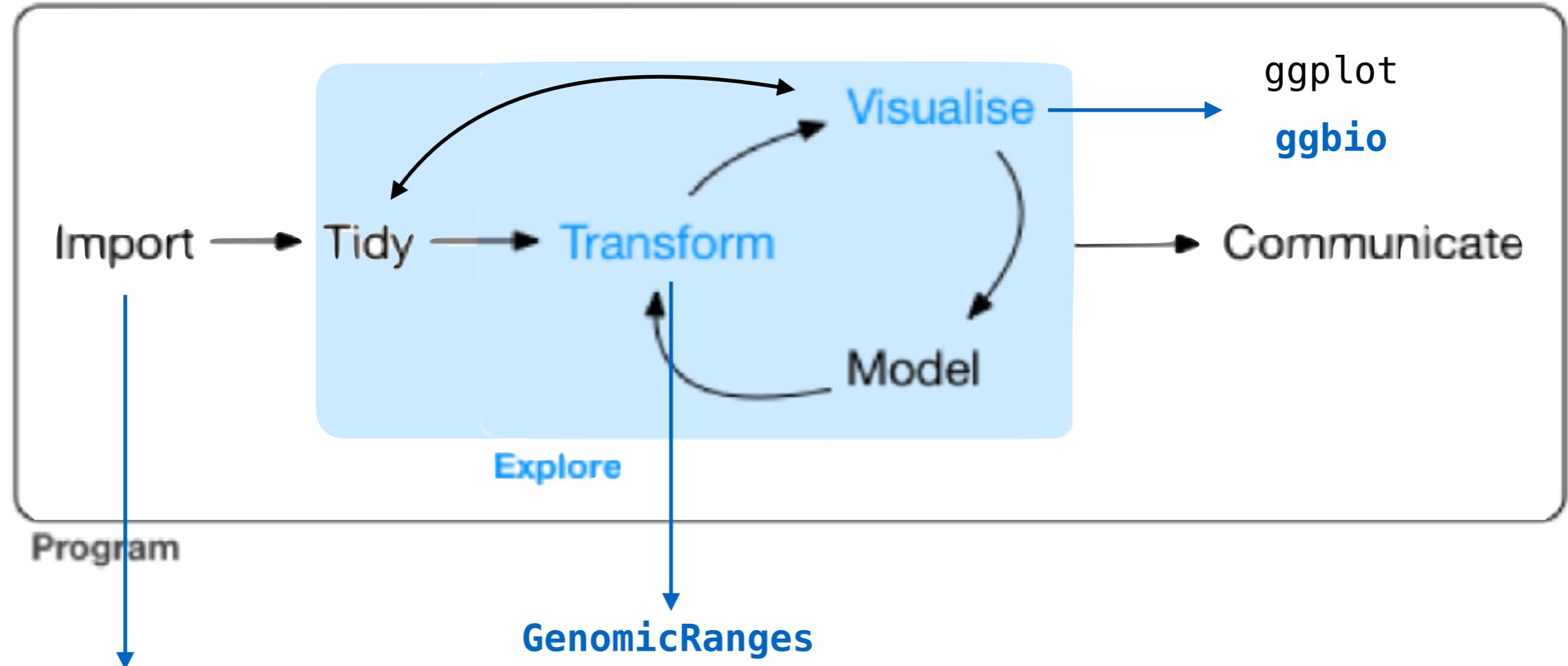


Visualización de datos genómicos

Visualizar para seguir limpiando



```
data.table::fread(), data.table=FALSE)
```

R for Data Science

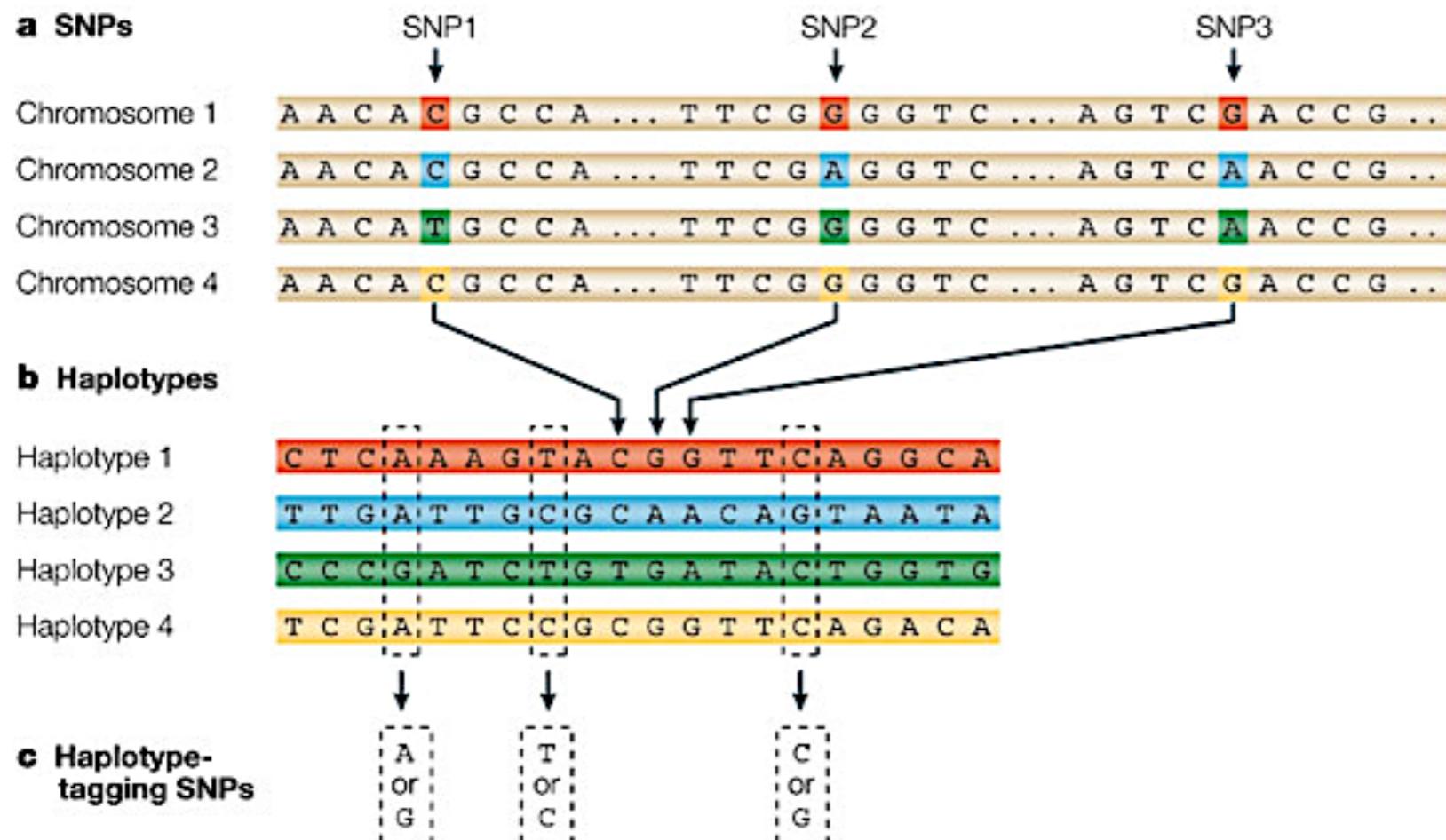
Garrett Grolemund
Hadley Wickham

Visualización de la estructura de poblaciones



Para cada individuo se miden entre miles y millones de variantes

Data



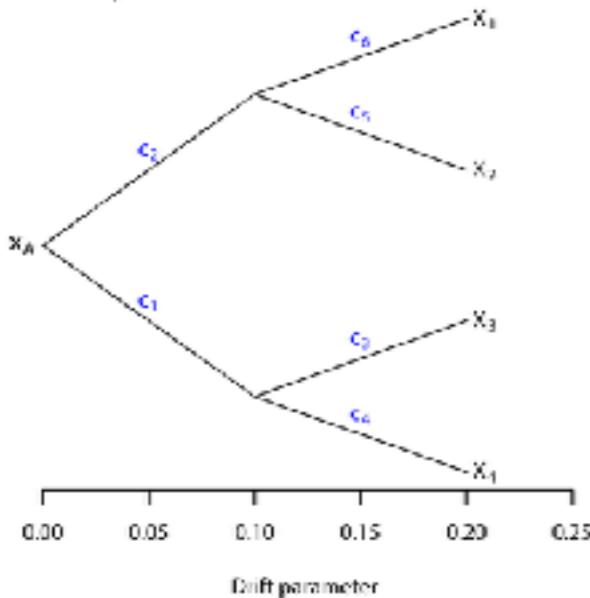
Whole Genome
Sequencing

WGS después de
identificar variantes
(millones)

Chip
(miles)

Un árbol no alcanza para representar las relaciones entre las poblaciones

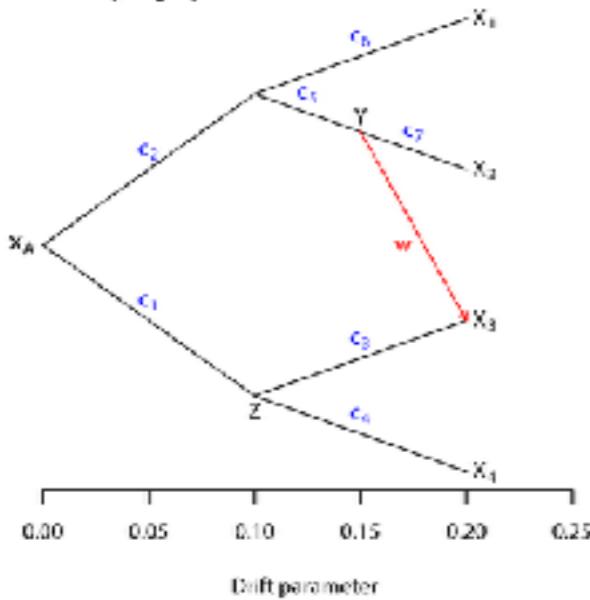
A. Example tree



B. Covariance matrix for tree in A.

X_1	$c_1 + c_5$	c_2	0	0
X_2	c_2	$c_3 + c_6$	0	0
X_3	0	0	$c_1 + c_3$	c_1
X_4	0	0	c_1	$c_3 + c_6$

C. Example graph



D. Covariance matrix for graph in C.

X_1	$c_1 + c_5$	c_2	$w c_2$	0
X_2	c_2	$c_3 + c_5 + c_6$	$w(c_3 + c_5)$	0
X_3	$w c_2$	$w(c_2 + c_5)$	$w^2(c_1 + c_3)$ $+(1-w)^2(c_1 + c_3)$	$(1-w)c_1$
X_4	0	0	$(1-w)c_1$	$c_3 + c_6$

