

# REPORTE PRÁCTICA 6: Genómica poblacional y ancestría

Jacqueline Vasquez

12-04-2021

## Contents

4. ANÁLISIS DE DIVERSIDAD . . . . .	1
4.4 Gráficos de heterogocidad individual, diversidad de nucleótidos y Desequilibrio de ligamiento (LD). . . . .	1
5. ANÁLISIS DE ESTRUCTURA POBLACIONAL . . . . .	7
5.7 - Gráficos de PCA con R . . . . .	7
6. ANÁLISIS DE ADMIXTURE . . . . .	9
6.3 - Gráficos de ADMIXTURE para 2, 4 y 6 poblaciones . . . . .	9

## 4. ANÁLISIS DE DIVERSIDAD

```
EU <- read_delim("EU.geno.ld", delim = "\t")
```

### 4.4 Gráficos de heterogocidad individual, diversidad de nucleótidos y Desequilibrio de ligamiento (LD).

```
##
## -- Column specification -----
## cols(
##   CHR = col_double(),
##   POS1 = col_double(),
##   POS2 = col_double(),
##   N_INDV = col_double(),
##   `R^2` = col_double()
## )
```

```
OC <- read_delim("OC.geno.ld", delim = "\t")
```

```
##
## -- Column specification -----
## cols(
##   CHR = col_double(),
##   POS1 = col_double(),
##   POS2 = col_double(),
##   N_INDV = col_double(),
##   `R^2` = col_double()
## )
```

```
US <- read_delim("US.geno.ld", delim = "\t")
```

```
##
## -- Column specification -----
## cols(
```

```

## CHR = col_double(),
## POS1 = col_double(),
## POS2 = col_double(),
## N_INDV = col_double(),
## `R^2` = col_double()
## )

EU$dist <- ceiling((EU$POS2 - EU$POS1)/1000)*1000
OC$dist <- ceiling((OC$POS2 - OC$POS1)/1000)*1000
US$dist <- ceiling((US$POS2 - US$POS1)/1000)*1000

EU2 <- group_by(EU,dist) %>%
  summarise(meanR2 = mean(`R^2`))

OC2 <- group_by(OC,dist) %>%
  summarise(meanR2 = mean(`R^2`))

US2 <- group_by(US,dist) %>%
  summarise(meanR2 = mean(`R^2`))

dd <- bind_rows(EU2,OC2,US2)

dd$pop <- rep(c("EU","OC","US"),c(nrow(EU2),nrow(OC2),nrow(US2)))

write_csv(dd,"EU_OC_US.windowed.ld.csv")

# Heterogenecidad individual
het <- read_delim("EU_OC_US.het",delim = "\t")

##
## -- Column specification -----
## cols(
##   INDV = col_character(),
##   `O(HOM)` = col_double(),
##   `E(HOM)` = col_double(),
##   N_SITES = col_double(),
##   F = col_double()
## )

het

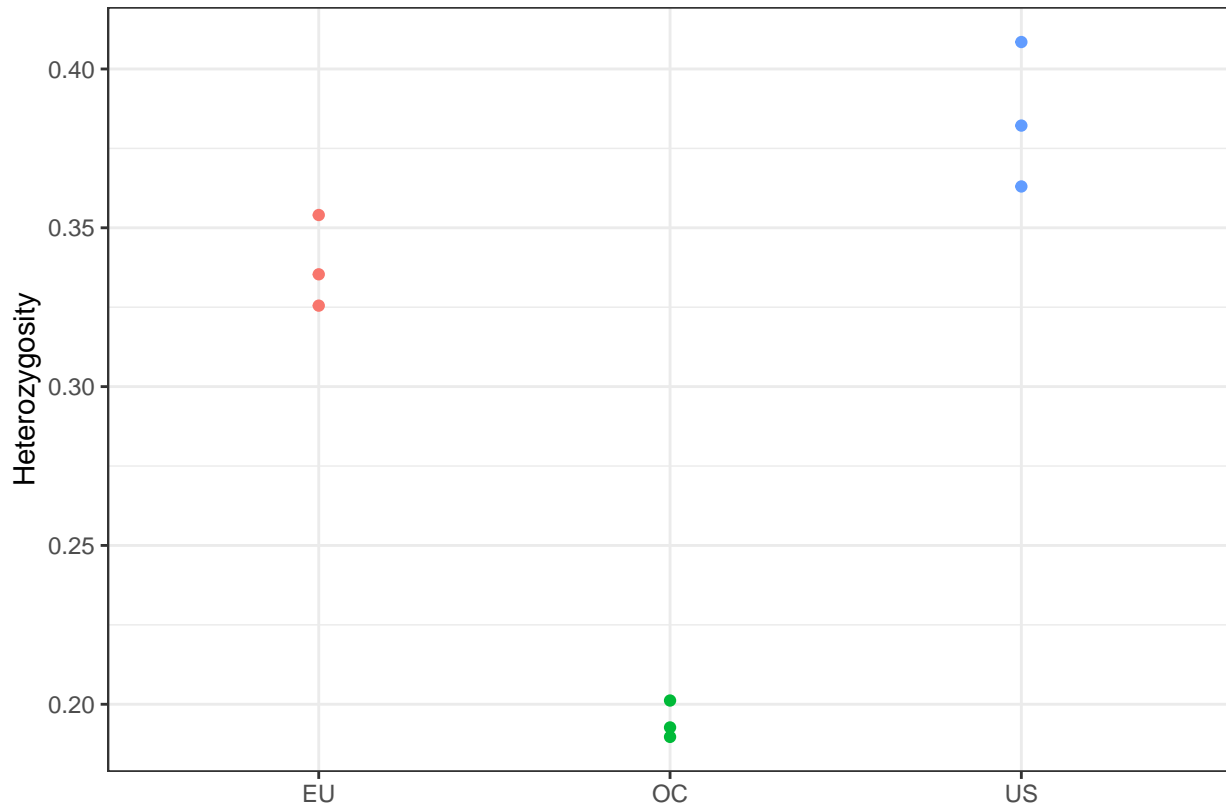
## # A tibble: 9 x 5
##   INDV                `O(HOM)` `E(HOM)` N_SITES      F
##   <chr>              <dbl>    <dbl>   <dbl>  <dbl>
## 1 2_WG0341511-DNA_A02_5408    8443    7747.   12703 0.140
## 2 3_WG0341511-DNA_A03_5416    8209    7749.   12708 0.0927
## 3 5_WG0341511-DNA_A05_5450    8571    7749.   12707 0.166
## 4 FR07958834              10157    7754.   12715 0.484
## 5 FR07958842              10268    7757.   12719 0.506
## 6 FR07958850              10301    7754.   12713 0.514
## 7 GNB12-1                 10618    9753.   16669 0.125
## 8 GNB12-10                9860     9754.   16669 0.0154
## 9 GNB12-11               10299    9754.   16670 0.0788

het$Heterozygosity <- 1-(het$`O(HOM)`/het$N_SITES)
het$Population <- c(rep("EU",3),rep("OC",3),rep("US",3))

