

Introducción a la Genómica

UNAL nov 2017

Alejandro Cáceres
ISGlobal, Barcelona

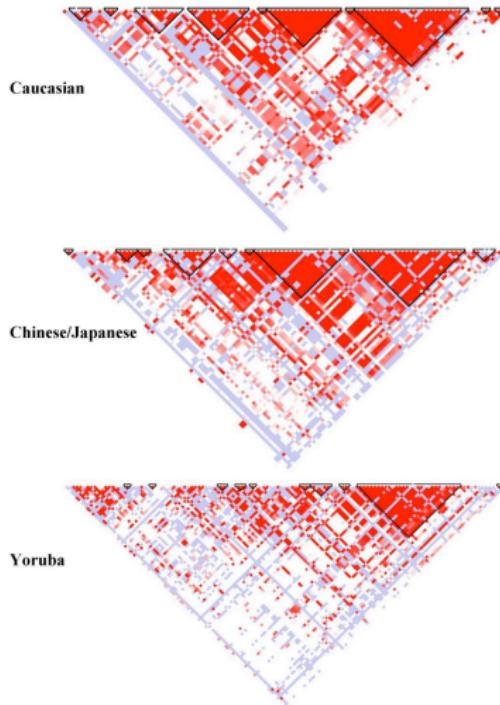
November 11, 2017

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.

Con `snpStats` se puede calcular el LD entre dos SNPs

```
library(snpStats)

load("datos/NewsnpsSNPstats.RData")
ls()

## [1] "NewsnpsSNPstats"
```

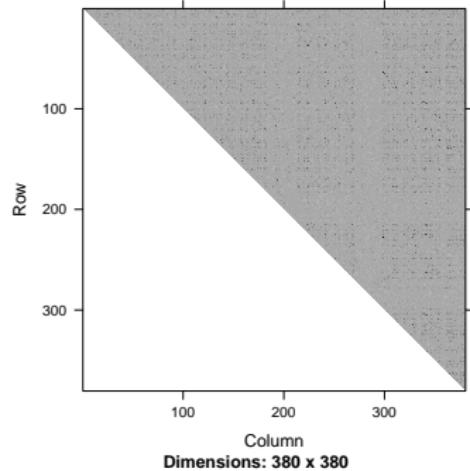
Con.snpStats se puede calcular el LD entre dos SNPs

```
snps <- NewsnpsSNPstats[,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 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 ...

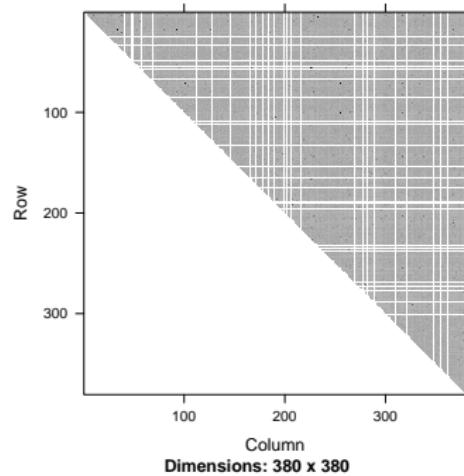
GBR <- pops=="GBR"
head(GBR)

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

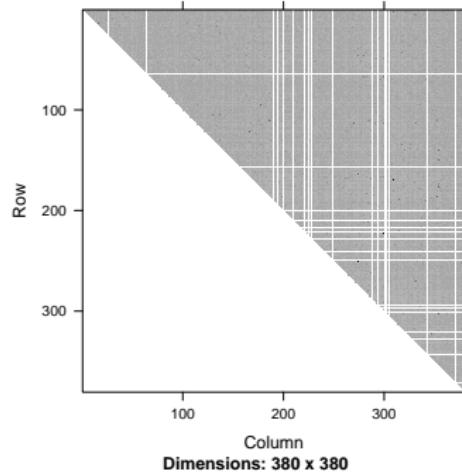
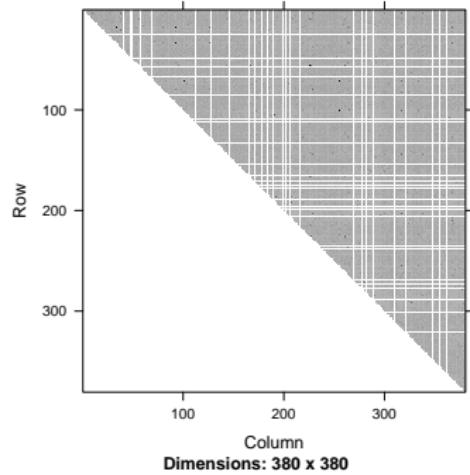
LD

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



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)
```