Treemix: intenta representar la estructura usando grafos dirigidos

Genes mirror geography within Europe

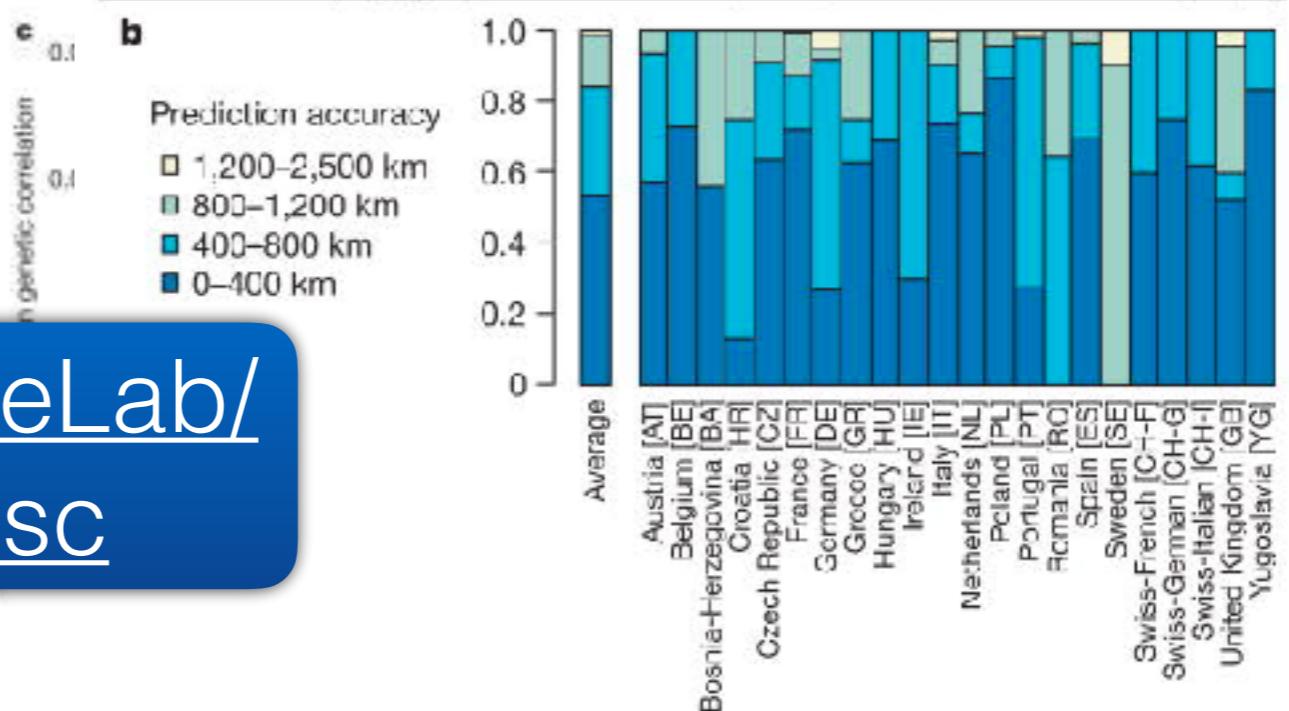
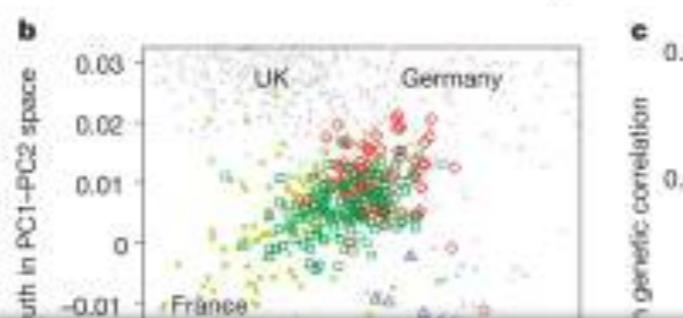
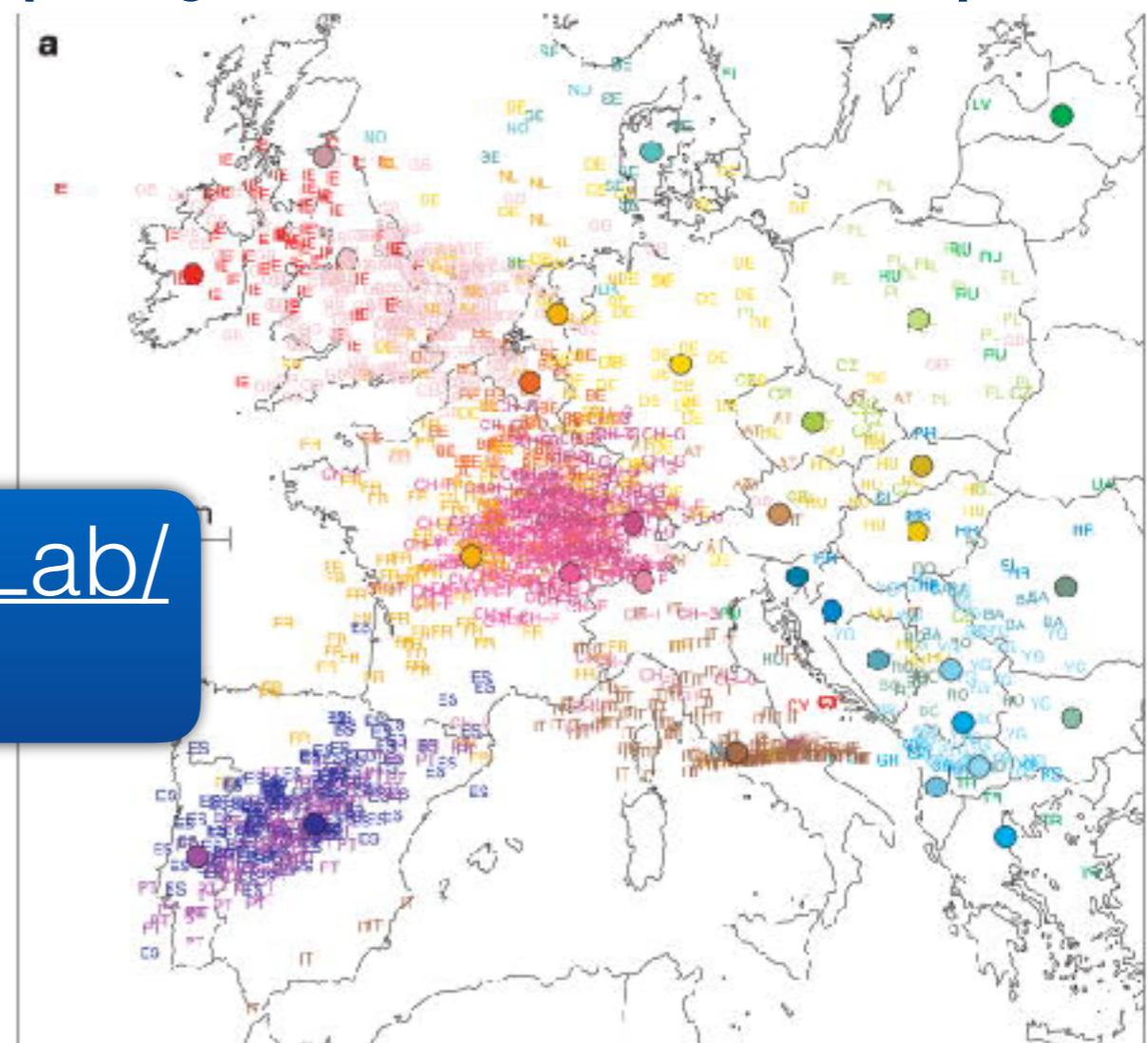
Novembre et al. 2008

An R package to help in plotting
PCA results nicely

[https://github.com/NovembreLab/
PCAviz](https://github.com/NovembreLab/PCAviz)

[https://github.com/NovembreLab/
Novembre_etal_2008_misc](https://github.com/NovembreLab/
Novembre_etal_2008_misc)

△ Italian-speaking Swiss ▲ Italian



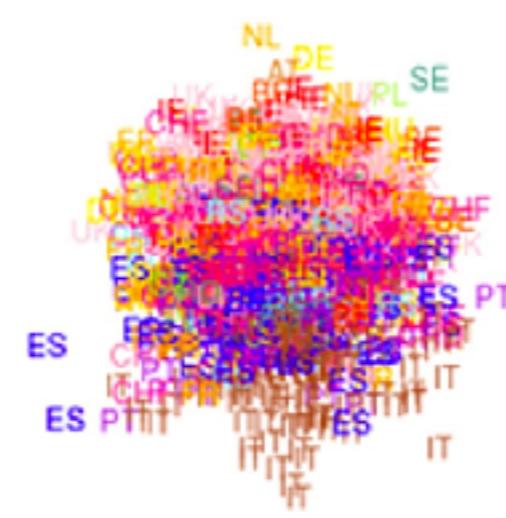
Para ver relaciones a diferentes escalas hay que realizar PCAs distintas