```

```
A <- ggplot(het,aes(x = Population, y = Heterozygosity, col = Population)) +
  geom_point()+
  theme_bw()+
  theme(legend.position = "none")+
  xlab("")
```

A



*# Diversidad de nucleótidos*

```
pi_EU <- read_delim("EU.windowed.pi",delim = "\t")
```

```
##
## -- Column specification -----
## cols(
##   CHROM = col_double(),
##   BIN_START = col_double(),
##   BIN_END = col_double(),
##   N_VARIANTS = col_double(),
##   PI = col_double()
## )
```

```
pi_EU
```

```
## # A tibble: 4,903 x 5
##   CHROM BIN_START BIN_END N_VARIANTS      PI
##   <dbl>   <dbl>   <dbl>     <dbl>   <dbl>
## 1     1         1     200000         2 0.00000467
## 2     1    1000001    1200000         1 0.00000167
## 3     1    1200001    1400000         1 0.00000167
## 4     1    2200001    2400000         2 0.00000467
```

```
## 5      1      2800001 3000000      1 0.00000267
## 6      1      3000001 3200000      1 0.00000267
## 7      1      3200001 3400000      4 0.00001
## 8      1      3400001 3600000      3 0.000006
## 9      1      3600001 3800000      2 0.00000533
## 10     1      4000001 4200000      1 0.00000267
## # ... with 4,893 more rows
```

```
pi_OC <- read_delim("OC.windowed.pi",delim = "\t")
```

```
##
## -- Column specification -----
## cols(
##   CHROM = col_double(),
##   BIN_START = col_double(),
##   BIN_END = col_double(),
##   N_VARIANTS = col_double(),
##   PI = col_double()
## )
```

```
pi_OC
```

```
## # A tibble: 3,236 x 5
##   CHROM BIN_START BIN_END N_VARIANTS      PI
##   <dbl>   <dbl>   <dbl>     <dbl>   <dbl>
## 1     1         1       200000         2 0.00000533
## 2     1      200001     400000         1 0.000003
## 3     1     1200001    1400000         1 0.00000267
## 4     1     1800001    2000000         1 0.00000167
## 5     1     2200001    2400000         1 0.00000267
## 6     1     3200001    3400000         3 0.000006
## 7     1     3400001    3600000         1 0.00000167
## 8     1     4400001    4600000         2 0.00000333
## 9     1     4800001    5000000         1 0.000003
## 10    1     9000001    9200000         1 0.00000267
## # ... with 3,226 more rows
```

```
pi_US <- read_delim("US.windowed.pi",delim = "\t")
```

```
##
## -- Column specification -----
## cols(
##   CHROM = col_double(),
##   BIN_START = col_double(),
##   BIN_END = col_double(),
##   N_VARIANTS = col_double(),
##   PI = col_double()
## )
```

```
pi_US
```