Análisis de mixtura de haplotipos

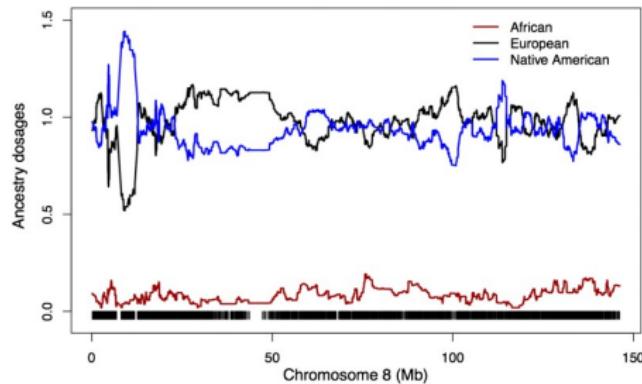
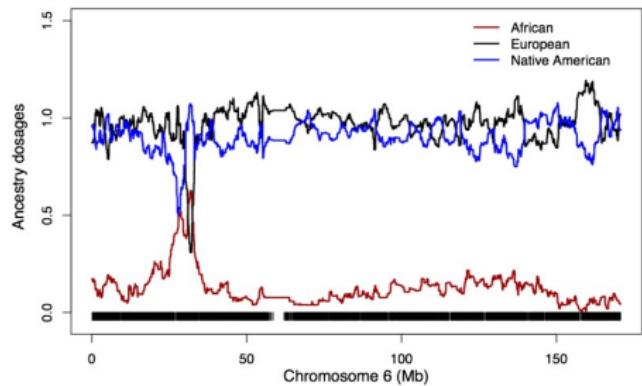
- ▶ En un población mestiza los bloques de LD se mezclan de acuerdo con la recombinación de los individuos. Para analizar estas poblaciones hace falta de métodos que ayuden ver los patrones de mixtura.
- ▶ Structure of haplotypes and local ancestry (Yongtau Guan)
Genetics. 2014, 196(3): 625642.

Software para hacer los análisis

The screenshot shows a Mozilla Firefox browser window with the title bar "elai-manual.pdf - Mozilla Firefox". The address bar contains the URL "www.haplotype.org/download/elai/elai-manual.pdf". The page content is the "ELAI user manual" by Yongtao Guan from Baylor College of Medicine, Version 1.01, dated 6 April 2016. Below the title, there is a table of contents.

Contents	
1 Copyright	2
2 What ELAI Can Do	2
3 A simple example	2
4 Input file formats	3
4.1 Genotype file format	3
4.2 Phased genotype file format	3
4.3 SNP position file format	4
5 Running ELAI	4
6 Output Files	6
6.1 Log file: <code>prefix.log</code>	6
6.2 SNP information file: <code>prefix.snpinfo.txt</code>	6
6.3 Mean level ancestry: <code>prefix.mca</code>	6

Mapa cromosómico para identificar la dosis de ancestría en la población



Heterocigosidad

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

se calcula como

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

Heterocigosidad

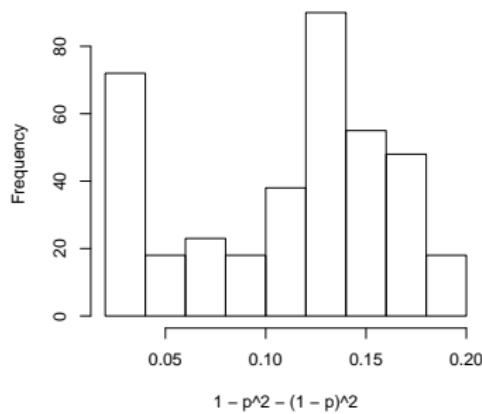
```
sumSnps <- col.summary(NewsnpsSNPstats)
head(sumSnps)
```