PC1: 1.63% explained



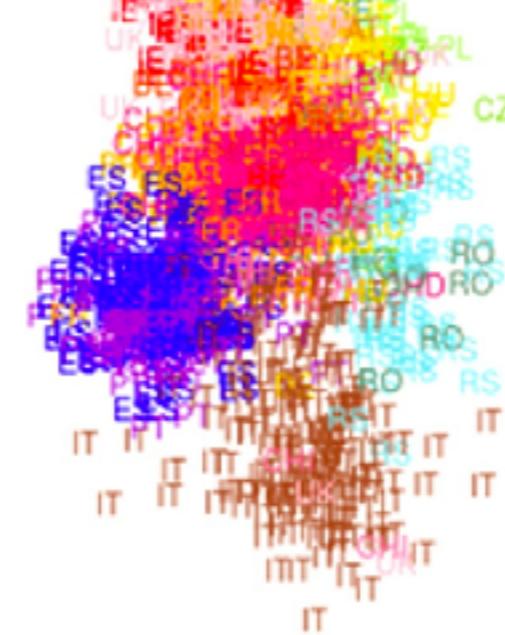
100 loci

PC1: 0.42% explained



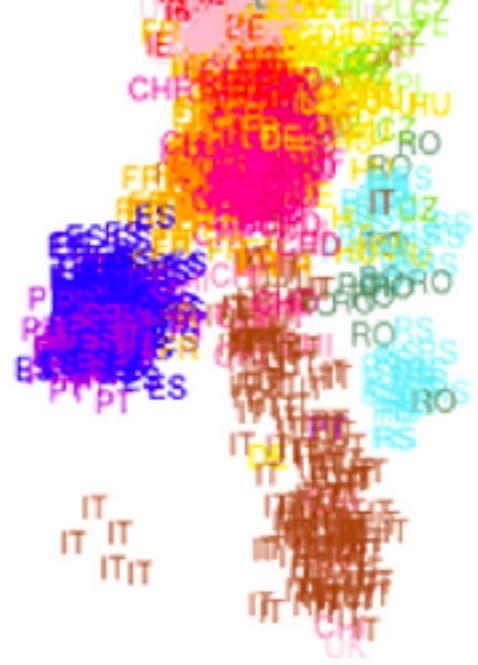
1,000 loci

PC1: 0.31% explained



10,000 loci

PC1: 0.29% explained



100,000 loci

PC2: 1.57% explained



PC2: 0.36% explained



PC2: 0.17% explained

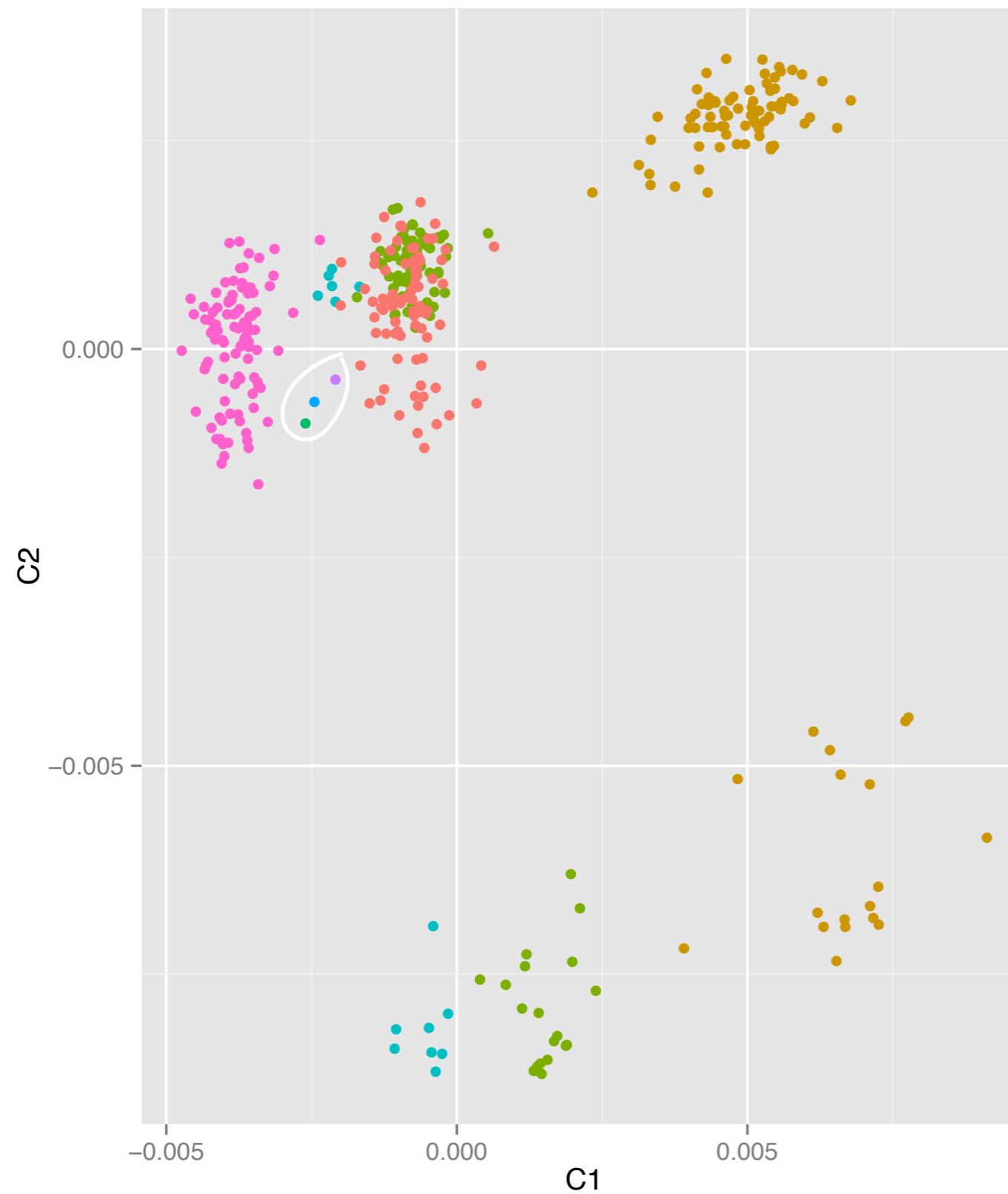


PC2: 0.15% explained

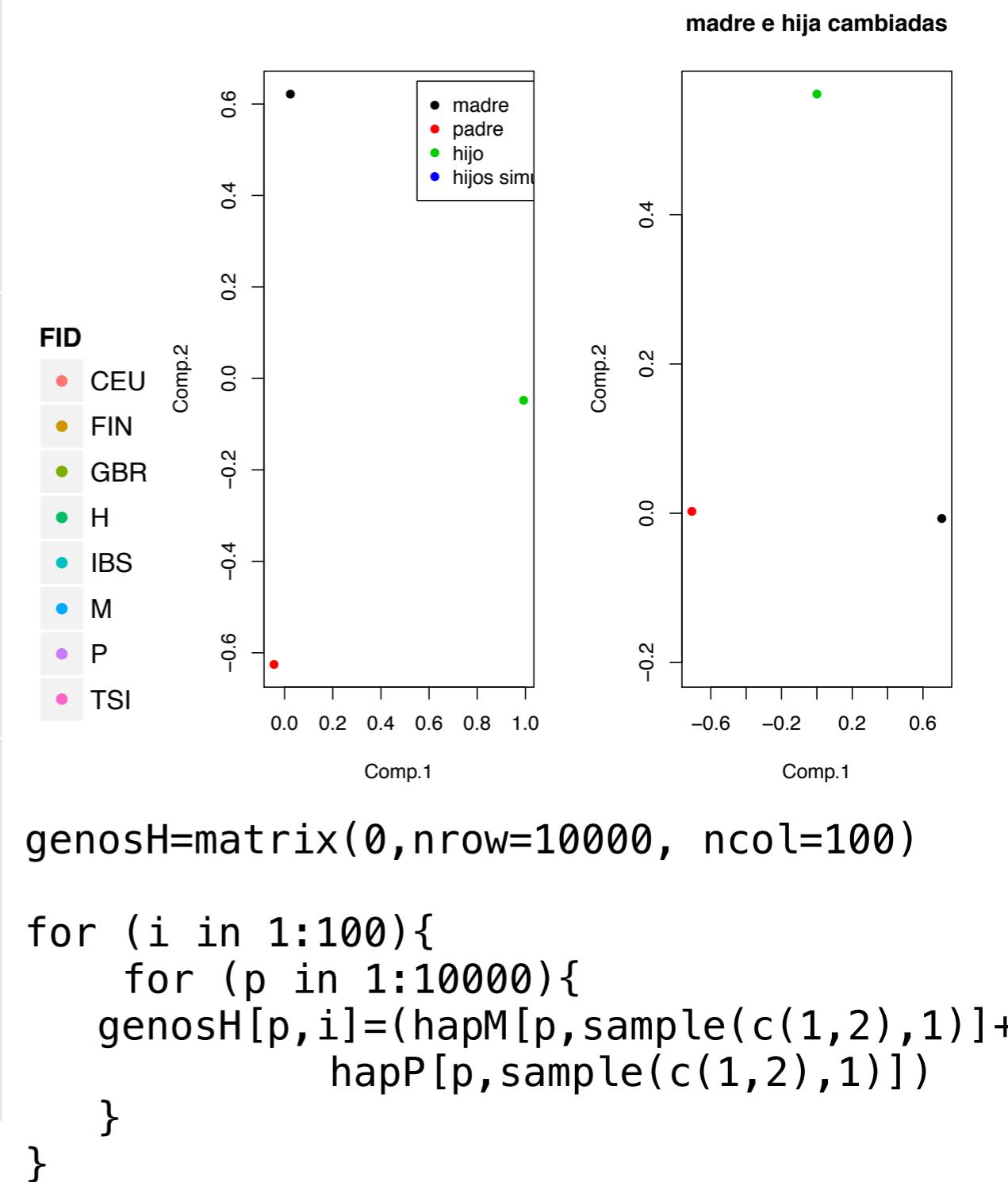
Current Opinion in Genetics & Development

