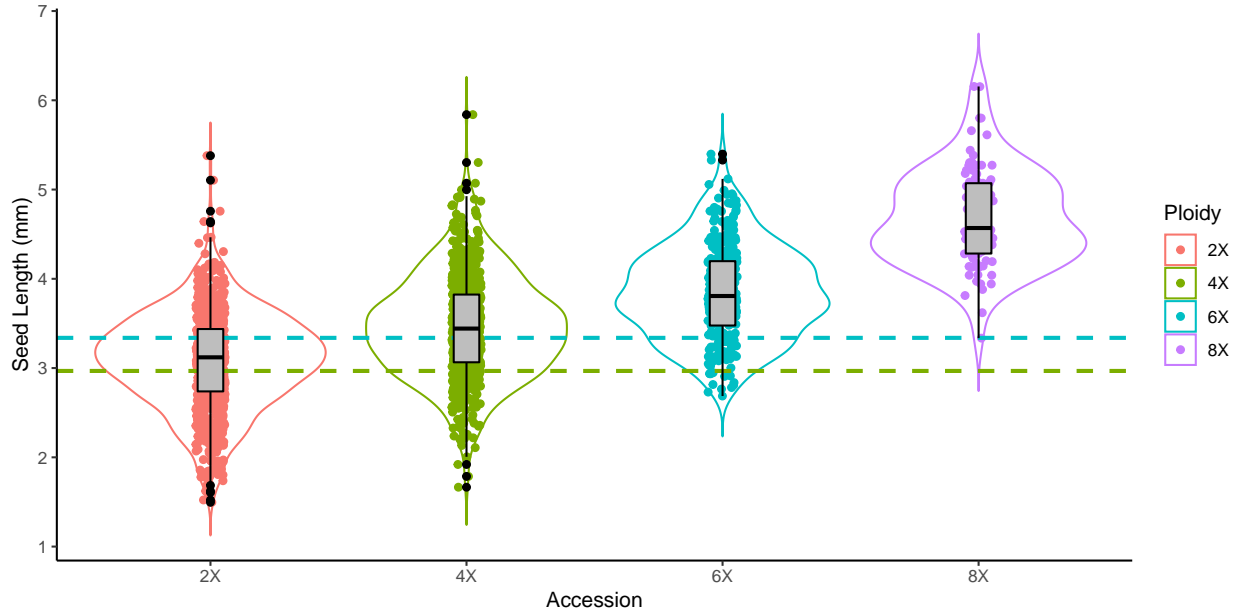


```
ggsave("./output/seed_length_accession.pdf")
```

## Saving 9 x 4.5 in image

```
# Distribution across ploidy
vp2 <- ggplot(mydata3, aes(x=Ploidy, y=Length, color=Ploidy)) +
  geom_violin(trim = FALSE) +
  geom_point(position = position_jitterdodge(seed = 1, dodge.width = 0.9)) +
  labs(y = "Seed Length (mm)", x = "Accession") +
  geom_hline(yintercept = QUATRO_length_mean, linetype = 2, color = ggplot_colors[2], size = 1) +
  geom_hline(yintercept = BEACON_length_mean, linetype = 2, color = ggplot_colors[3], size = 1) +
  geom_boxplot(width=0.1, color = 'black', fill = "gray") +
  theme_classic()
```

vp2



```
ggsave("./output/seed_length_ploidy.pdf")
```

```
## Saving 9 x 4.5 in image
```

```
# Regress the seed length on ploidy level
Model2 <- lm(Length ~ Ploidy, data = mydata3)
# summary(Model2)
# kable(anova(Model2))
kable(HSD.test(Model2, "Ploidy")$groups)
```

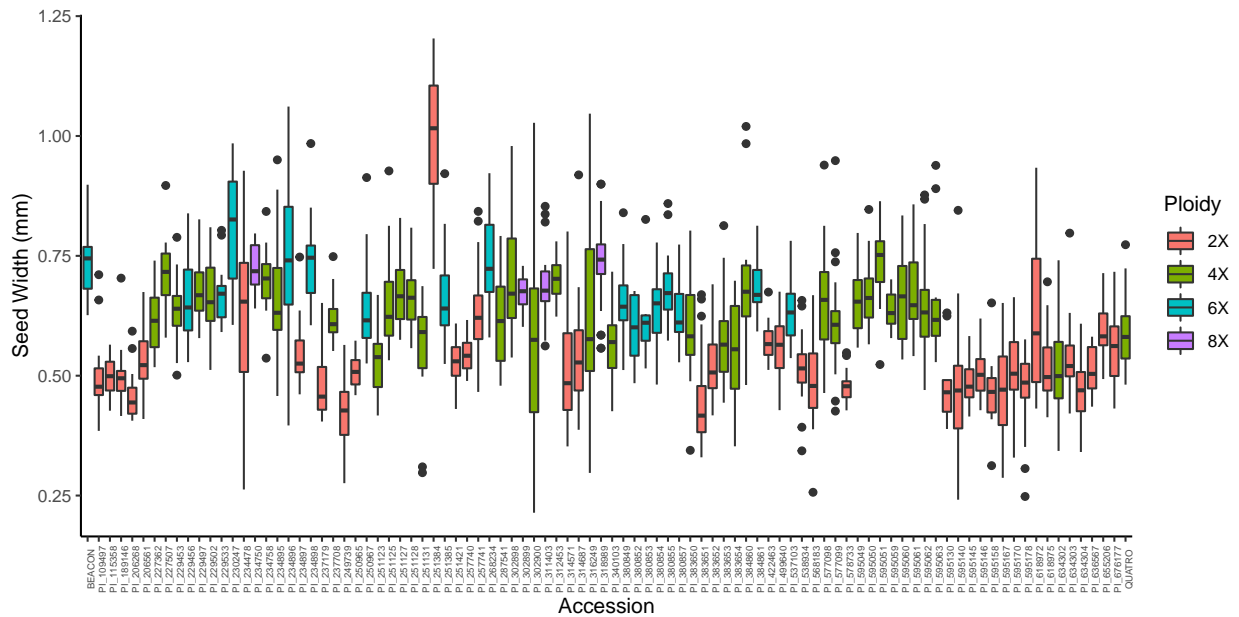
	Length	groups
8X	4.676767	a
6X	3.844965	b
4X	3.463788	c
2X	3.087498	d

