

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) 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, '/Seeds_data_restuls.csv'))
ploidy <- read.csv(paste0(mydir, '/pic_list_with_ploidy.csv'))

# 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))

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

```

Table 1: Coefficient of Variation

| Ploidy | Accession | sampled_seeds | mean_area | sd_area | cv_area |
|--------|---------------|---------------|-----------|-----------|-----------|
| 2X | PI_109497_009 | 18 | 565.2222 | 83.41149 | 14.757292 |
| 2X | PI_115358_005 | 30 | 615.9667 | 62.15691 | 10.090954 |
| 2X | PI_189146_008 | 22 | 498.0909 | 57.45052 | 11.534144 |
| 2X | PI_206561_002 | 38 | 668.7105 | 52.23332 | 7.811050 |
| 2X | PI_234478_016 | 23 | 728.6522 | 274.50014 | 37.672315 |
| 2X | PI_234897_035 | 36 | 769.6389 | 89.71229 | 11.656413 |
| 2X | PI_237179_017 | 13 | 380.3077 | 59.10497 | 15.541355 |
| 2X | PI_249739_010 | 25 | 476.2800 | 81.20218 | 17.049252 |
| 2X | PI_250965_018 | 27 | 711.1111 | 41.99115 | 5.905005 |
| 2X | PI_251384_019 | 21 | 1397.9524 | 260.68630 | 18.647724 |
| 2X | PI_251421_037 | 19 | 683.4737 | 66.04407 | 9.663001 |
| 2X | PI_257740_036 | 20 | 605.2500 | 26.85218 | 4.436543 |
| 2X | PI_257741_006 | 28 | 679.4643 | 110.61461 | 16.279679 |
| 2X | PI_314571_011 | 15 | 553.4000 | 98.10111 | 17.726981 |
| 2X | PI_314687_026 | 16 | 667.0000 | 132.01212 | 19.791922 |
| 2X | PI_383651_003 | 27 | 449.2222 | 70.80707 | 15.762147 |
| 2X | PI_383652_004 | 25 | 511.2800 | 43.94269 | 8.594643 |
| 2X | PI_499640_020 | 22 | 742.4091 | 72.23484 | 9.729789 |
| 2X | PI_538934_023 | 19 | 447.1053 | 52.80772 | 11.811026 |
| 2X | PI_568183_033 | 16 | 522.1875 | 89.85931 | 17.208246 |
| 2X | PI_578733_034 | 18 | 604.3333 | 64.62380 | 10.693403 |
| 2X | PI_595130_032 | 11 | 542.0000 | 98.08262 | 18.096424 |
| 2X | PI_595140_014 | 20 | 453.1500 | 121.39465 | 26.789065 |
| 2X | PI_595145_031 | 15 | 678.4000 | 55.15407 | 8.130022 |
| 2X | PI_595146_030 | 15 | 641.8667 | 83.58389 | 13.022002 |
| 2X | PI_595158_029 | 11 | 523.7273 | 109.95189 | 20.994112 |
| 2X | PI_595167_028 | 16 | 541.6875 | 102.81032 | 18.979637 |