1	Pärnu	Yellow
10	Viljandi	Green
11	Tartu	Dark Blue
12	Valga	Light Green
13	Põlva	Black
14	Võru	Purple

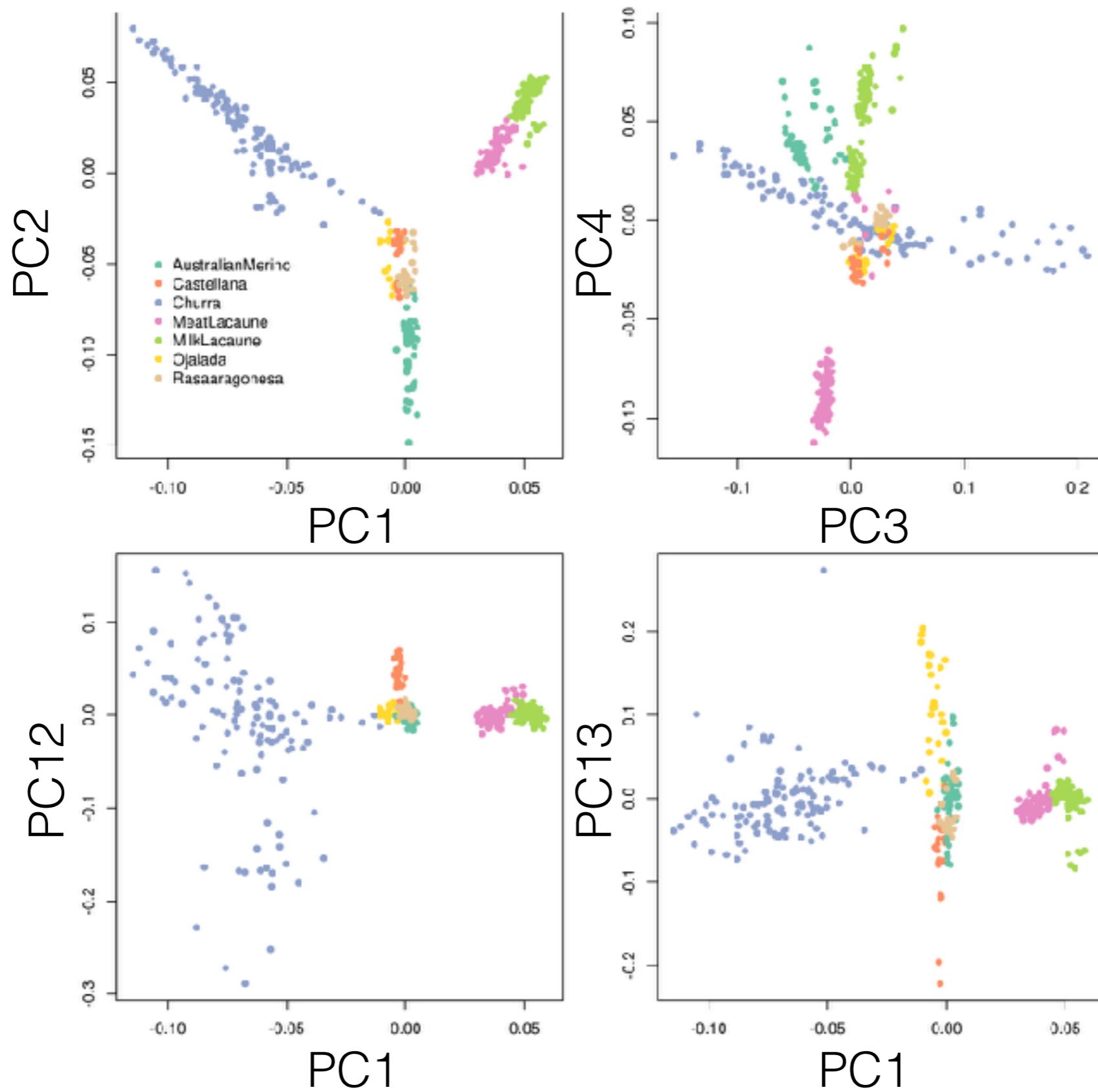
PCA de individuos como herramienta para detectar errores

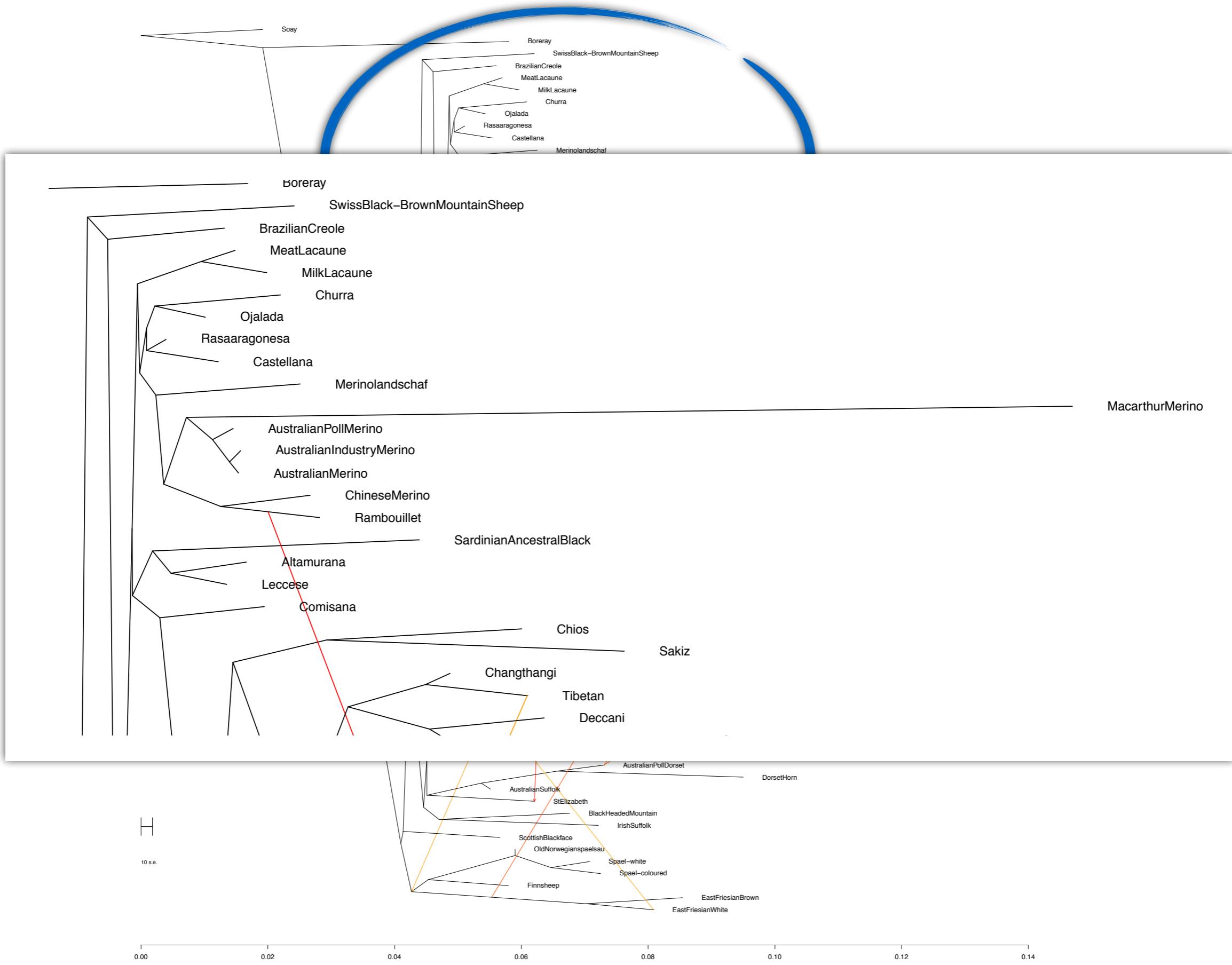


Comparación con hijos simulados

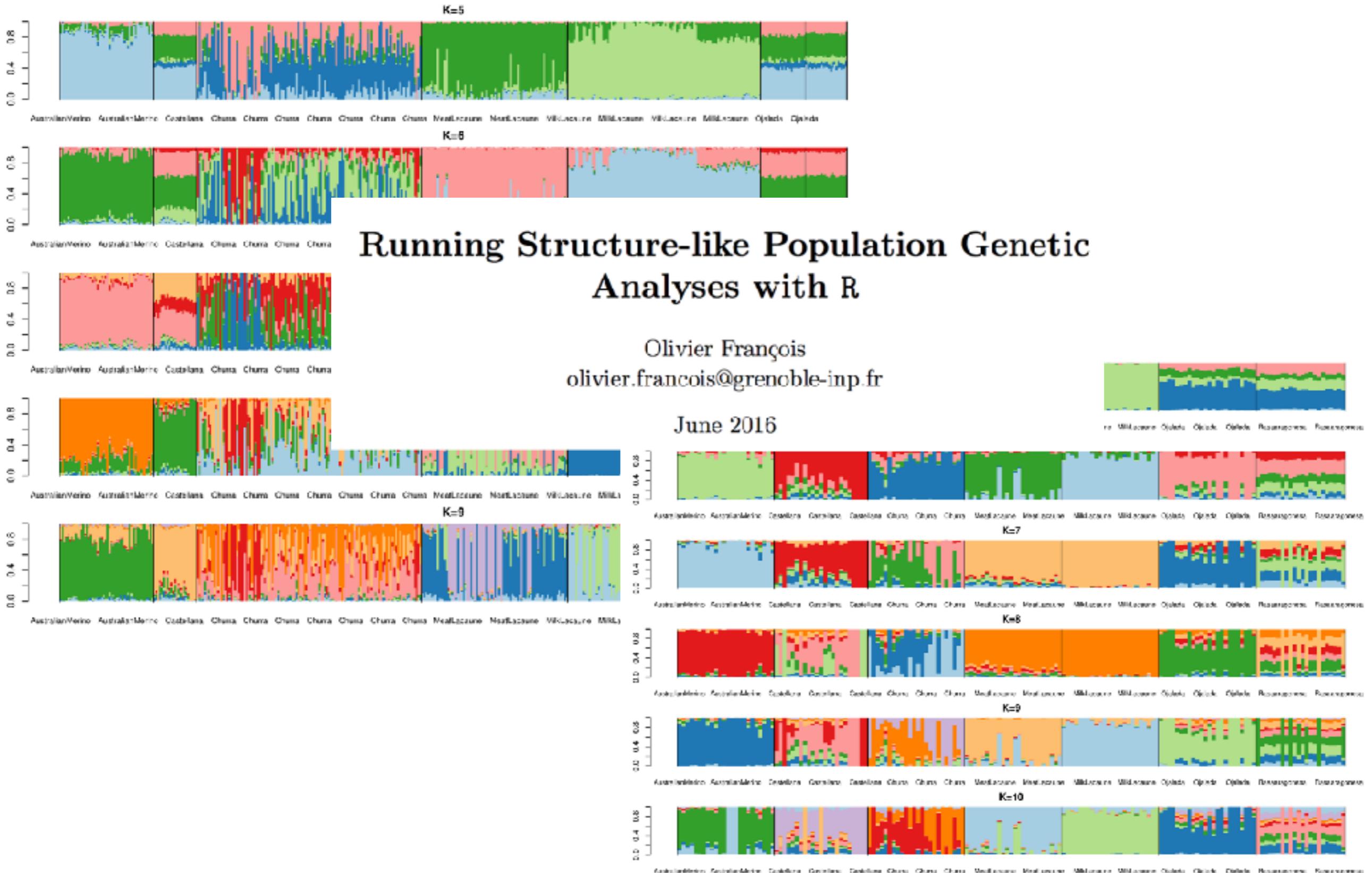


Cómo se relacionan Castellana, Churra y Ojalada entre sí?

