

Introducción a la Genómica

UNAL nov 2017

Alejandro Cáceres
ISGlobal, Barcelona

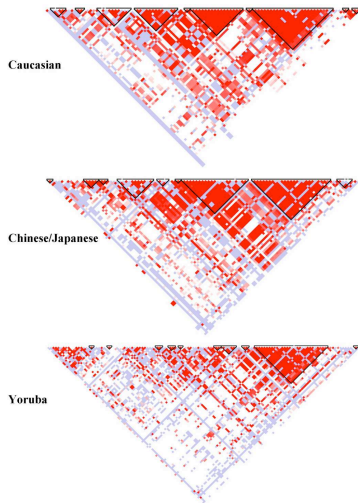
October 13, 2017

Genética de poblaciones

Los datos genómicos de SNPs nos sirven para ver la variabilidad genética entre poblaciones

- ▶ Estructura de LD en cada población
- ▶ Heterocigosidad por población
- ▶ Índice de fijación
- ▶ Patrones de recombinación
- ▶ Indicadores de selección
- ▶ Filogenia

Estructura de LD entre poblaciones



Cómo podemos extraer esa estructura

	A	T	Total
C	x_{CA}	x_{CT}	q_C
A	x_{AA}	x_{AT}	q_A
Total	p_A	p_T	1

el LD esta dado por $D = p_A * q_C - x_{CA}$

Recordemos que la fase se pierde por lo que el LD tiene que ser estimado modelando la probabilidad de una fase particular.

LD

Con snpStats se puede calcular el LD entre dos SNPs

```
library(snpStats)
```

```
load("datos/NewsnpSNPstats.RData")
```

```
ls()
```

```
## [1] "NewsnpSNPstats"
```

LD

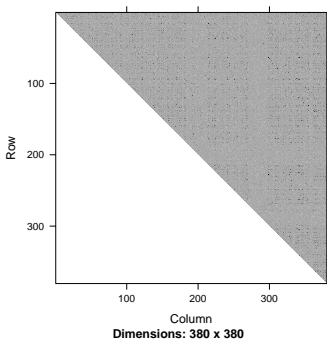
Con snpStats se puede calcular el LD entre dos SNPs

```
snps <- NewsnpSNPstats[,c(90,91)]  
ld(snps, stats=c("D.prime", "R.squared"),depth=1)  
  
## $D.prime  
## 2 x 2 sparse Matrix of class "dgCMatrix"  
##           rs561646592 rs34845889  
## rs561646592          . 0.01220497  
## rs34845889           . .  
##  
## $R.squared  
## 2 x 2 sparse Matrix of class "dgCMatrix"  
##           rs561646592 rs34845889  
## rs561646592          . 8.640143e-05  
## rs34845889           . .
```

LD

Podemos también calcular el LD entre todos los SNPs

```
LD <- ld(NewsnpsSNPstats, stats=c("D.prime", "R.squared"), depth=379)  
image(LD$R.squared, lwd=0)
```



LD

LD para una población

Seleccionemos los individuos de la población GBR

```
ids <- read.table("datos/20130606_g1k.ped", sep="\t", header=TRUE)

rownames(ids) <- ids$Individual.ID
pops <- ids[rownames(NewsnpsSNPstats),]$Population
head(pops)

## [1] GBR GBR GBR GBR GBR GBR
## 26 Levels: ACB ASW BEB CDX CEU CHB CHS CLM ESN FIN GBR GIH GWD IBS .

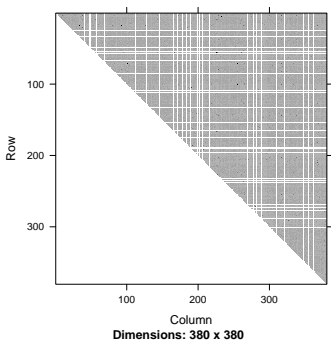
GBR <- pops=="GBR"
head(GBR)

## [1] TRUE TRUE TRUE TRUE TRUE TRUE
```

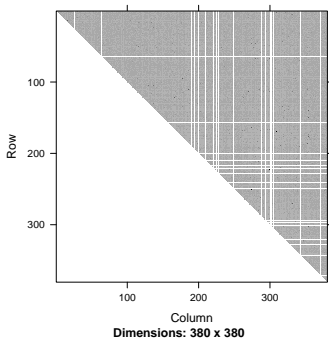
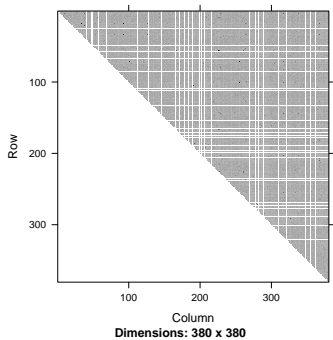

LD

LD para GRB

```
snpsGBR <- NewsnpsSNPstats[GBR,]  
LDGBR <- ld(snpsGBR, stats=c("D.prime", "R.squared"),depth=379)  
image(LDGBR$R.squared, lwd=0)
```