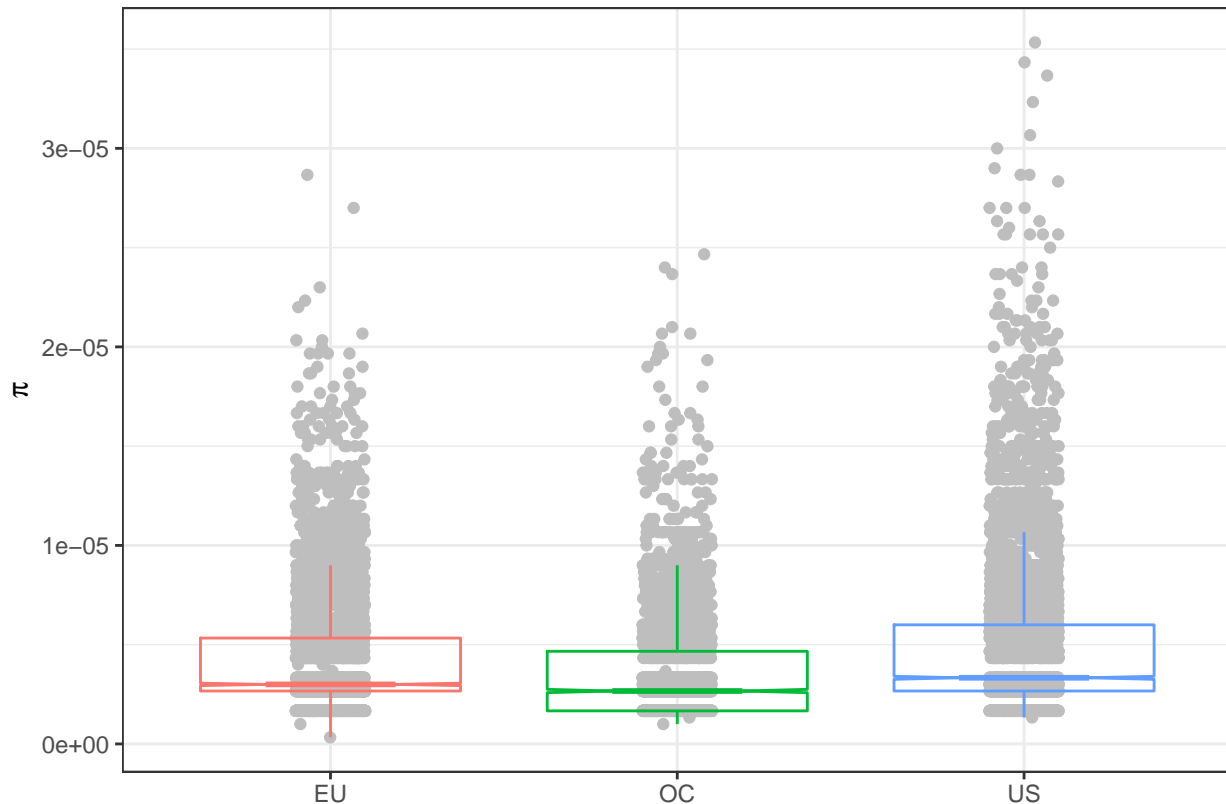
```
## # A tibble: 4,774 x 5
##   CHROM BIN_START BIN_END N_VARIANTS      PI
##   <dbl>   <dbl>   <dbl>     <dbl>   <dbl>
## 1     1      200001     400000         1 0.000003
## 2     1     1000001    1200000         2 0.000006
## 3     1     1600001    1800000         1 0.00000167
```

```
## 4      1    1800001 2000000          1 0.000003
## 5      1    2200001 2400000          1 0.000003
## 6      1    2400001 2600000          1 0.000003
## 7      1    2800001 3000000          1 0.000003
## 8      1    3000001 3200000          2 0.000006
## 9      1    3200001 3400000          6 0.0000177
## 10     1    3400001 3600000          2 0.000006
## # ... with 4,764 more rows
```

```
pi_all <- bind_rows(pi_EU,pi_OC,pi_US)
pi_all$Population<-c(rep("EU",nrow(pi_EU)),rep("OC",nrow(pi_OC)),rep("US",nrow(pi_US)))
```

```
B <- ggplot(pi_all,aes(x = Population, y = PI, col = Population))+
  geom_jitter(col = "grey",width = 0.1)+
  geom_boxplot(notch = T, alpha = 0,outlier.shape = NA)+
  theme_bw()+
  theme(legend.position = "none")+
  xlab("")+
  ylab(expression(pi))
```

B



```
# Desequilibrio de ligamiento
ld <- read_csv("EU_OC_US.windowed.ld.csv")
```

```
##
## -- Column specification -----
## cols(
##   dist = col_double(),
##   meanR2 = col_double(),
##   pop = col_character()
```

```
## )
```

```
ld
```

```
## # A tibble: 256 x 3
```

```
##   dist meanR2 pop  
##   <dbl> <dbl> <chr>
```

```
## 1  1000  0.891 EU
```

```
## 2  2000  0.9   EU
```

```
## 3  3000  0.781 EU
```

```
## 4  4000  0.75  EU
```

```
## 5  5000  0.25  EU
```

```
## 6  6000  1     EU
```

```
## 7  7000  1     EU
```

```
## 8  8000  0.85  EU
```