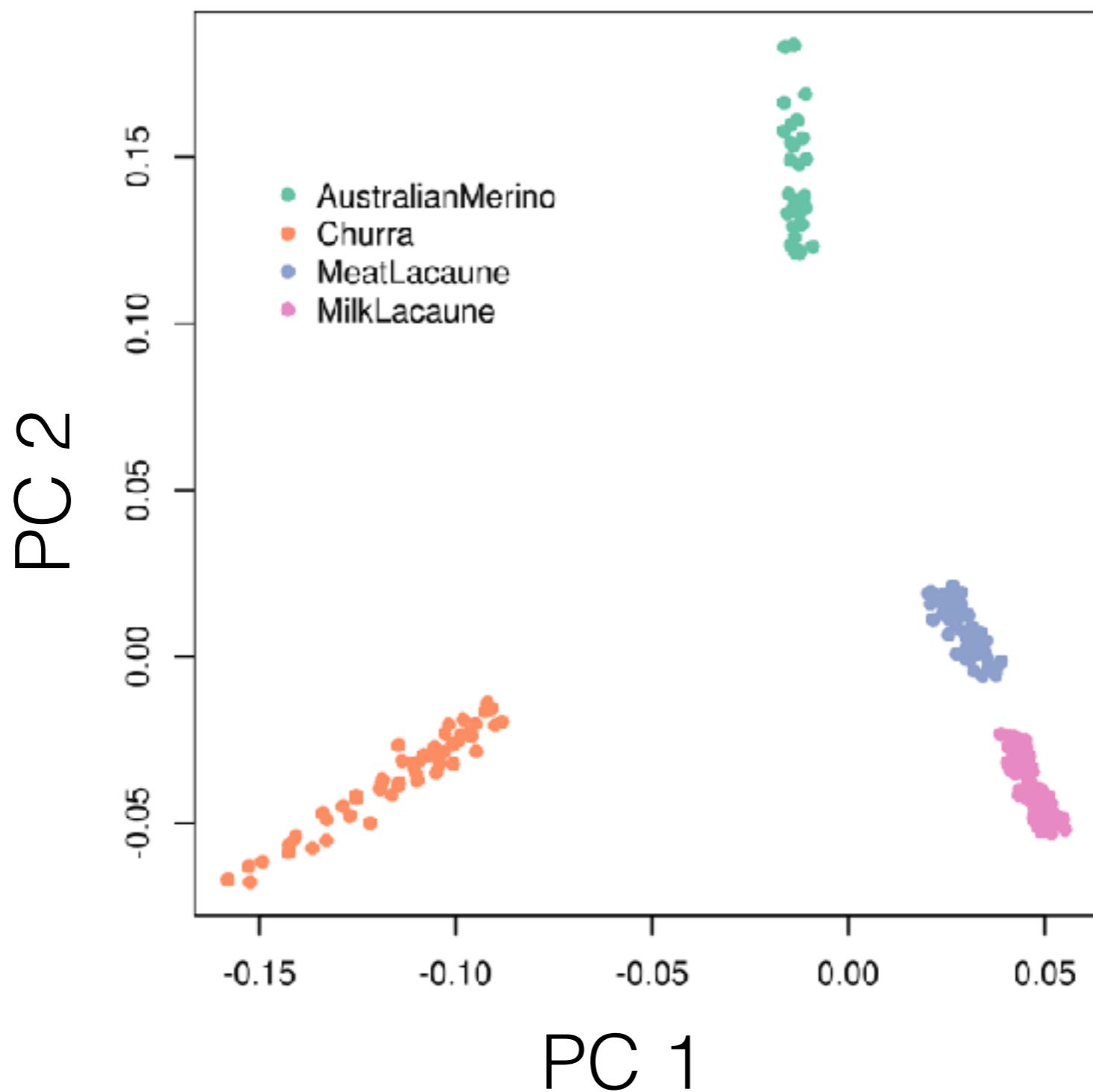




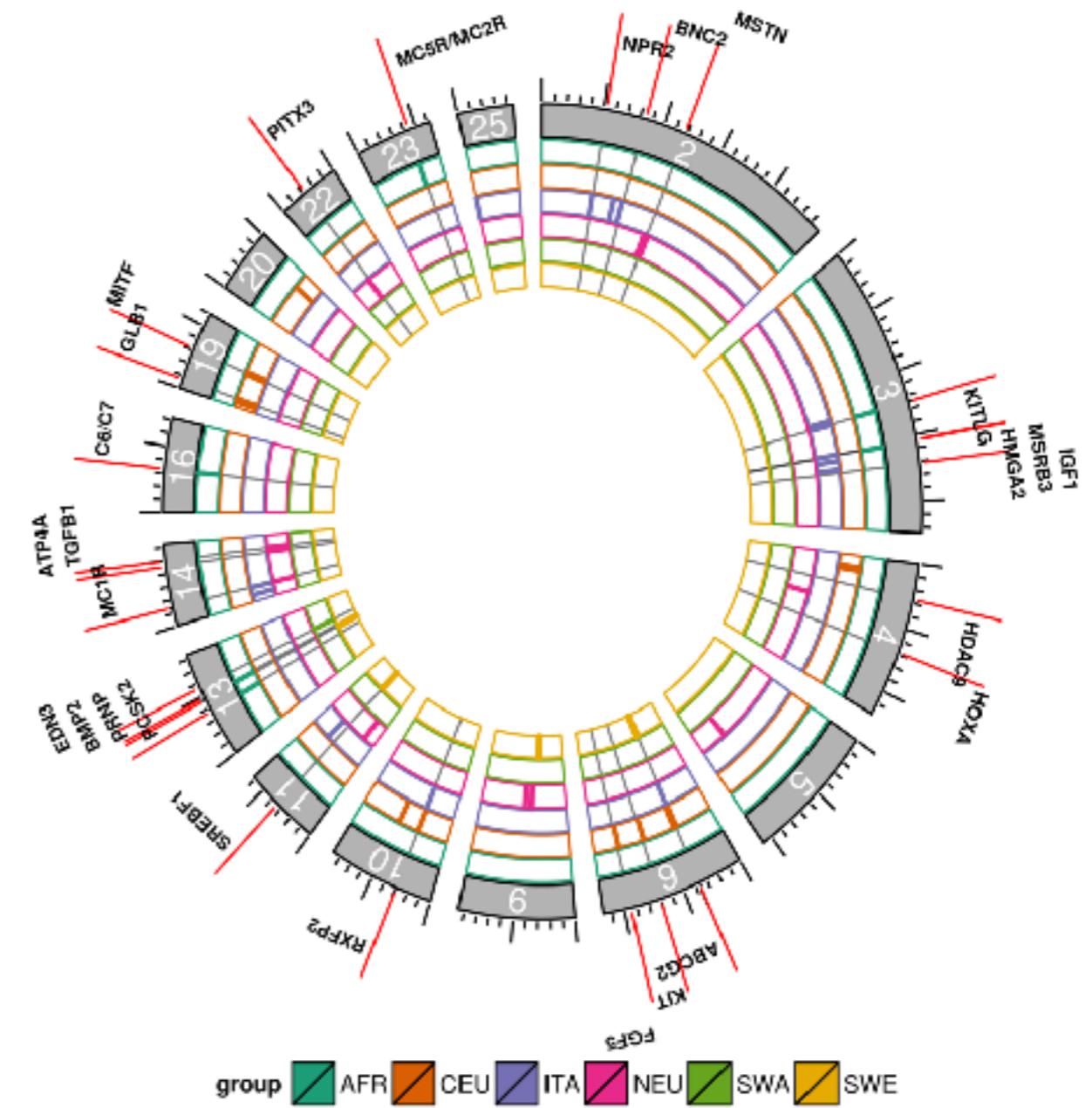
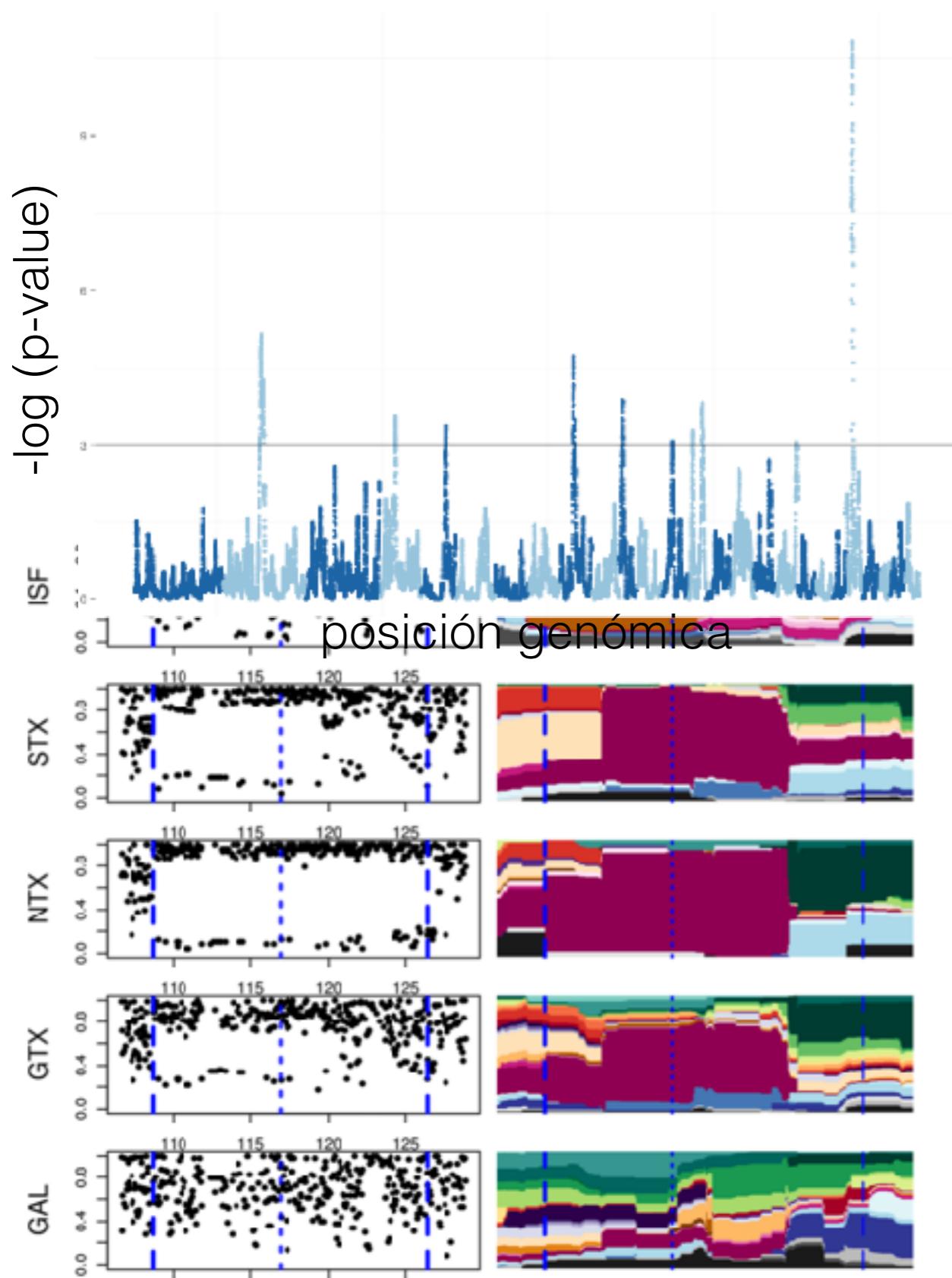
ADMIXTURE: Clustering Local