## Seed width

```
# Mean width of commercial varieties
BEACON_width_mean <- mean(mydata3[which(mydata3$Accession == "BEACON"), "Width"])
QUATRO_width_mean <- mean(mydata3[which(mydata3$Accession == "QUATRO"), "Width"])

# Distribution of the seed length across accessions
bp2 <- ggplot(mydata3, aes(x=Accession, y=Width, fill = Ploidy)) +
  geom_boxplot() +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, size = 5)) +
  labs(y = "Seed Width (mm)", x = "Accession")
```

bp2

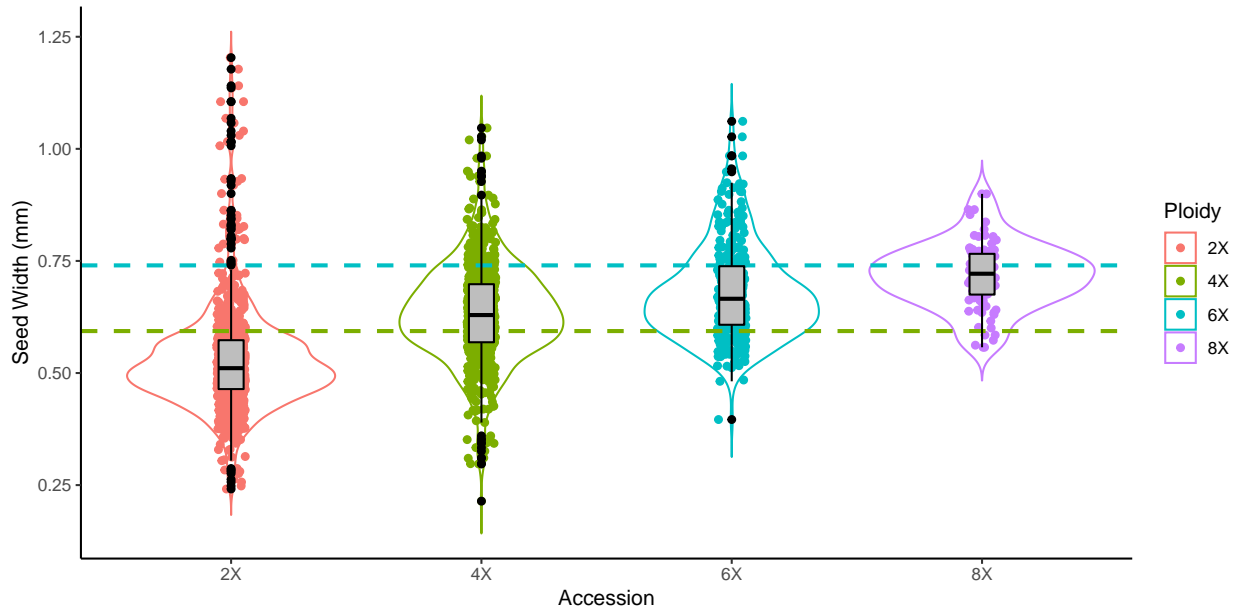


```
ggsave("./output/seed_width_accession.pdf")
```

## Saving 9 x 4.5 in image

```
# Distrubution across ploidy
vp2 <- ggplot(mydata3, aes(x=Ploidy, y=Width, color=Ploidy)) +
  geom_violin(trim = FALSE) +
  geom_point(position = position_jitterdodge(seed = 1, dodge.width = 0.9)) +
  labs(y= "Seed Width (mm)", x = "Accession") +
  geom_hline(yintercept = QUATRO_width_mean, linetype = 2, color = ggplot_colors[2], size = 1 ) +
  geom_hline(yintercept = BEACON_width_mean, linetype = 2, color = ggplot_colors[3], size = 1) +
  geom_boxplot(width=0.1, color = 'black', fill = "gray") +
  theme_classic()

vp2
```



```
ggsave("./output/seed_width_ploidy.pdf")
```

```
## Saving 9 x 4.5 in image
```

```
# Regress the seed width using a Type II ANOVA (main effects)
Model3 <- lm(Width ~ Ploidy, data = mydata3)
# summary(Model3)
# kable(anova(Model3))
kable(HSD.test(Model3, "Ploidy")$groups)
```

	Width	groups
8X	0.7173019	a
6X	0.6788670	b
4X	0.6334165	c
2X	0.5317359	d

## Results

Our results indicate that there are statistical differences in seed area, length and width between four different ploidy levels of Festuca at the  $\alpha = 5$ .

## Notes

- Please read the methods.
- These results coincide with the previous results.
- Some accession has more variability than other, do you want to keep them all? Look at the plots and let me know if there is any accession you want to remove or include.



- The Github repository has been updated. I would only talk about the latest Python code in the methods and would not mention the Matlab one.
- Let me know if there are some questions or concerns.

## References

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

Wickham, H. (2009) ggplot2: elegant graphics for data analysis. Springer New York.