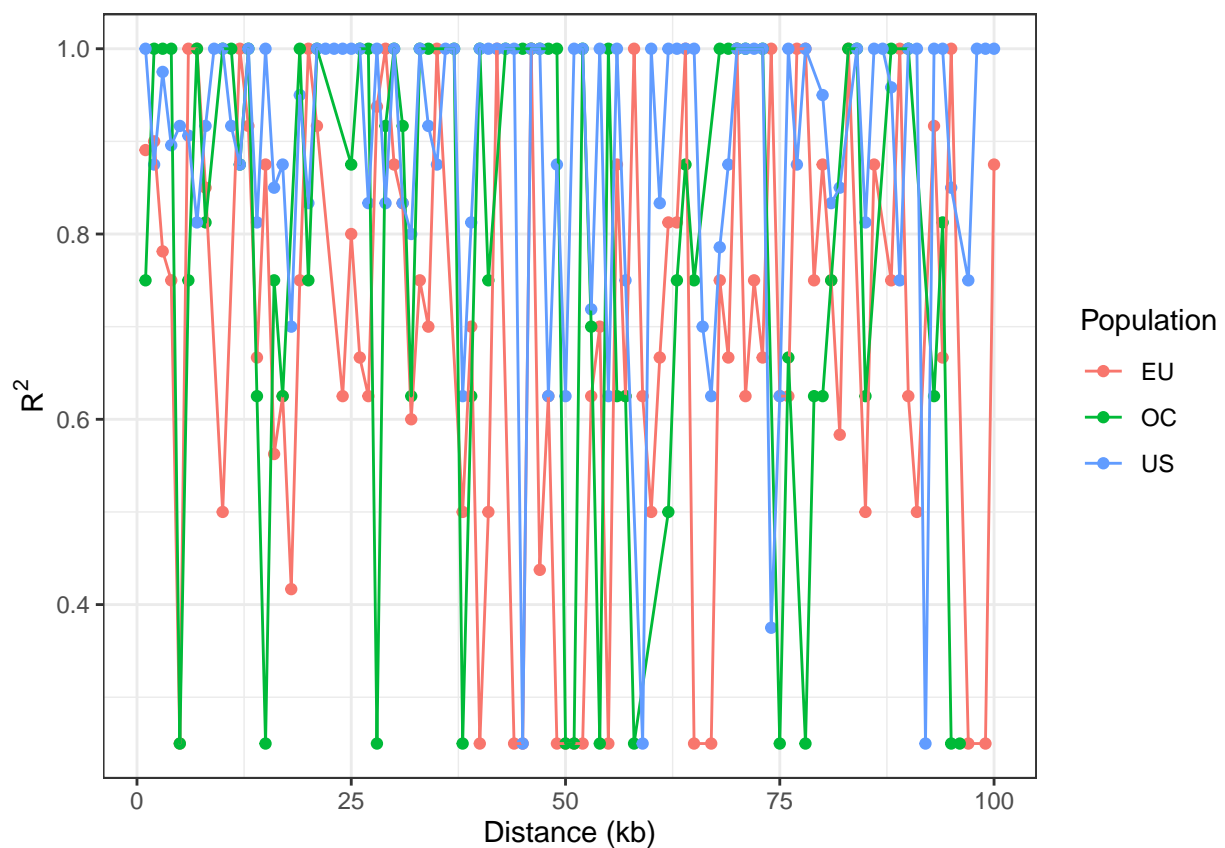
```
## 9 10000  0.5   EU
```

```
## 10 12000 1     EU
```

```
## # ... with 246 more rows
```