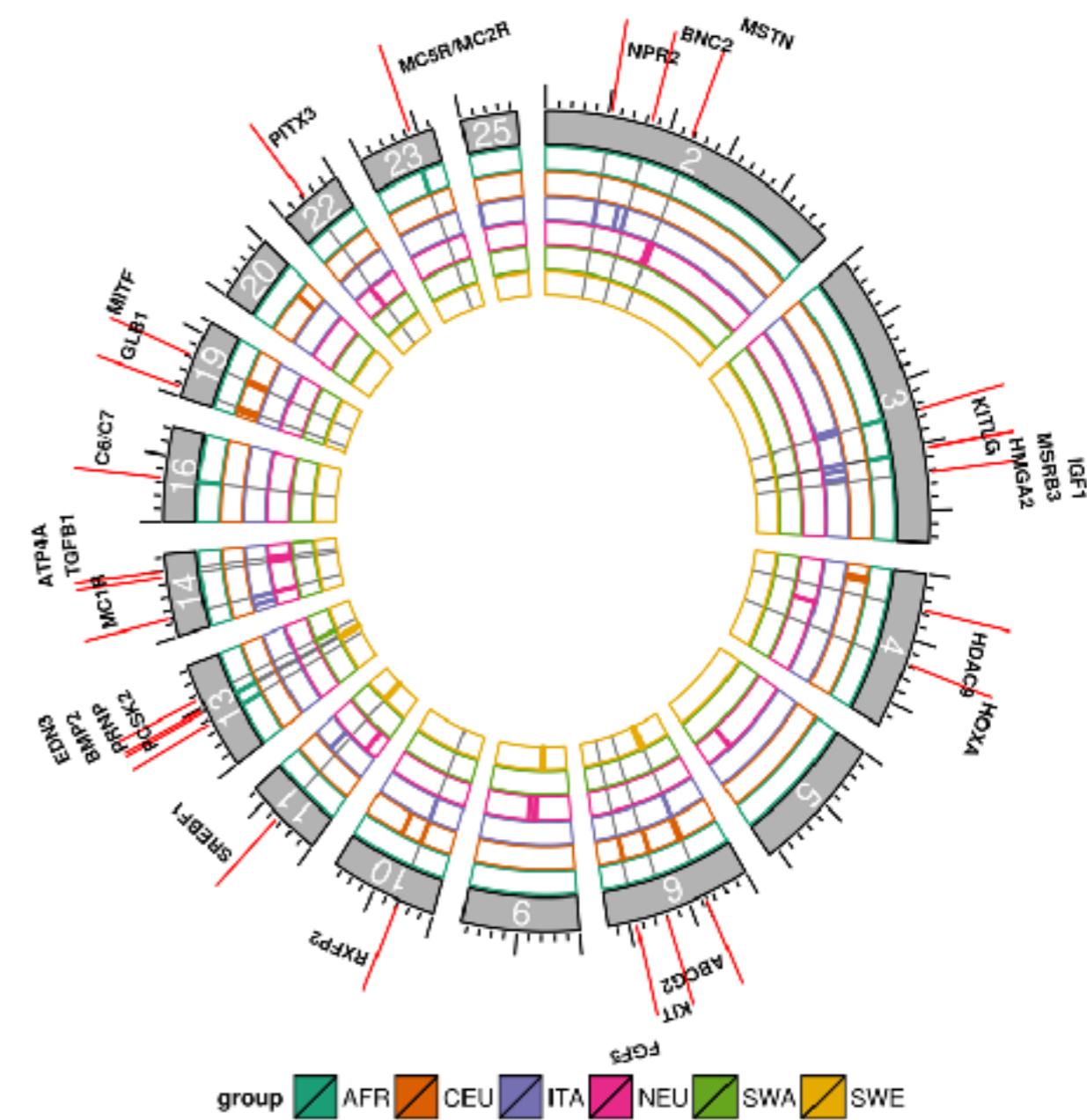
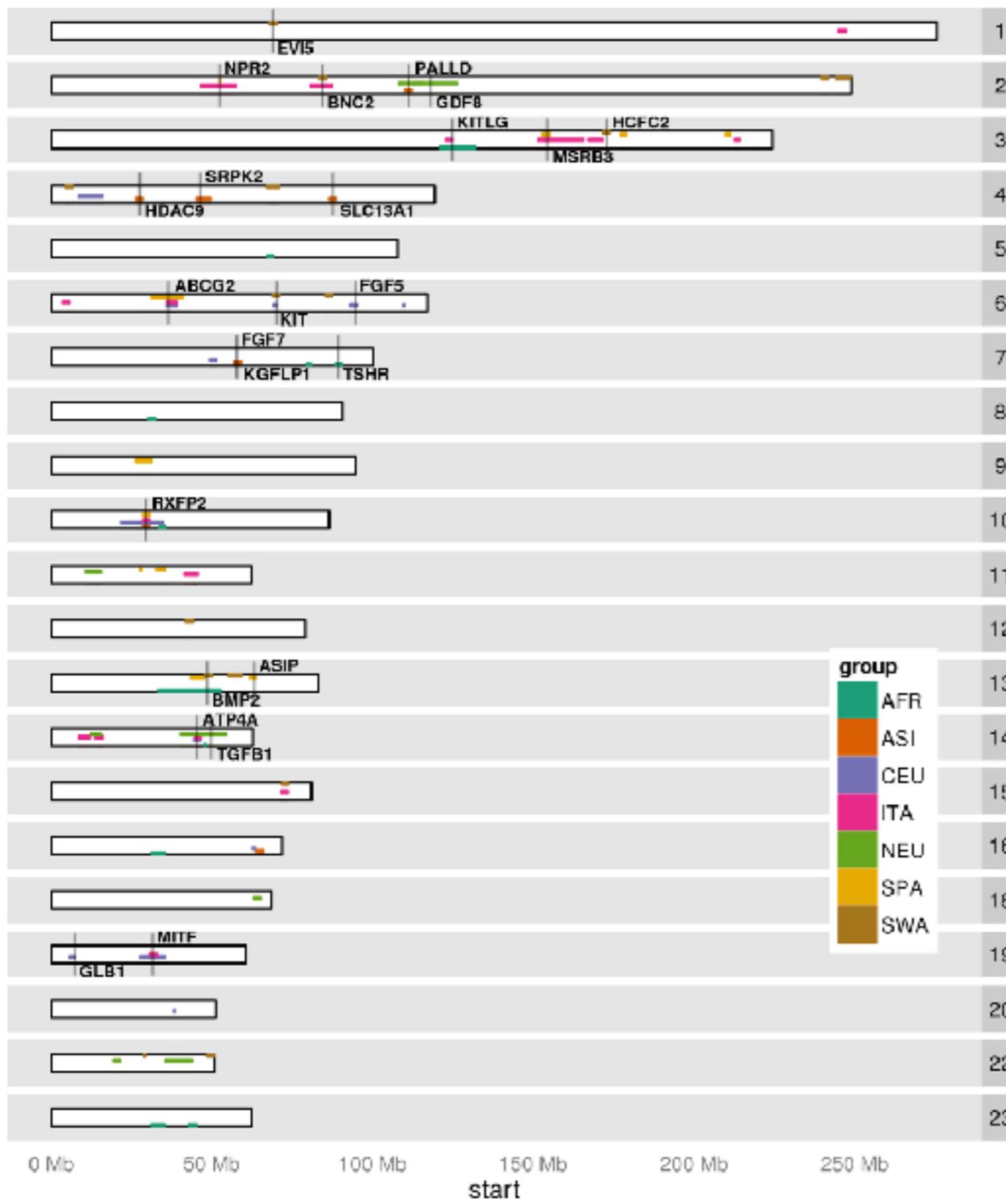
Juego de datos final



hapFLK: Un test de detección de selección basado en la estructura de las poblaciones