	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)
```

Histogram of $1 - p^2 - (1 - p)^2$



Heterocigosidad

Heterocigosidad en una población

```
lv <- levels(pops)
x <- lv[1]
x

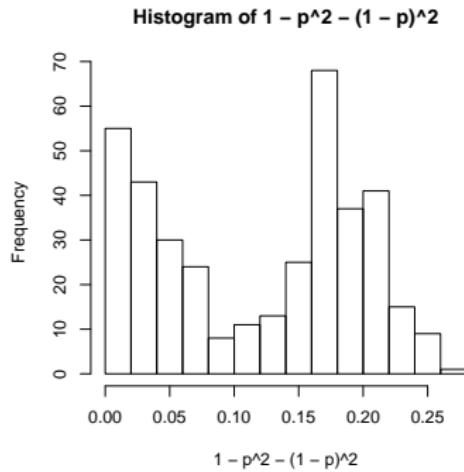
## [1] "ACB"

whichpop <- pops==x
POPsnps <- NewsnpsSNPstats[whichpop,]
sumSnps <- col.summary(POPsnps)
p <- sumSnps$MAF
```

Heterocigosidad

Heterocigosidad en una población

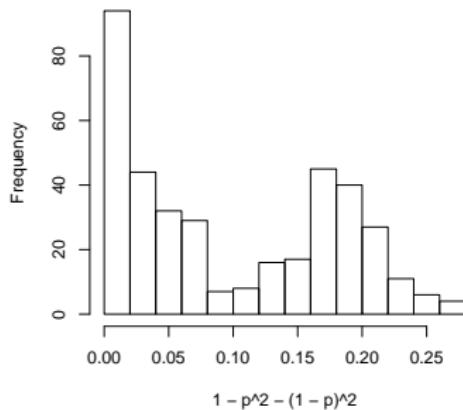
```
hist(1-p^2-(1-p)^2)
```



Histograma de heterocigosidad para YRI

```
## [1] "YRI"
```

Histogram of $1 - p^2 - (1 - p)^2$



Histograma de heterocigosidad para YRI

```
lv <- levels(pops)
x <- lv[26]
x
whichpop <- pops==x
POPsnps <- NewsnpsSNPstats[whichpop,]
sumSnps <- col.summary(POPsnps)
p <- sumSnps$MAF
hist(1-p^2-(1-p)^2)
```

Heterocigosidad

Heterocigosidad por poblaciones

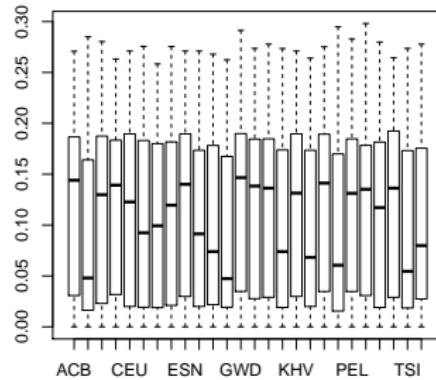
```
hetpop<-sapply(levels(pops), function(x)
{
  whichpop <- pops==x
  POPsnps <- NewsnpsSNPstats[whichpop,]
  sumSnps <- col.summary(POPsnp)
  p <- sumSnps$MAF
  1-p^2-(1-p)^2
})
hetpop[1:5,1:5]

##          ACB         ASW         BEB         CDX         CEU
## [1,] 0.1948242 0.01625907 0.1965251 0.16608857 0.1569738
## [2,] 0.1440430 0.12254770 0.2401298 0.07243612 0.1735027
## [3,] 0.2341580 0.07860790 0.2230936 0.20857903 0.1226916
## [4,] 0.1614041 0.15049718 0.1592077 0.16608857 0.2130395
## [5,] 0.1614041 0.10816985 0.2230936 0.13920684 0.1816141
```

Heterocigosidad

Heterocigosidad por poblaciones