Ejercicio LD para una población YRI



LD

Ejercicio LD para una población YRI

```
YRI <- pops=="YRI"  
snpsYRI <- NewsnpsSNPstats[YRI,]  
LDYRI <- ld(snpsYRI, stats=c("D.prime", "R.squared"),depth=379)  
image(LDYRI$R.squared, lwd=0)
```

Heterocigosidad

- ▶ La heterocigosidad es la proporción de heterocigotos que hay en una población.
- ▶ Mayor tasa de heterocigosidad mayor es la variabilidad genética

se calcula como

$$2pq = 1 - p^2 - q^2 \quad (1)$$

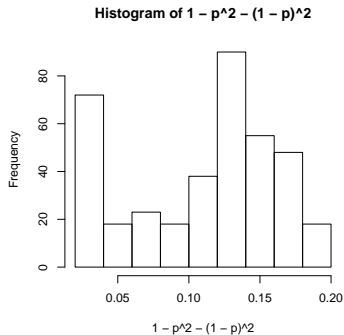
Heterecigosidad

```
sumSnps <- col.summary(NewsnpSNPstats)
head(sumSnps)
```

##		Calls	Call.rate	Certain.calls		RAF	MAF
##	rs555347111	2504	1	1	0.07368211	0.07368211	0.85
##	rs542617372	2504	1	1	0.07607827	0.07607827	0.85
##	rs17763596	2504	1	1	0.07767572	0.07767572	0.85
##	rs144873025	2504	1	1	0.08047125	0.08047125	0.84
##	rs184461291	2504	1	1	0.07607827	0.07607827	0.85
##	rs566494525	2504	1	1	0.06709265	0.06709265	0.86
##		P.AB	P.BB	z.HWE			
##	rs555347111	0.1417732	0.002795527	1.930779			
##	rs542617372	0.1313898	0.010383387	-3.271542			
##	rs17763596	0.1273962	0.013977636	-5.548735			
##	rs144873025	0.1417732	0.009584665	-2.102509			
##	rs184461291	0.1449681	0.003594249	1.561671			
##	rs566494525	0.1309904	0.001597444	2.321653			

Heterocigosidad

```
p<-sumSnps$MAF  
hist(1-p^2-(1-p)^2)
```



Heterecigosidad

Heterocigosidad por poblaciones

LD

Ejercicio LD para una población YRI

```
YRI <- pops=="YRI"  
snpsYRI <- NewsnpsSNPstats[YRI,]  
LDYRI <- ld(snpsYRI, stats=c("D.prime", "R.squared"),depth=379)  
image(LDYRI$R.squared, lwd=0)
```

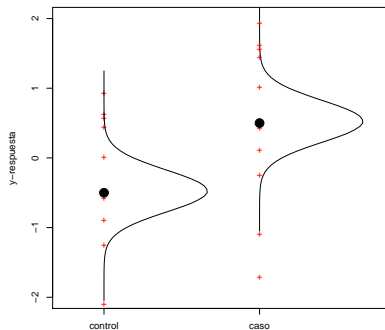

Índice de fijación

El F_{ST} (fixation index) que mide la proporción de variabilidad genética debida a diferencias entre poblaciones

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} \quad (2)$$

σ_S^2 es la varianza entre poblaciones y σ_T^2 es la varianza total

Índice de fijación



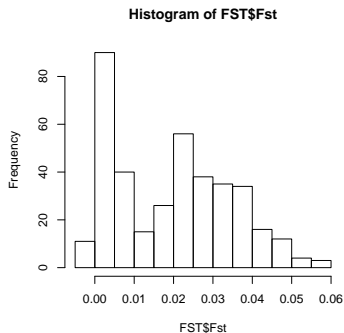
$$F = \frac{\sigma_{\text{caso-control}}^2}{\sigma_{\text{todos}}^2} \quad (3)$$

Qué tan diferenciables son los grupos?

Fst

Distribución de Fst en la región de *MAPT* para los 1000 genomas

```
FST <- Fst(NewsnpsSNPstats,pops)  
hist(FST$Fst)
```



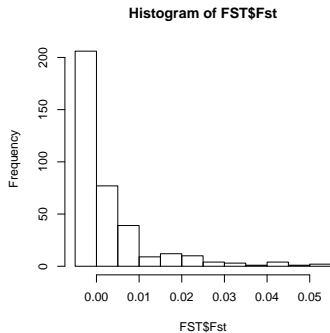
Fst

Fst promedio estimado para las poblaciones del HapMap

	CEU	YRI	JPT
YRI	0.153		
JPT	0.111	0.190	
CHB	0.110	0.192	0.007

Ejercicio: calcular Fst entre YRI y CEU en la region de *MAPT*

```
## [1] 0.003333865
```



Fst

Ejercicio: calcular Fst entre YRI y CEU en la region de *MAPT*

```
CEUandYRI <- pops%in%c("CEU","YRI")
snpsCEUandYRI<-NewsnpsSNPstats[CEUandYRI,]
popsCEUandYRI <- pops[CEUandYRI]
FST <- Fst(snpsCEUandYRI,popsCEUandYRI)
hist(FST$Fst)
mean(FST$Fst, na.rm=TRUE)
```