ggbio: extensión de ggplot para datos genómicos



Predicción de peso en bovinos

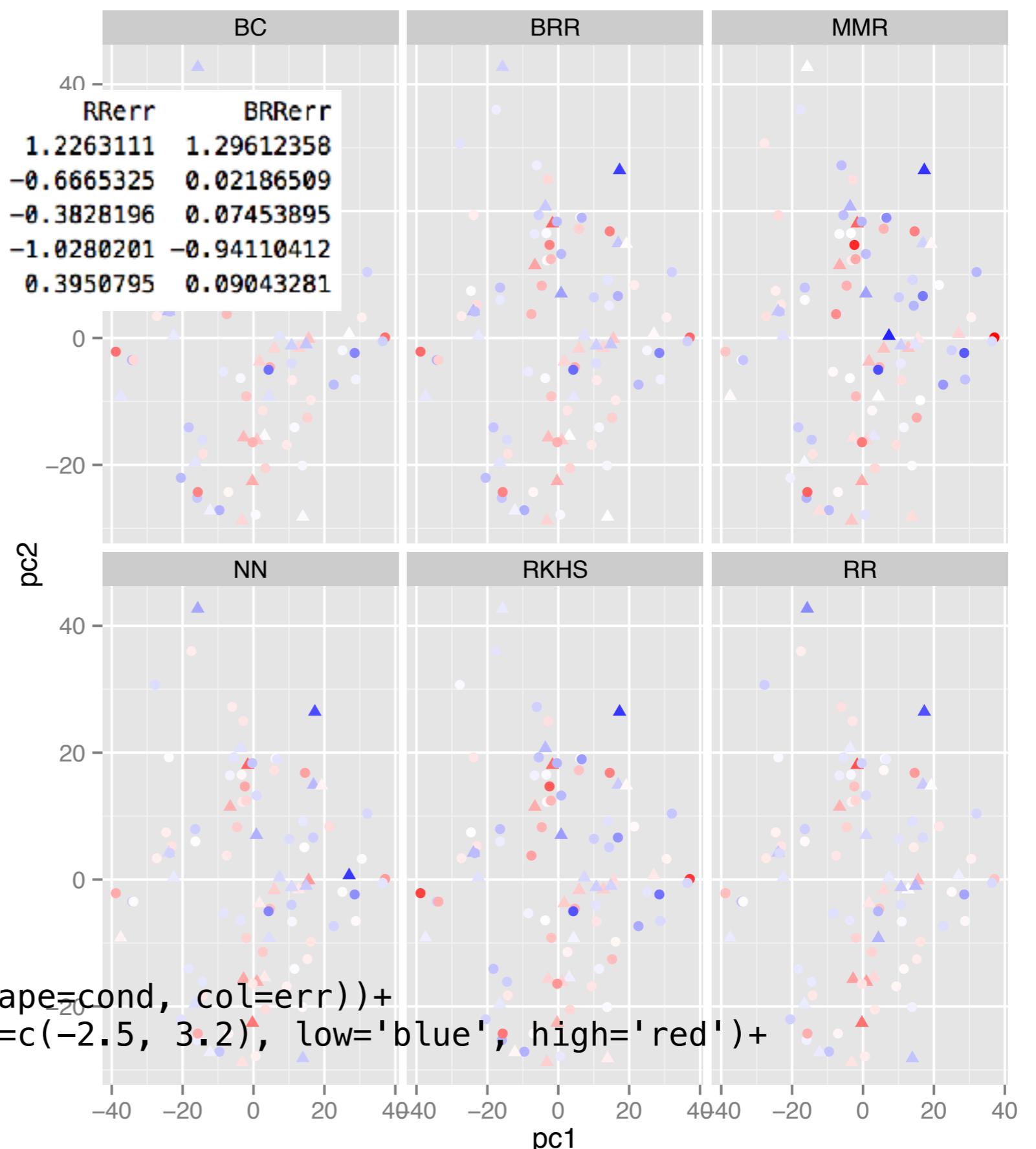
	pc1	pc2	cond	MMRerr	RRerr	BRRerr
1	2.2789088	-2.086765	tst	0.6601694	1.2263111	1.29612358
2	-2.0968440	-14.172238	trn	0.5256149	-0.6665325	0.02186509
3	-14.9727082	-8.166578	trn	-0.3556436	-0.3828196	0.07453895
4	-0.5641897	-15.721319	trn	-0.2475793	-1.0280201	-0.94110412
5	23.6647659	32.574513	trn	0.4357870	0.3950795	0.09043281

86 filas

	pc1	pc2	cond	err	met
1	3.0967912	-15.426518	tst	0.429320647	MMR
2	16.2053443	-9.823568	trn	0.486568827	MMR
3	0.9743909	-16.142086	tst	1.285450352	MMR
4	-5.5013597	19.325279	trn	-0.005661114	MMR
5	-27.7517997	30.709115	trn	-0.381243886	MMR
6	15.2213491	-12.553756	trn	-0.024340599	MMR
7	-15.8889519	-25.244321	trn	-0.381473282	MMR
8	-23.8508043	19.325335	trn	0.462749415	MMR
9	-3.1671041	12.252044	trn	0.346699421	MMR
10	12.7392667	-1.580064	tst	0.754603012	MMR

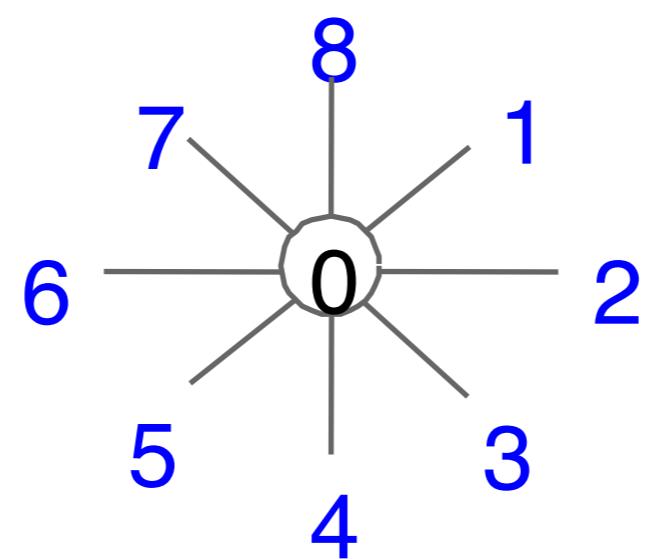
558 filas

```
plot1<-ggplot(data=dfgrafica)+  
  geom_point(aes(x=pc1,y=pc2,shape=cond, col=err))+  
  scale_colour_gradient2(limits=c(-2.5, 3.2), low='blue', high='red')+  
  facet_wrap(~ met)
```