| 2X | PI_595170_022 | 25 | 550.0800 | 93.94277 | 17.078019 |
| 2X | PI_595178_021 | 19 | 486.1053 | 115.25040 | 23.708939 |
| 2X | PI_618972_024 | 22 | 626.4545 | 66.78232 | 10.660362 |
| 2X | PI_618975_025 | 21 | 649.2381 | 94.28728 | 14.522758 |
| 2X | PI_634303_027 | 18 | 625.3889 | 90.75056 | 14.511060 |
| 2X | PI_634304_013 | 17 | 470.2941 | 77.09553 | 16.393046 |
| 2X | PI_636567_012 | 26 | 694.5385 | 49.77649 | 7.166844 |
| 2X | PI_655206_015 | 25 | 725.8000 | 76.03234 | 10.475660 |
| 4X | PI_227362_045 | 20 | 945.5000 | 50.45842 | 5.336692 |
| 4X | PI_227507_047 | 18 | 1056.2778 | 85.33243 | 8.078598 |
| 4X | PI_229453_042 | 16 | 1082.5000 | 77.83658 | 7.190446 |
| 4X | PI_229497_046 | 18 | 1074.2222 | 42.88638 | 3.992319 |
| 4X | PI_229502_041 | 16 | 1012.3750 | 167.57242 | 16.552406 |
| 4X | PI_234758_043 | 20 | 1209.3000 | 66.74074 | 5.518956 |
| 4X | PI_234895_053 | 18 | 794.9444 | 140.98705 | 17.735460 |
| 4X | PI_237708_039 | 16 | 765.3125 | 65.42754 | 8.549127 |
| 4X | PI_251125_052 | 22 | 938.9545 | 70.65306 | 7.524652 |
| 4X | PI_251131_068 | 10 | 703.5000 | 239.12352 | 33.990549 |

| Ploidy | Accession | sampled_seeds | mean_area | sd_area | cv_area |
|--------|---------------|---------------|-----------|-----------|-----------|
| 4X | PI_287541_065 | 14 | 860.7857 | 99.71203 | 11.583839 |
| 4X | PI_302898_064 | 10 | 949.1000 | 101.53429 | 10.697954 |
| 4X | PI_302900_063 | 16 | 612.8125 | 209.10578 | 34.122311 |
| 4X | PI_312453_051 | 16 | 1272.7500 | 98.88074 | 7.769062 |
| 4X | PI_316249_044 | 20 | 692.8000 | 211.79822 | 30.571336 |
| 4X | PI_340103_054 | 15 | 749.1333 | 60.91665 | 8.131616 |
| 4X | PI_383650_049 | 20 | 685.7000 | 137.87336 | 20.106950 |
| 4X | PI_383653_038 | 24 | 718.6250 | 80.58688 | 11.214037 |
| 4X | PI_383654_048 | 15 | 547.4000 | 126.66198 | 23.138835 |
| 4X | PI_384860_050 | 16 | 800.5000 | 250.27718 | 31.265107 |
| 4X | PI_577098_066 | 18 | 644.1667 | 100.98296 | 15.676527 |
| 4X | PI_577099_067 | 20 | 675.3000 | 124.38061 | 18.418571 |
| 4X | PI_595049_062 | 22 | 867.0455 | 67.17956 | 7.748101 |
| 4X | PI_595050_061 | 17 | 867.0000 | 55.64396 | 6.417988 |
| 4X | PI_595051_055 | 14 | 940.1429 | 101.15324 | 10.759348 |
