

# About Me

Safietou Fall

## Contents

<b>Who I am and where I came from</b>	<b>1</b>
<b>Research Interests</b>	<b>2</b>
Influential papers . . . . .	4
The mathematics behind my research . . . . .	4
My computing experience . . . . .	4
What I hope to get out of this class . . . . .	5
<b>Evaluating some R code</b>	<b>5</b>
<b>Citations</b>	<b>8</b>

## Who I am and where I came from

My name is **Safietou Fall**. I am from **Senegal** (West Africa). I am a second year **PhD student** in the department of **Soil and Crop Science** under the supervision of Dr Geoffrey Morris.

I hold a Bachelor's degree in Biotechnology Engineering from the University of Sherbrooke (Quebec, Canada) and a Master's degree in Plant and Microbial Biotechnology from Cheikh Anta Diop University of Dakar (Dakar, Senegal).

Upon graduating from the Thierno Saidou Nourou Tall high school of Dakar, I decided to pursue a career in Biotechnology Engineering. I therefore successfully received training in this field from 2010 to 2015, which gave me the title of Biotechnology Engineer, member of the Quebec Order of Engineers. I then completed a Master's degree in Plant and Microbial Biotechnology (BIOVEM) at Cheikh Anta Diop University of Dakar. Throughout my journey, I have had the opportunity to design or participate in the development of several research projects in the field of microbial and plant biology. This allowed me to gain experience in research and enterprise.

In January 2020, I joined Dr Morris's Lab in Kansas State University (KSU), to complete a PhD degree as part of the GENMIL (pearl millet project) funded by SMIL/USAID. Since Dr. Morris got a new position at Colorado State University (CSU) in the fall of 2020, I transferred from KSU to CSU to continue my research activities under his supervision.

When I'm off campus, I usually stay home. I spend a lot of time on the phone talking with my family on WhatsApp. In fact, I have my 6 year old son and my husband in Senegal. I have organized my research activities so that I have the opportunity to travel and visit as often as possible, especially since all my field trials take place in Senegal. I am therefore very homely but my favorite activities are:

1. Swimming

2. Watching movies
3. Cooking

## Research Interests

The **GENMIL project** aims at accelerating the development of a combination of pearl millet innovations to put into farmers hands that at finite will result in sustainable productivity enhancement of PM for the food security and income generation in Senegal. More specifically, my thesis project aims to **support the Senegalese program on the development of drought-resistant varieties**.The research goals:

1. **Undercover the genetic basis of drought adaptive traits in the Senegalese germplasm, and**
2. **Develop an elite gene pool to enhance the rate of genetic gain.**

Here is a photo of one of my field trials in progress in Senegal.



## Influential papers

I spent a lot of time working through (Bernardo (2002)) and learned a lot about basics of plant breeding and genetics. In this book, (Bernardo (2002)) described the fundamental of crop improvement with very nice examples.

Another paper that inspired me a lot while defining my research goals is (Diack et al. (2020)). As, (Diack et al. (2020)) is the first paper published on genetic basis of flowering time in the Senegalese pearl millet germplasm.

## The mathematics behind my research

There are two fundamental equations that I use all the time in my research. The first is the breeder's equation.

$$Gy = \frac{k\sigma_A^2}{y\sqrt{(\sigma_e^2/rt) + (\sigma_G^2/t) + (\sigma_g^2)}}$$

It described how to improve response du selection (genetic gain).

The second equation below described the heritability of a trait:

$$h^2 = \frac{\sigma_G^2}{\sigma_p^2} = \frac{\sigma_A^2 + \sigma_D^2 + \sigma_I^2}{\sigma_p^2}$$

## My computing experience

My first experience with language programming was in undergrad using Matlab. Later, I also learned to use Visual basic. Here is some Matlab code that I found off the web:

```
nyear = 1975;
if (mod(nyear, 400) == 0)
    fprintf('%6u is a leap year', nyear)
elseif (mod(nyear,4) == 0) & (mod(nyear,100) ~= 0)
    fprintf('%6u is a leap year', nyear)
else
    fprintf('%6u is not a leap year', nyear)
end
```