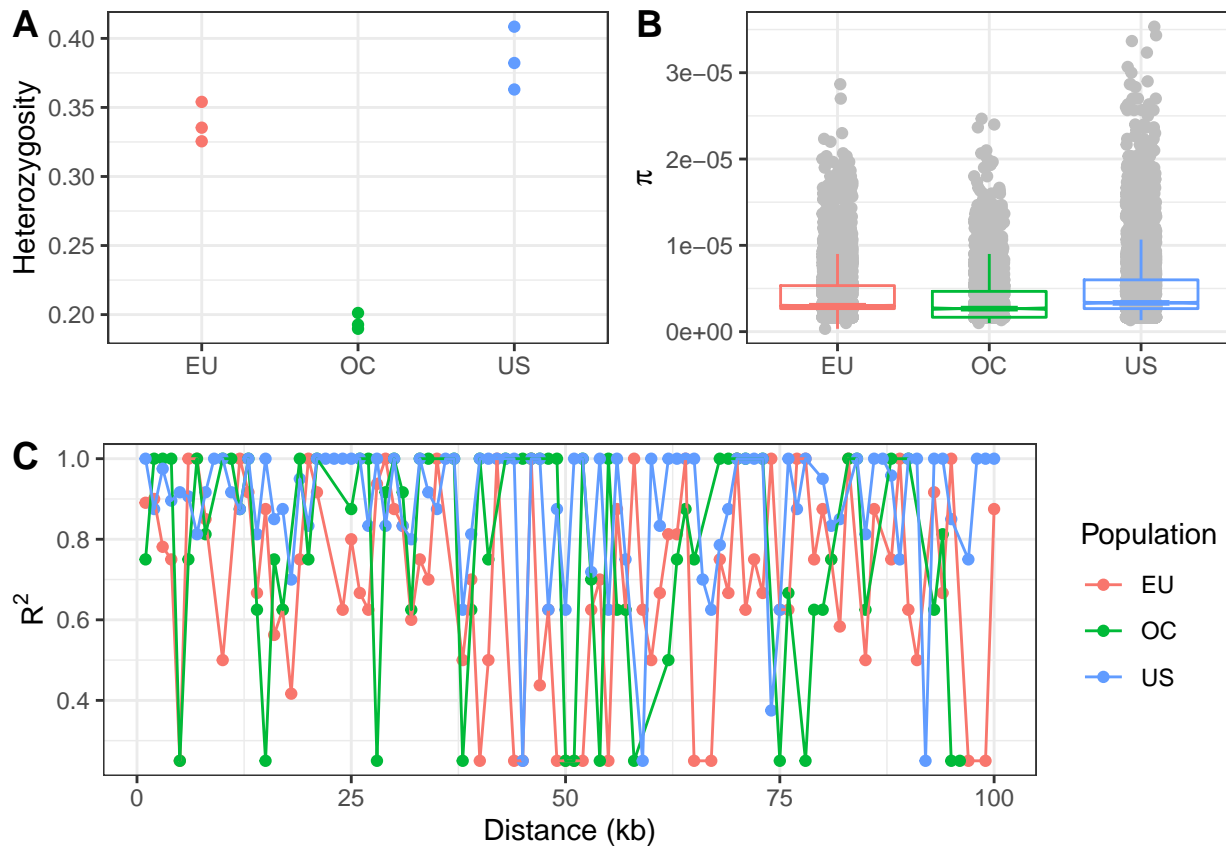
```
C <- ggplot(ld,aes(x = dist/1000, y = meanR2, col = pop)) +  
  geom_point()+  
  geom_line()+  
  theme_bw()+  
  xlab("Distance (kb)") +  
  ylab(expression(R2))+  
  scale_colour_discrete(name = "Population")
```

```
C
```



```
# GRAFICOS PANELES MULTIPLES
```

```
top_row <- plot_grid(A,B,labels = "AUTO")
plot_grid(top_row,C,nrow = 2,labels = c("", "C"))
```



## 5. ANÁLISIS DE ESTRUCTURA POBLACIONAL

```
#GRAFICOS DE PCA
```

```
pca1 <- read_delim("EU_OC_US.FilteredPrunedUnrel.eigenvec", delim = " ", col_names = F)
```

### 5.7 - Gráficos de PCA con R

```
##
## -- Column specification -----
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double(),
##   X6 = col_double()
## )
```

```
head(pca1)
```

```
## # A tibble: 5 x 6
```

```
##   X1                X2                X3                X4                X5                X6
```

```
##   <chr>                <chr>                <dbl> <dbl> <dbl> <dbl>
## 1 3_WG0341511-DNA_A03_54~ 3_WG0341511-DNA_A03_5~ 0.854 -0.286 0.0340 0.00175
## 2 FR07958842             FR07958842             -0.474 -0.767 -0.0280 -0.0304
## 3 GNB12-1                 GNB12-1                 -0.125 0.354 0.300 -0.747
## 4 GNB12-10                GNB12-10                -0.0875 0.335 -0.810 0.128
## 5 GNB12-11                GNB12-11                -0.151 0.305 0.502 0.651
```

```
colnames(pca1) <- c("Population","Individual",paste0("PC",c(1:4)))
head(pca1)
```