| 4X | PI_595059_060 | 14 | 806.8571 | 53.36645 | 6.614114 |
| 4X | PI_595060_059 | 23 | 740.0435 | 116.79449 | 15.782112 |
| 4X | PI_595061_058 | 20 | 779.4000 | 84.74072 | 10.872558 |
| 4X | PI_595062_057 | 19 | 824.3684 | 85.11770 | 10.325202 |
| 4X | PI_595063_056 | 19 | 866.7368 | 64.86554 | 7.483879 |
| 4X | PI_634302_040 | 24 | 651.2917 | 77.84794 | 11.952855 |
| 5X | PI_234751_087 | 22 | 1020.7273 | 62.81809 | 6.154248 |
| 5X | PI_251126_085 | 18 | 867.8889 | 108.96890 | 12.555628 |
| 5X | PI_274619_086 | 16 | 1310.9375 | 482.72897 | 36.823188 |
| 5X | PI_311405_089 | 16 | 877.8125 | 101.97269 | 11.616682 |
| 5X | PI_314522_090 | 16 | 895.1875 | 76.99023 | 8.600459 |
| 5X | PI_314523_084 | 15 | 1121.2000 | 103.32280 | 9.215376 |
| 5X | PI_315448_083 | 14 | 1011.0714 | 64.16381 | 6.346120 |
| 5X | PI_371896_082 | 10 | 1189.5000 | 70.12093 | 5.894992 |
| 5X | PI_380850_088 | 14 | 1031.2143 | 90.81018 | 8.806141 |
| 5X | PI_538933_092 | 15 | 1104.4000 | 95.76519 | 8.671241 |
| 6X | PI_229456_070 | 19 | 1038.5263 | 103.93340 | 10.007777 |
| 6X | PI_229533_073 | 20 | 1203.5500 | 125.84890 | 10.456474 |
| 6X | PI_230247_079 | 21 | 1288.3333 | 77.41985 | 6.009303 |
| 6X | PI_250967_075 | 22 | 1068.7727 | 74.89035 | 7.007136 |
| 6X | PI_251385_072 | 22 | 1030.5909 | 106.43785 | 10.327846 |
| 6X | PI_380849_078 | 20 | 921.4500 | 62.20973 | 6.751286 |
| 6X | PI_380852_076 | 13 | 838.0769 | 115.58724 | 13.791961 |
| 6X | PI_380853_071 | 15 | 926.5333 | 55.37775 | 5.976876 |
| 6X | PI_380854_069 | 15 | 865.2667 | 52.24320 | 6.037815 |
| 6X | PI_380855_077 | 20 | 1019.9500 | 46.79347 | 4.587820 |
| 6X | PI_380857_074 | 25 | 974.0400 | 66.73298 | 6.851153 |
| 6X | PI_384861_081 | 14 | 1054.4286 | 62.98561 | 5.973435 |
| 6X | PI_537103_080 | 17 | 930.0588 | 118.10719 | 12.698895 |
| 8X | PI_234750_094 | 13 | 1616.3846 | 74.06027 | 4.581847 |
| 8X | PI_318989_093 | 10 | 1591.1000 | 89.61455 | 5.632238 |

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

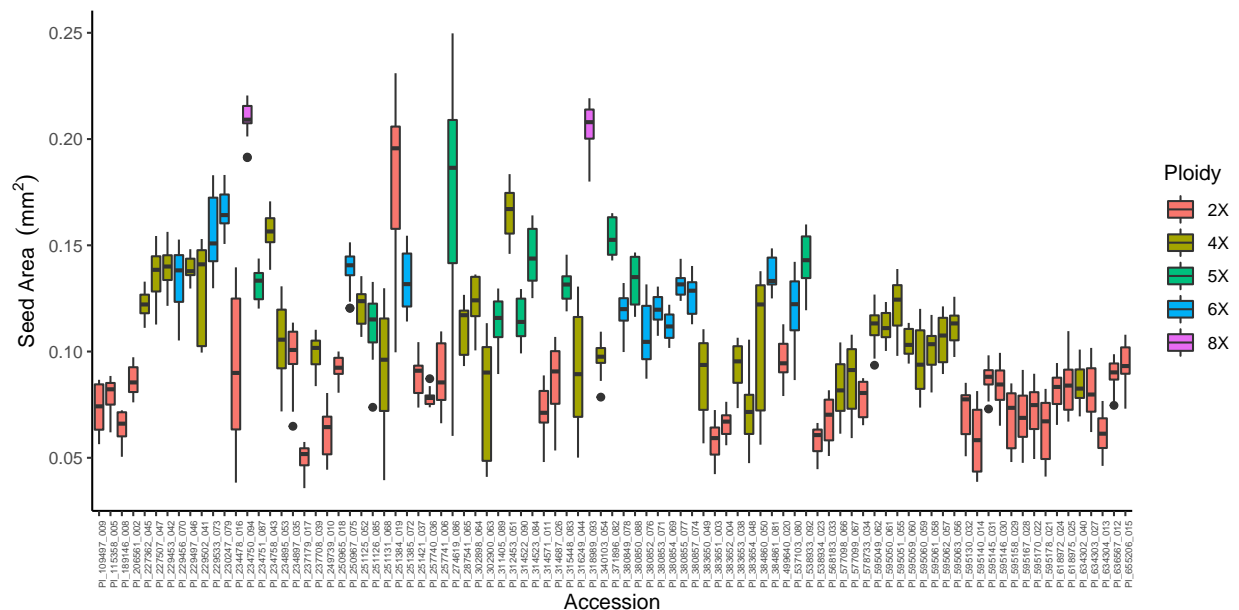
# Apply conversion factors
mydata3$Area <- mydata3$Area / (W_mm*H_mm)
mydata3$Length <- mydata3$Length / px2mm
mydata3$Width <- mydata3$Width / px2mm
```