I started programming in R in graduate school. Here is some R code that I wrote :

```
#generating random variables

#simple function for generating vector of random values from uniform normal distribution
runif(n=10,min = 0,max=10)
#need whole values?
round(runif(n=10,min = 0,max=10))
hist(round(runif(n=10,min = 0,max=10)))

#if our sample normally distributed
#Shapiro wilks tests:
#H_Null: data is normally distributed
#H_Alt: data is not normally distributed
```

```

shapiro.test(round(runif(n=10,min = 0,max=10)))
#what if we increase the sample size?
#central limit theorem
#samples =>30 are expected to have a mean approximately equal to the population mean
shapiro.test(round(runif(n=30,min = 0,max=100)))
#each value from uniform normal distribution has equal chance of being selected
hist(round(runif(n=1000,min = 0,max=10)))

##
#rnorm will select from a normal distribution
#abs() will give absolute values forcing only positive data
dat=data.frame(v1=abs(rnorm(n=100, mean = 10, sd =1 )),
               v2=abs(rnorm(n=100, mean = 25, sd =1 )))
)
#View distributions
library(ggplot2)
library(reshape2)
#reshape
dat_melt<- melt(dat)
head(dat_melt)
ggplot(dat_melt,aes(x=value, fill=variable)) + geom_density(alpha=0.25)
#formally test distribution
shapiro.test(dat$v1)
shapiro.test(dat$v2)
#compare the distribution of v1 to v2
#H_Null mean v1 = mean v2
#H_Alt mean v1 Not equal to mean v2
#T-test of unpaired data
t.test(dat$v1, dat$v2, alternative = "two.sided", var.equal = T)
#paired data
t.test(dat$v1, dat$v2, paired = TRUE, alternative = "two.sided",var.equal = T)

```

## What I hope to get out of this class

Give me a bullet list of three things:

- Be more familiar with bioinformatic tools (R programming, Unix, Summit);
- Be comfortable to work with genomic data;
- Learn how to align orthologous genes and identify putative variants ( variance analysis using SNPEFF).

## Evaluating some R code

```

#Example generate distribution
#rnorm will select from a normal distribution
#abs() will give absolute values forcing only positive data
dat=data.frame(v1=abs(rnorm(n=100, mean = 10, sd =1 )),
               v2=abs(rnorm(n=100, mean = 25, sd =1 )))
)