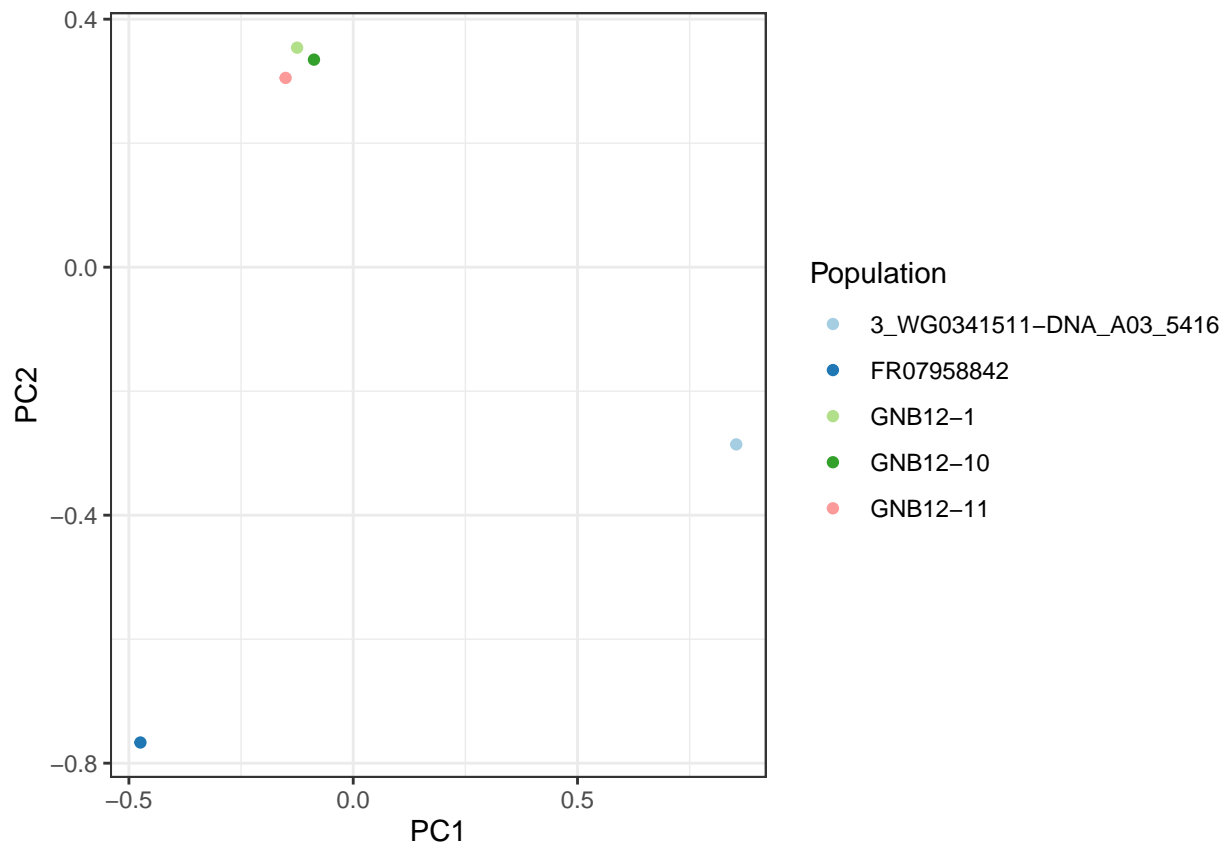
```
## # A tibble: 5 x 6
##   Population            Individual            PC1    PC2    PC3    PC4
##   <chr>                <chr>                <dbl> <dbl> <dbl> <dbl>
## 1 3_WG0341511-DNA_A03_54~ 3_WG0341511-DNA_A03_5~ 0.854 -0.286 0.0340 0.00175
## 2 FR07958842             FR07958842             -0.474 -0.767 -0.0280 -0.0304
## 3 GNB12-1                 GNB12-1                 -0.125 0.354 0.300 -0.747
## 4 GNB12-10                GNB12-10                -0.0875 0.335 -0.810 0.128
## 5 GNB12-11                GNB12-11                -0.151 0.305 0.502 0.651
```

```
mycols <- c("#a6cee3",
            "#1f78b4",
            "#b2df8a",
            "#33a02c",
            "#fb9a99",
            "#e31a1c",
            "#fdbf6f",
            "#ff7f00",
            "#cab2d6")
```

```
D <- ggplot(pca1,aes(x = PC1,y = PC2,col = Population))+
  geom_point()+
  theme_bw()+
  scale_colour_manual(values = mycols)
```

```
D
```





## 6. ANÁLISIS DE ADMIXTURE

```
# GRAFICOS DE ADMIXTURE PARA 2, 4 y 6 POBLACIONES
source("Admixture_plot.R")

pops <- read_delim("EU_OC_US.Thinned.fam", delim = " ", col_names = F)
```