```
boxplot(hetpop)
```



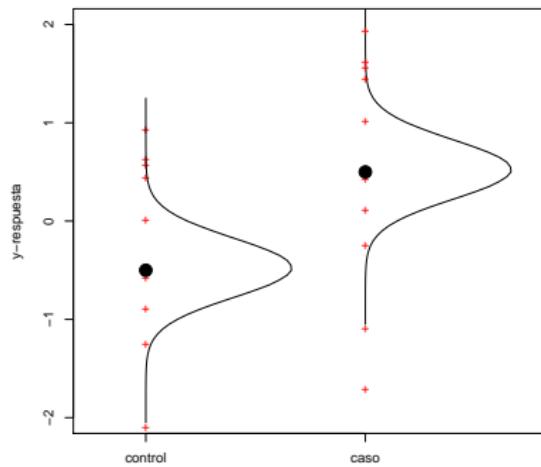
Índice de fijación

El Fst (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



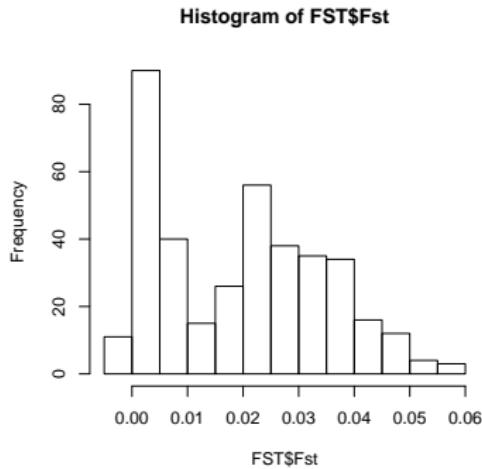
$$F = \frac{\sigma_{caso-control}^2}{\sigma_{todos}^2} \quad (3)$$

Qué tan differentiables son los grupos?

Fst

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

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



Fst

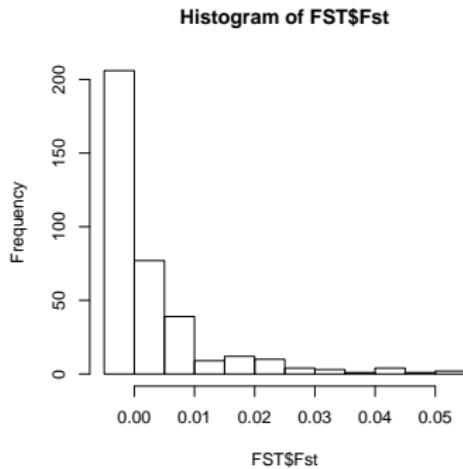
Fst promedio estimado para las poblaciones del HapMap

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

Fst

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