Data visualization and analyses of variance

Seed area

```
# 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 ", (mm^2))), x = "Accession")
```

bp1



```
# Model1 <- lm(Area ~ Ploidy + Accession, data = mydata3)
Model1 <- lm(Area ~ Ploidy, data = mydata3)
summary(Model1)
```

```
##
## Call:
## lm(formula = Area ~ Ploidy, data = mydata3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.073489 -0.014611 -0.001362  0.011504  0.150069
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0809094  0.0009169   88.24  <2e-16 ***
## Ploidy4X      0.0277244  0.0014037   19.75  <2e-16 ***
## Ploidy5X      0.0528798  0.0021962   24.08  <2e-16 ***
## Ploidy6X      0.0514552  0.0018432   27.92  <2e-16 ***
## Ploidy8X      0.1268276  0.0052776   24.03  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02493 on 1706 degrees of freedom
## Multiple R-squared:  0.4822, Adjusted R-squared:  0.481
## F-statistic: 397.2 on 4 and 1706 DF,  p-value: < 2.2e-16
```

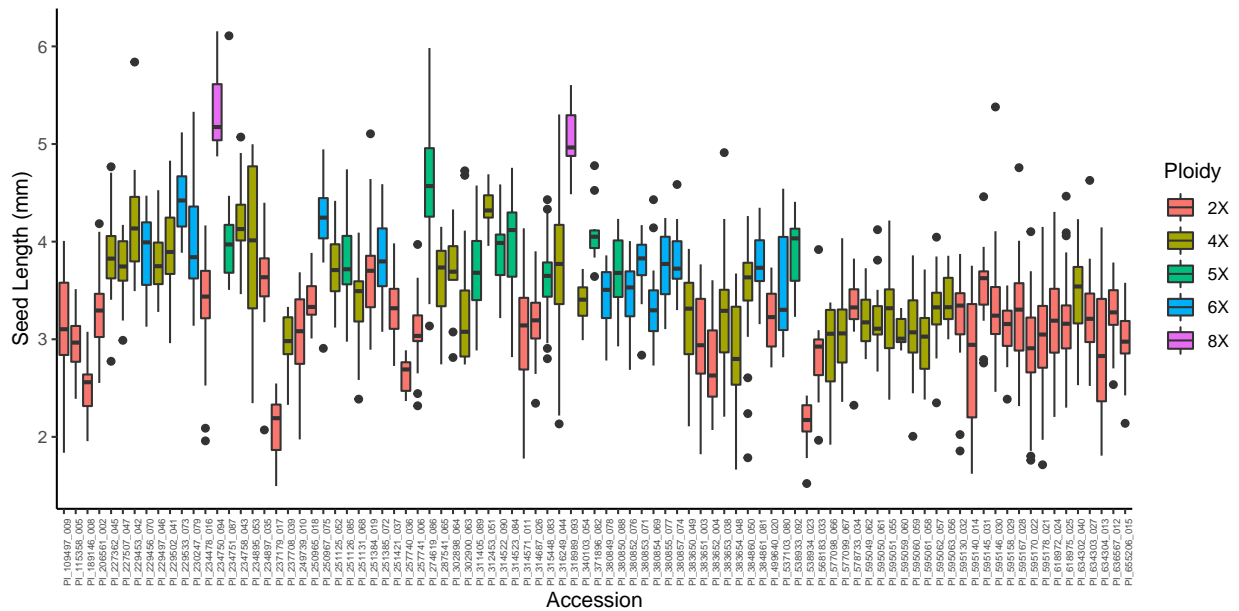
```
kable(anova(Model1))
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|------|-----------|-----------|---------|--------|
| Ploidy | 4 | 0.9870738 | 0.2467684 | 397.19 | 0 |
| Residuals | 1706 | 1.0599133 | 0.0006213 | NA | NA |
| \ | | | | | |
| \ | | | | | |

Seed 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
```