### 6.3 - Gráficos de ADMIXTURE para 2, 4 y 6 poblaciones

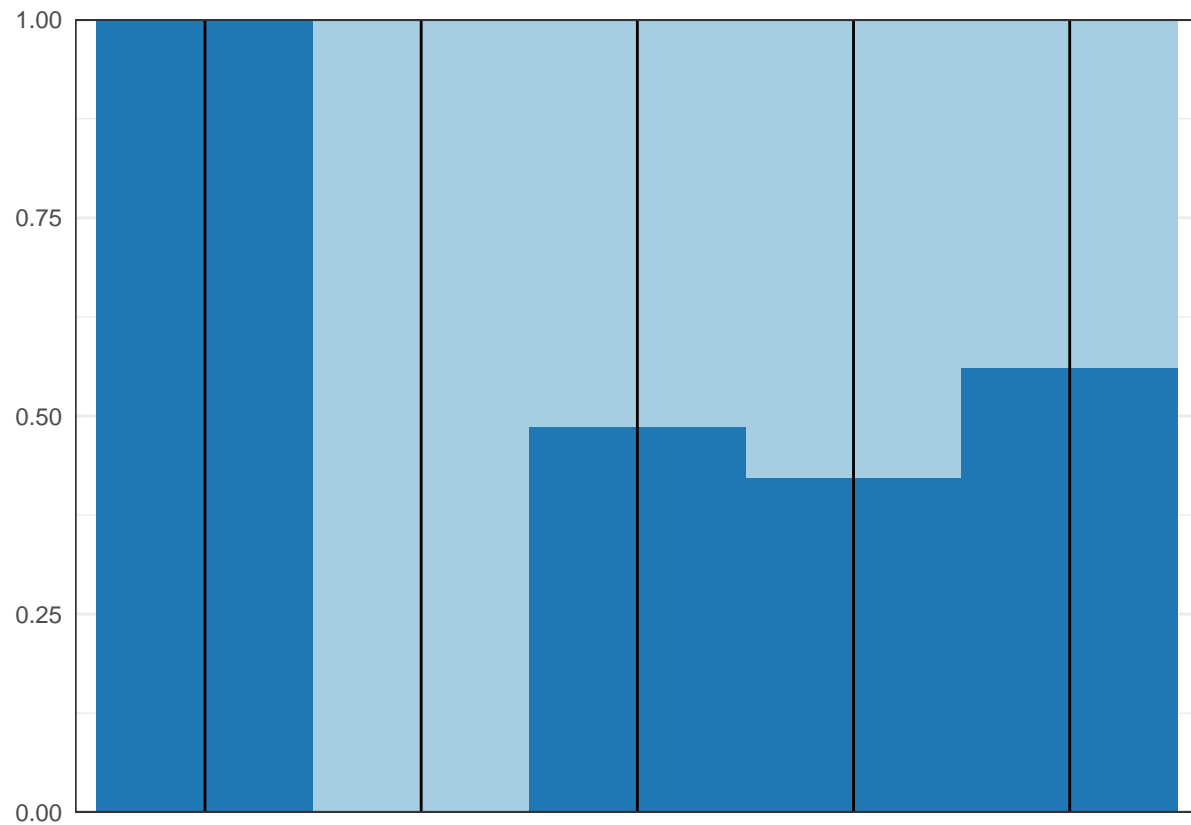
```
##
## -- Column specification -----
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double(),
##   X6 = col_double()
## )

K2 <- read_delim("EU_OC_US.Thinned.2.Q", delim = " ", col_names = F)

##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   X2 = col_double()
```

```
## )
```

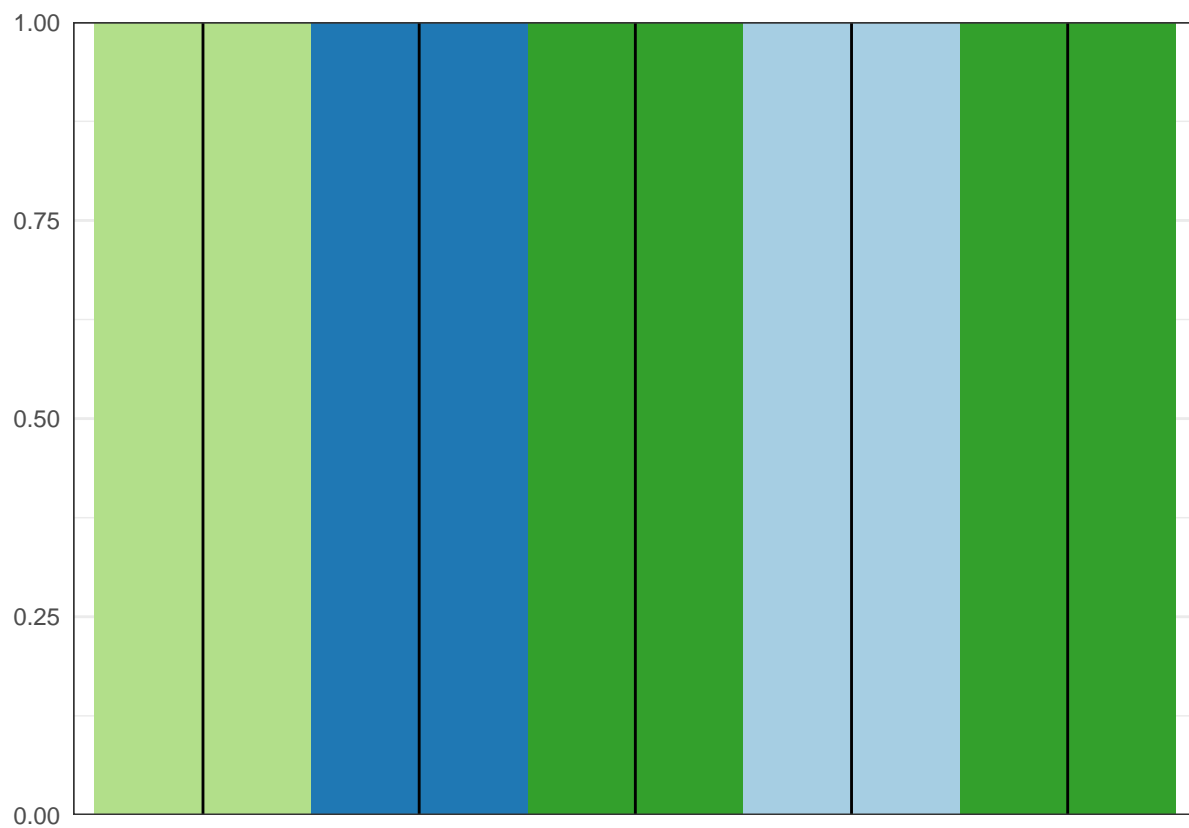
```
E <- admixtureplot(str_out = K2,k = 2, pops = pops, xaxis = F)
E
```



```
K4 <- read_delim("EU_OC_US.Thinned.4.Q", delim = " ", col_names = F)
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double()
## )
```