```
## [1] 0.0033333865
```



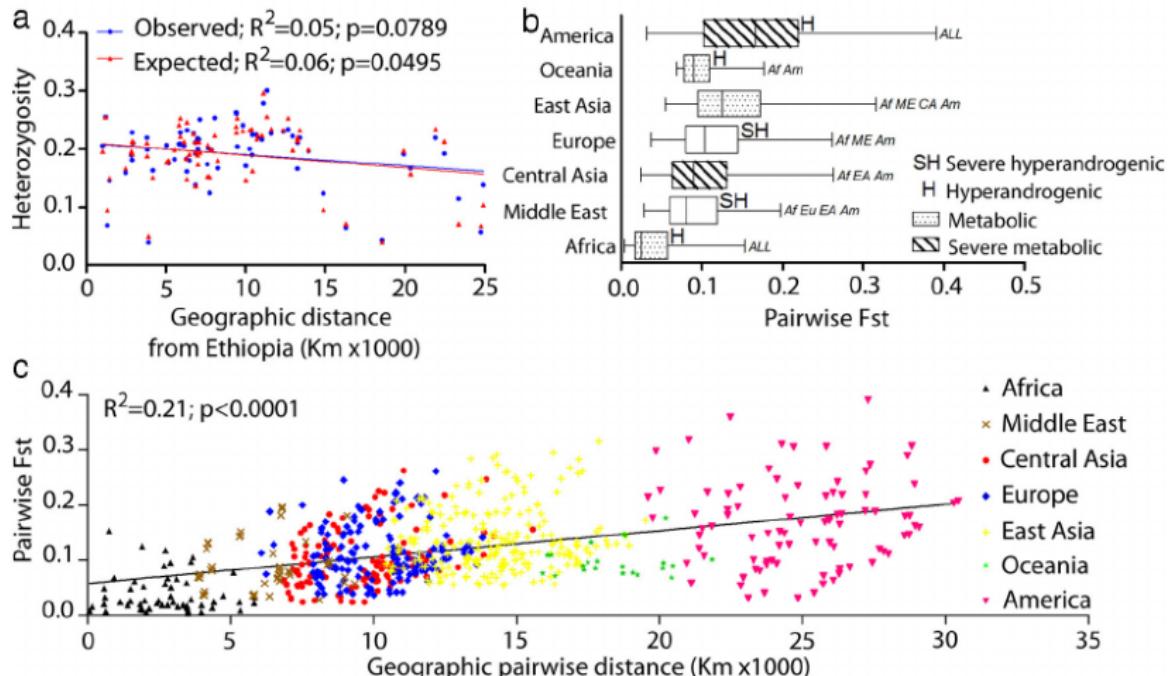
Fst

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)
```

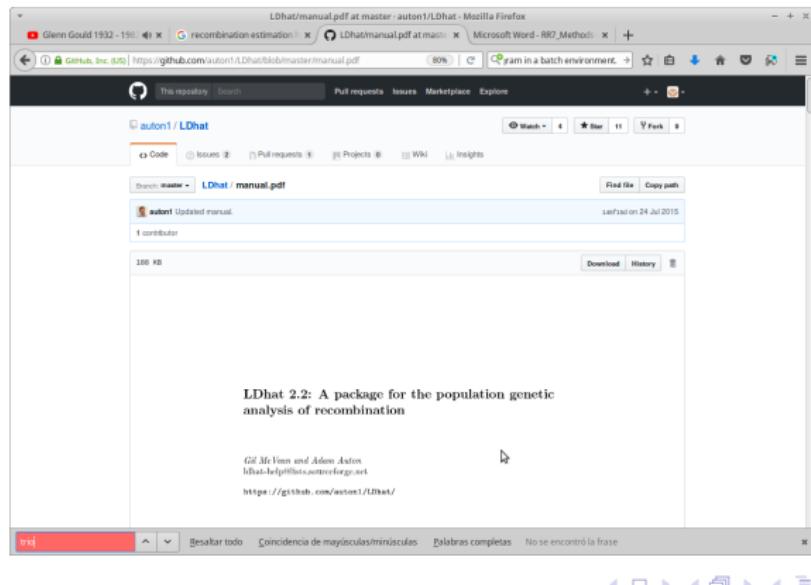
Fst y heterocigosidad en 1000 Genomas

Livio Casarini et al. Clin Endocrinol Metab, 2014, 99(11):E2412



Recombinación

- ▶ Otro análisis relevante que se puede hacer a partir de los patrones de LD es el mapa de recombinación
- ▶ LDhat es un software que ha sido utilizado para derivar los mapas de recombinación de CEU, YRI y CHB-JPT de HapMap II