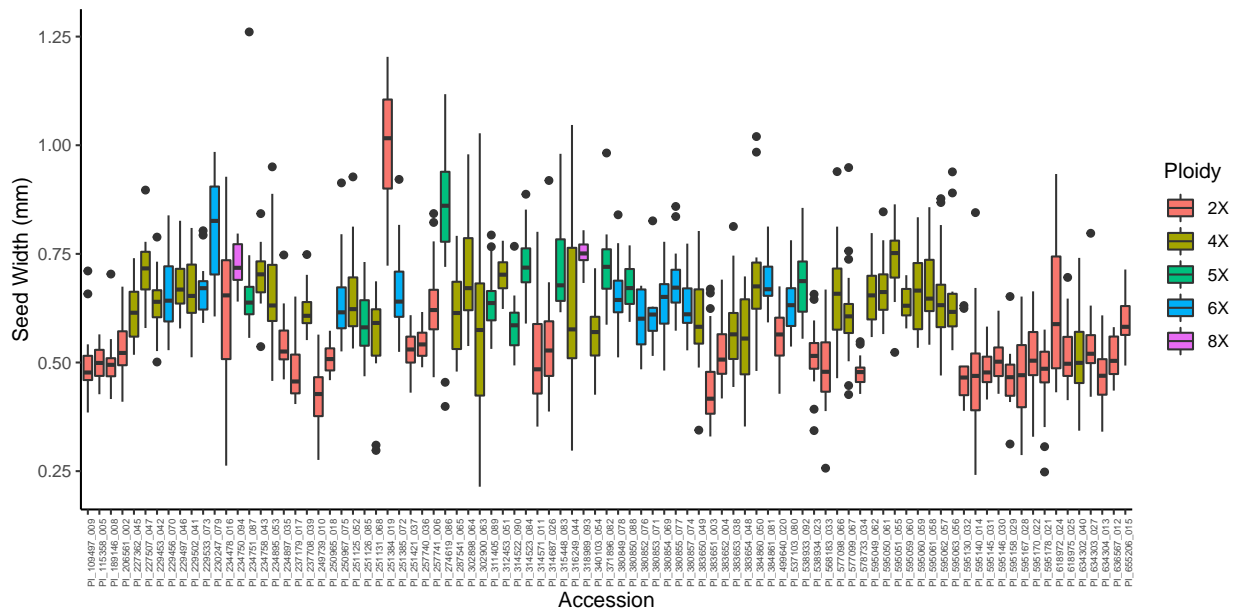
```
##  
## Call:  
## lm(formula = Length ~ Ploidy, data = mydata3)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -1.79906 -0.37155  0.00707  0.35580  2.37514   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   3.09412    0.02072  149.30  <2e-16 ***   
## Ploidy4X       0.37008    0.03173   11.66  <2e-16 ***   
## Ploidy5X       0.84633    0.04964   17.05  <2e-16 ***   
## Ploidy6X       0.72059    0.04166   17.30  <2e-16 ***   
## Ploidy8X       2.13010    0.11928   17.86  <2e-16 ***   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.5634 on 1706 degrees of freedom  
## Multiple R-squared:  0.3022, Adjusted R-squared:  0.3006   
## F-statistic: 184.7 on 4 and 1706 DF,  p-value: < 2.2e-16
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|------|----------|------------|----------|--------|
| Ploidy | 4 | 234.5362 | 58.6340528 | 184.7414 | 0 |
| Residuals | 1706 | 541.4581 | 0.3173846 | NA | NA |

Seed width

```
# Distribution of the seed width across accessions
bp3 <- 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")

bp3
```



```
# Regress the seed width using a Type II ANOVA (main effects)
Model3 <- lm(Width ~ Ploidy, data = mydata3)
summary(Model3)
```

```
##
## Call:
## lm(formula = Width ~ Ploidy, data = mydata3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42185 -0.06683 -0.01546  0.04814  0.67156
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.531867   0.004308  123.47 < 2e-16 ***
## Ploidy4X      0.104193   0.006595   15.80 < 2e-16 ***
## Ploidy5X      0.149710   0.010318   14.51 < 2e-16 ***
## Ploidy6X      0.129977   0.008660   15.01 < 2e-16 ***
## Ploidy8X      0.203081   0.024795    8.19 5.05e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1171 on 1706 degrees of freedom
## Multiple R-squared:  0.2183, Adjusted R-squared:  0.2164
## F-statistic: 119.1 on 4 and 1706 DF,  p-value: < 2.2e-16
```

```
kable(anova(Model3))
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|------|-----------|-----------|----------|--------|
| Ploidy | 4 | 6.532024 | 1.6330059 | 119.0769 | 0 |
| Residuals | 1706 | 23.395865 | 0.0137139 | NA | NA |

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.
- I did not scan the “controls” we used before because I did not have them.
- The Github repository has been updated. I would only talk about the latest Python code in the methods and would not mentioned 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.