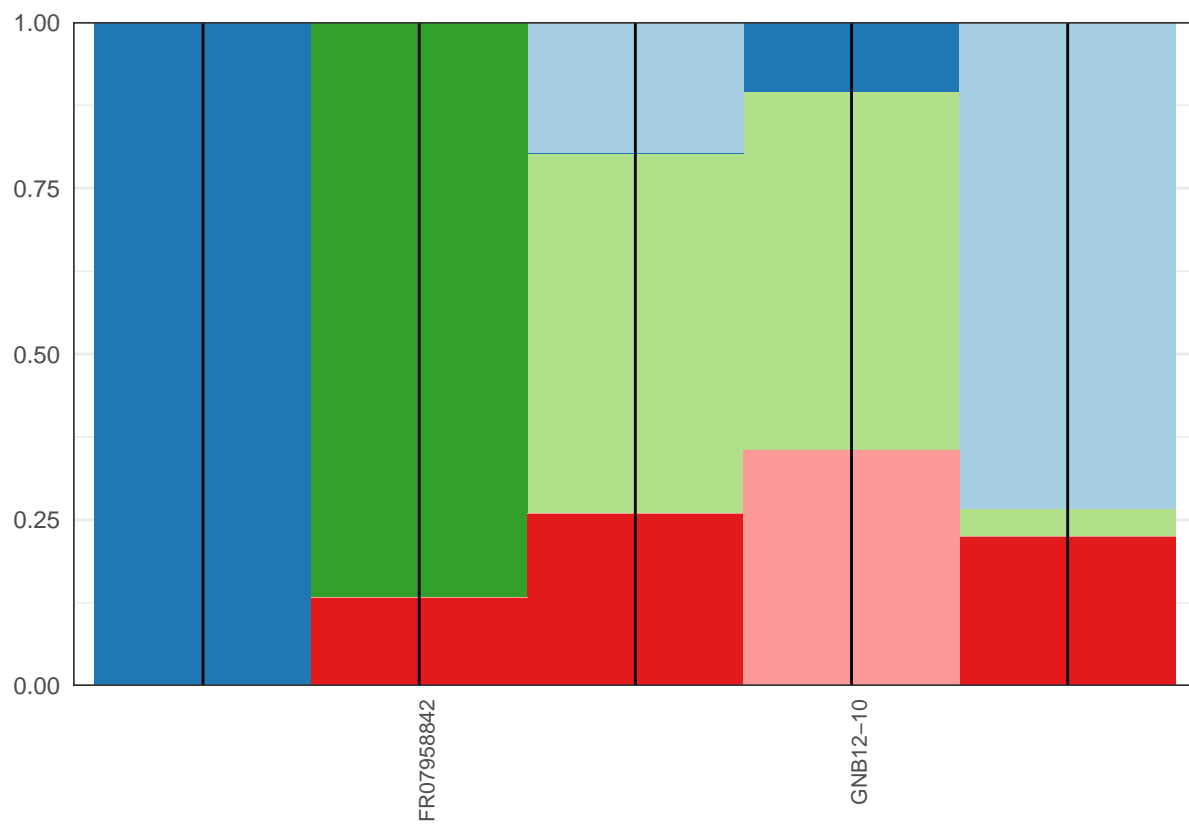
```
G <- admixtureplot(str_out = K4,k = 4, pops = pops, xaxis = F)
G
```



```
K6 <- read_delim("EU_OC_US.Thinned.6.Q", delim = " ", col_names = F)
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double(),
##   X5 = col_double(),
##   X6 = col_double()
## )
```

```
H <- admixtureplot(str_out = K6, k = 6, pops = pops, xaxis = T)
H
```



```
#GRÁFICOS
plot_grid(E,G,H,labels = "AUTO")
```