Recombinación

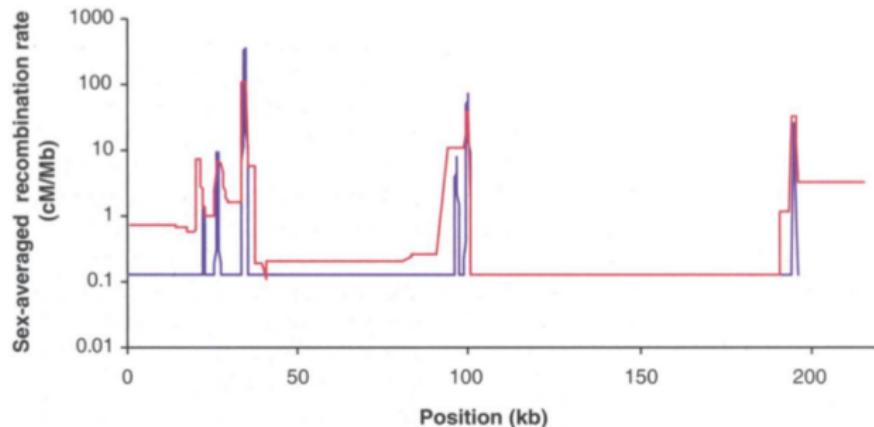


Fig. 2. Comparison between estimates of local recombination rates from population genetic data (red) and sperm analysis (blue) in the HLA region; data from (3). To convert the male crossing over rates to sex-averaged rates, we used the previous observation that the female crossing-over rate in this region is about four times that of males (42).

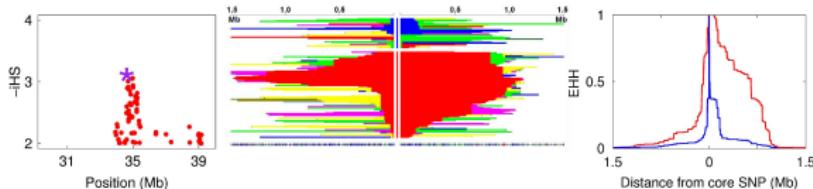
McVean et al. 2004. Science 304: 581-584.

Selección

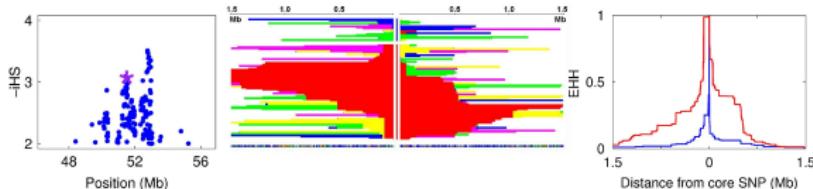
Haplótipos más extensos de lo usual pueden dar señales de selección. iHS propuesta por Voight et al. es una extensión de EHH (extended haplotype homozygosity): los haplotipos alrededor de un SNP en selección tienden a aumentar su homocigosidad rápidamente en comparación con la deriva genética.

Selección

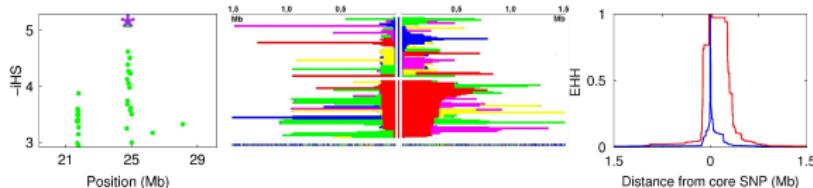
(a) East Asians, rs6060371 (in SPAG4), $p_d = 0.742$, 2.3 cM/Mb



(b) CEPH, rs996521 (in SNTG1), $p_d = 0.808$, 0.28 cM/Mb



(c) Yoruba, rs995647 (in NCOA1), $p_d = 0.492$, 0.62 cM/Mb

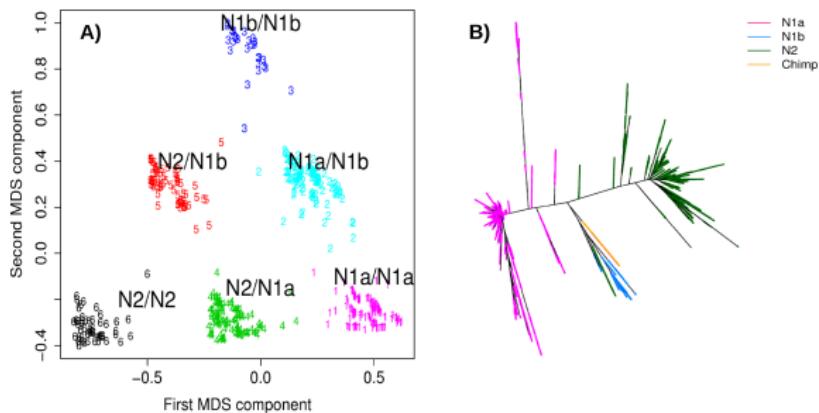


Voight et al. 2006. PLoS Genetics.