```

```
#View distributions
library(ggplot2)
library(reshape2)

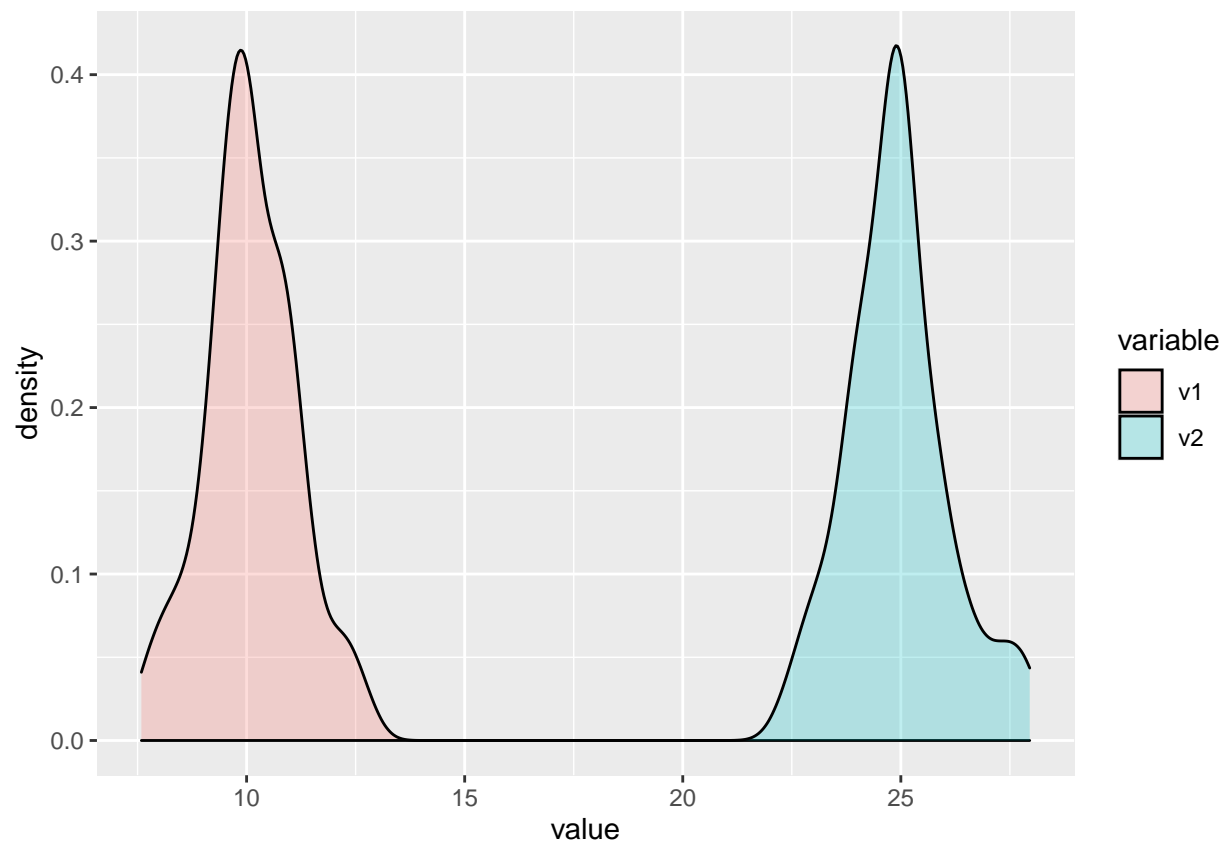
#reshape
dat_melt<- melt(dat)
```

```
## No id variables; using all as measure variables
```

```
head(dat_melt)
```

```
##   variable    value
## 1      v1 12.539690
## 2      v1 10.560777
## 3      v1  8.936212
## 4      v1 10.797856
## 5      v1  9.259056
## 6      v1  9.629084
```

```
ggplot(dat_melt,aes(x=value, fill=variable)) + geom_density(alpha=0.25)
```



```
#formally test distribution  
shapiro.test(dat$v1)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: dat$v1  
## W = 0.98943, p-value = 0.6187
```

```
shapiro.test(dat$v2)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: dat$v2  
## W = 0.97476, p-value = 0.05168
```

```
#compare the distribution of v1 to v2  
#H_Null mean v1 = mean v2  
#H_Alt mean v1 Not equal to mean v2
```

```
#T-test of unpaired data  
t.test(dat$v1, dat$v2, alternative = "two.sided", var.equal = T)
```

```
##  
## Two Sample t-test  
##  
## data: dat$v1 and dat$v2  
## t = -94.038, df = 198, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -15.14671 -14.52449  
## sample estimates:  
## mean of x mean of y  
## 10.0946 24.9302
```

```
#paired data  
t.test(dat$v1, dat$v2, paired = TRUE, alternative = "two.sided", var.equal = T)
```

```
##  
## Paired t-test  
##  
## data: dat$v1 and dat$v2  
## t = -96.954, df = 99, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -15.13922 -14.53198  
## sample estimates:  
## mean of the differences  
## -14.8356
```

## Citations

- Bernardo, Rex. 2002. *Breeding for Quantitative Traits in Plants*. Vol. 1. Stemma press Woodbury, MN.
- Diack, Oumar, Ghislain Kanfany, Mame Codou Gueye, Ousmane Sy, Amadou Fofana, Hamidou Tall, Desalegn D Serba, et al. 2020. “GWAS Unveils Features Between Early-and Late-Flowering Pearl Millets.” *BMC Genomics* 21 (1): 1–11.