Filogenia

Si tenemos una region específica en donde la estructura haplotípica es interesante podemos calcular la filogenia entre haplotipos (R package phyclust)

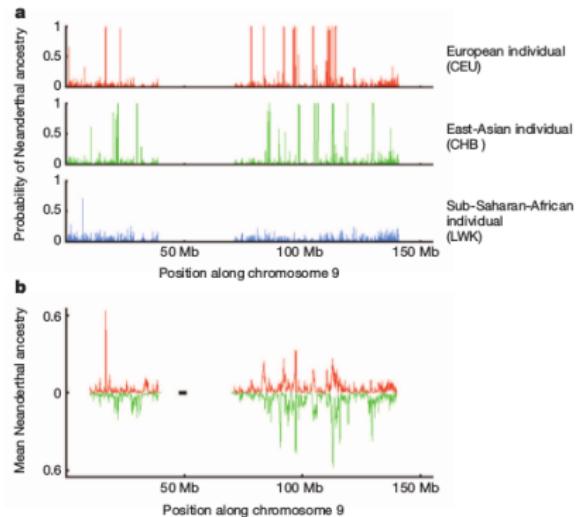
15q24 es una región con multiples haplotipos, entre duplicaciones segmentales que predisponen a duplicaciones y delecciones.



Caceres et al 2016 PLoS One

Ancestría Neardental

Otra cosa interesante que se puede hacer es derivar el contenido de alelos neandertales en las poblaciones



Sankararaman et al 2014. Nature

Ancestría Neardental

Table 1 | Genome-wide estimates of Neanderthal ancestry

Region	Population	Number of individuals	Neanderthal ancestry on autosomes (%)	Neanderthal ancestry on the X chromosome (%)
Europe	CEU	85	1.17 ± 0.08	0.21 ± 0.17
	FIN	93	1.20 ± 0.07	0.19 ± 0.14
	GBR	89	1.15 ± 0.08	0.20 ± 0.15
	IBS	14	1.07 ± 0.06	0.23 ± 0.18
	TSI	98	1.11 ± 0.07	0.25 ± 0.20
East Asia	CHB	97	1.40 ± 0.08	0.30 ± 0.21
	CHS	100	1.37 ± 0.08	0.27 ± 0.21
	JPT	89	1.38 ± 0.10	0.26 ± 0.21
America	CLM	60	1.14 ± 0.12	0.22 ± 0.16
	MXL	66	1.22 ± 0.09	0.21 ± 0.15
	PUR	55	1.05 ± 0.12	0.20 ± 0.15
Africa	LWK	97	0.08 ± 0.02	0.04 ± 0.07
	ASW	61	0.34 ± 0.22	0.07 ± 0.11

For each computationally phased genome in each population, we estimated the probability of Neanderthal ancestry at each SNP and the fraction of autosomal and X chromosome SNPs that are confidently of Neanderthal origin in each individual (marginal probability >90%). The table reports the average and standard deviation of this statistic across individuals within each population. ASW, people with African ancestry in Southwest United States; CEU, Utah residents with northern and western European ancestry (from the Centre d'Etude du Polymorphisme Humain (CEPH) collection); CHB, Han Chinese in Beijing, China; CHS, Han Chinese in South China; CLM, Colombians in Medellin, Colombia; FIN, Finnish in Finland; GBR, British from England and Scotland, UK; IBS, Iberian populations in Spain; JPT, Japanese in Tokyo, Japan; LWK, African Luhya in Webuye, Kenya; MXL, people with Mexican ancestry in Los Angeles, California; PUR, Puerto Ricans in Puerto Rico; TSI, Toscani in Italy.

Sankararaman et al 2014. Nature