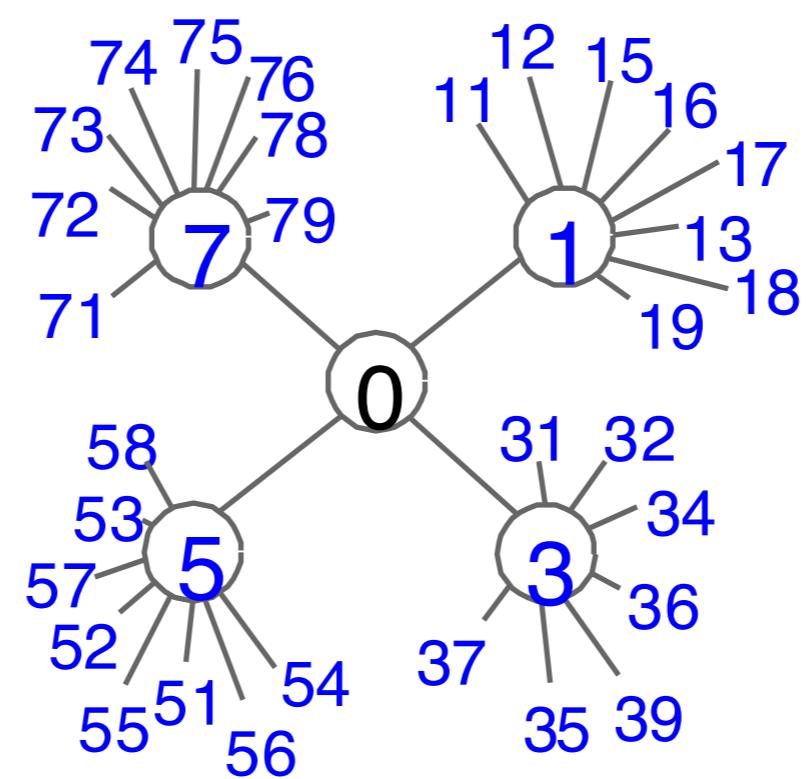


Visualización de evolución de virus

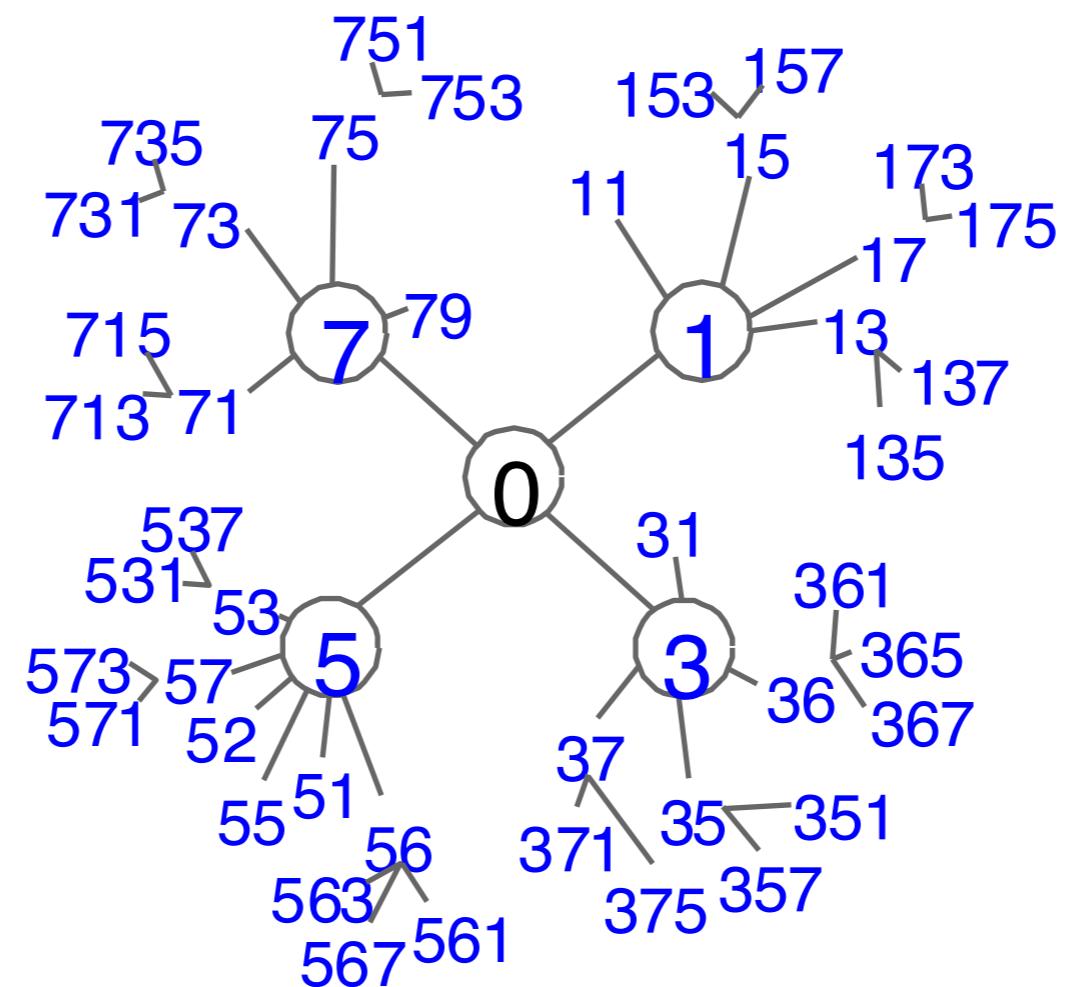
RNA Virus Quasispecies



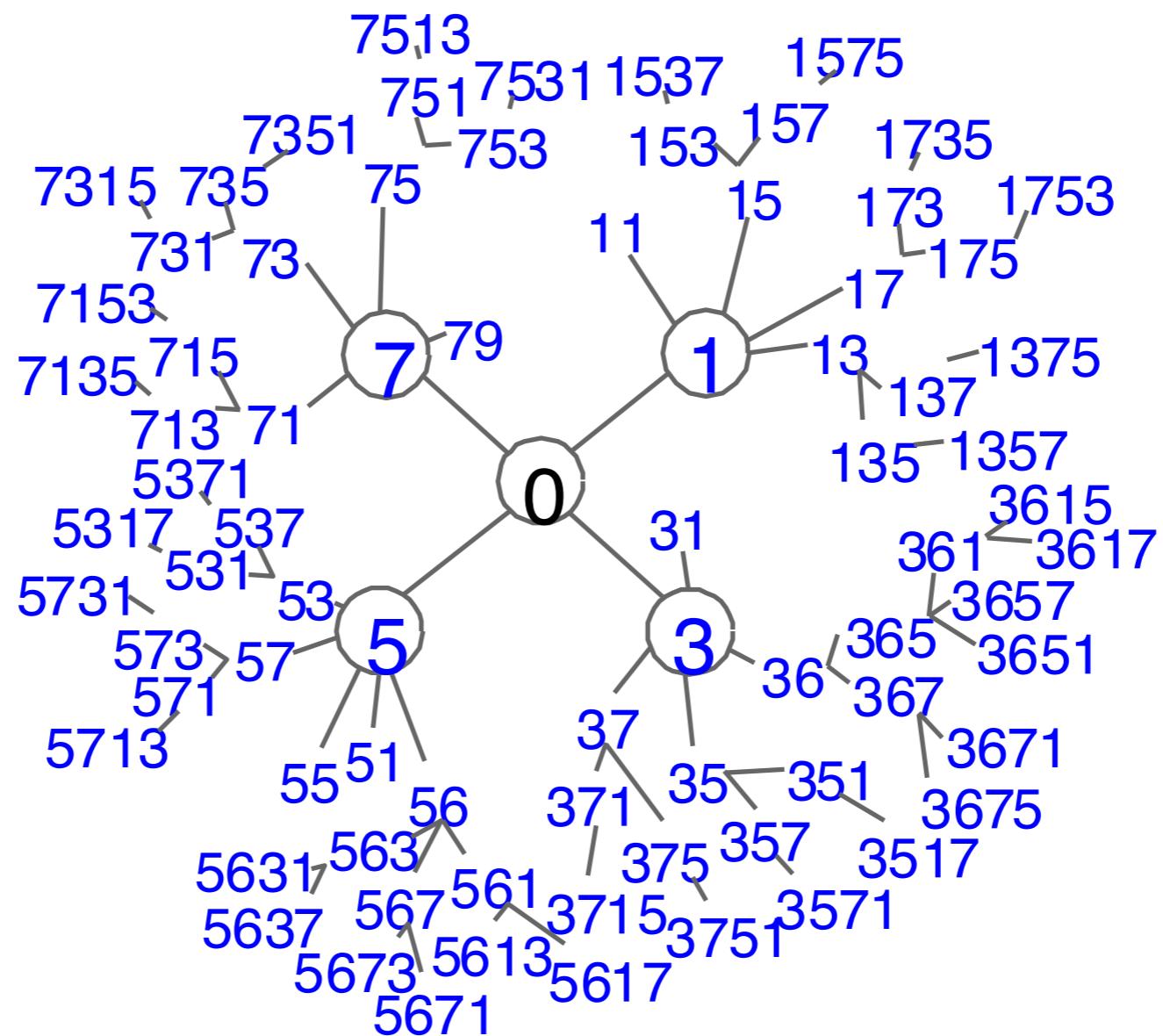
RNA Virus Quasispecies



RNA Virus Quasispecies



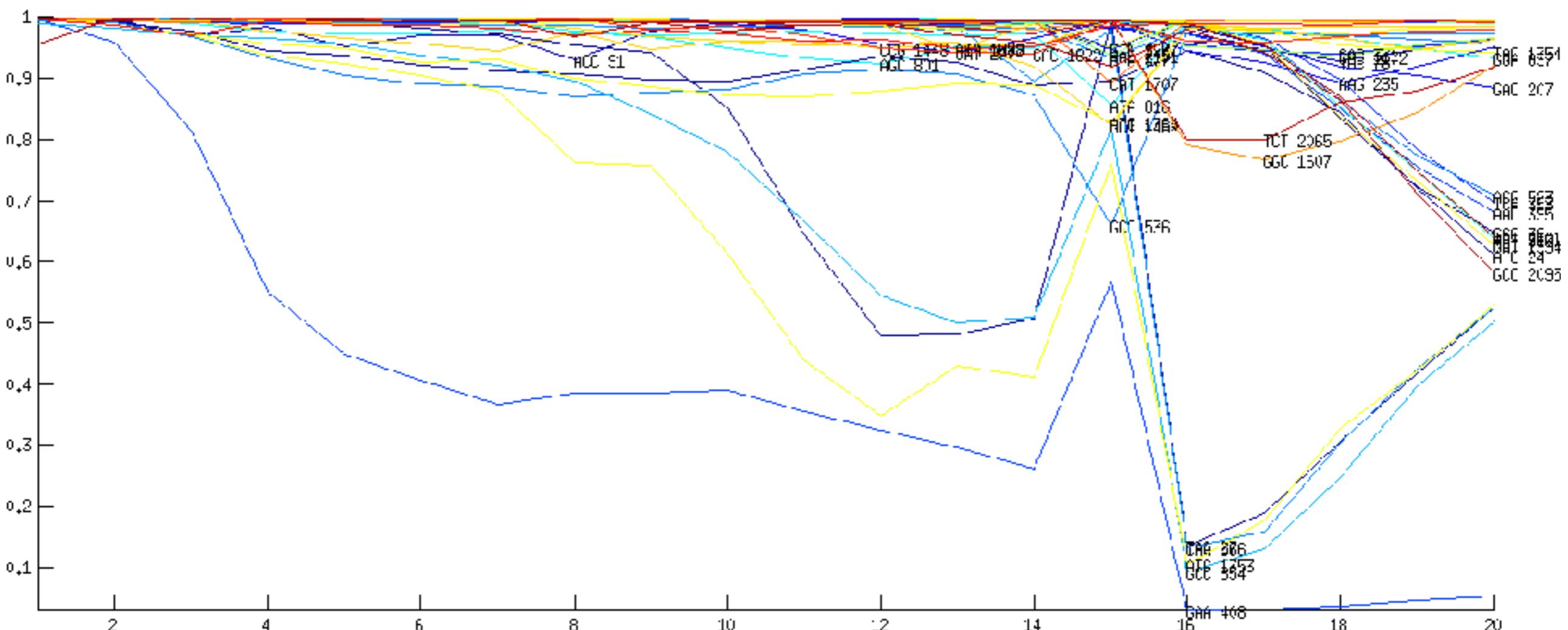
RNA Virus Quasispecies



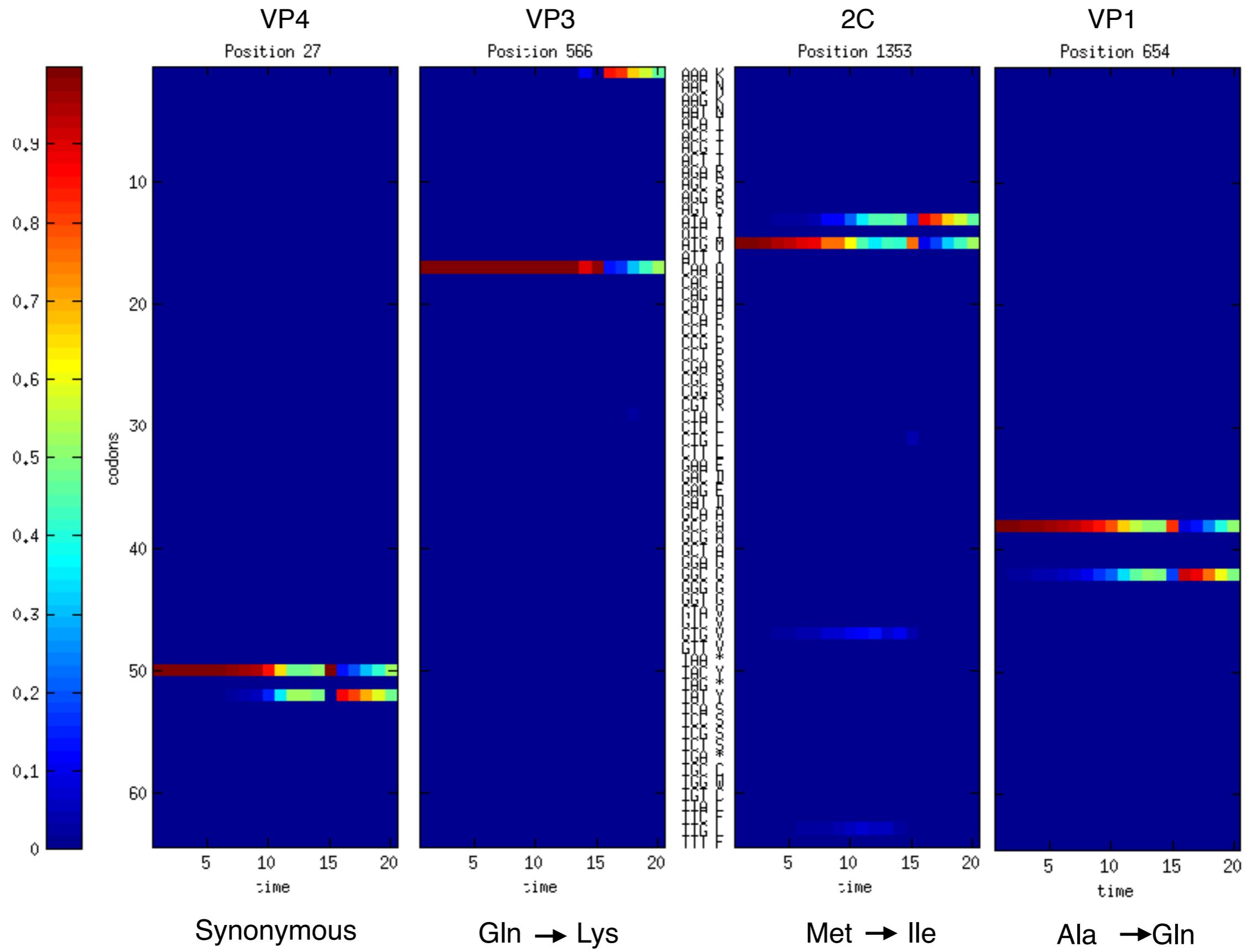
The Data

- Coxsackie virus
 - Low, Normal and High fidelity virus
 - Experimentally passaged 40 times in A549 cells
 - Whole-genome deep sequencing every 2nd passage

Frecuencia del codón dominante a través de los pasajes

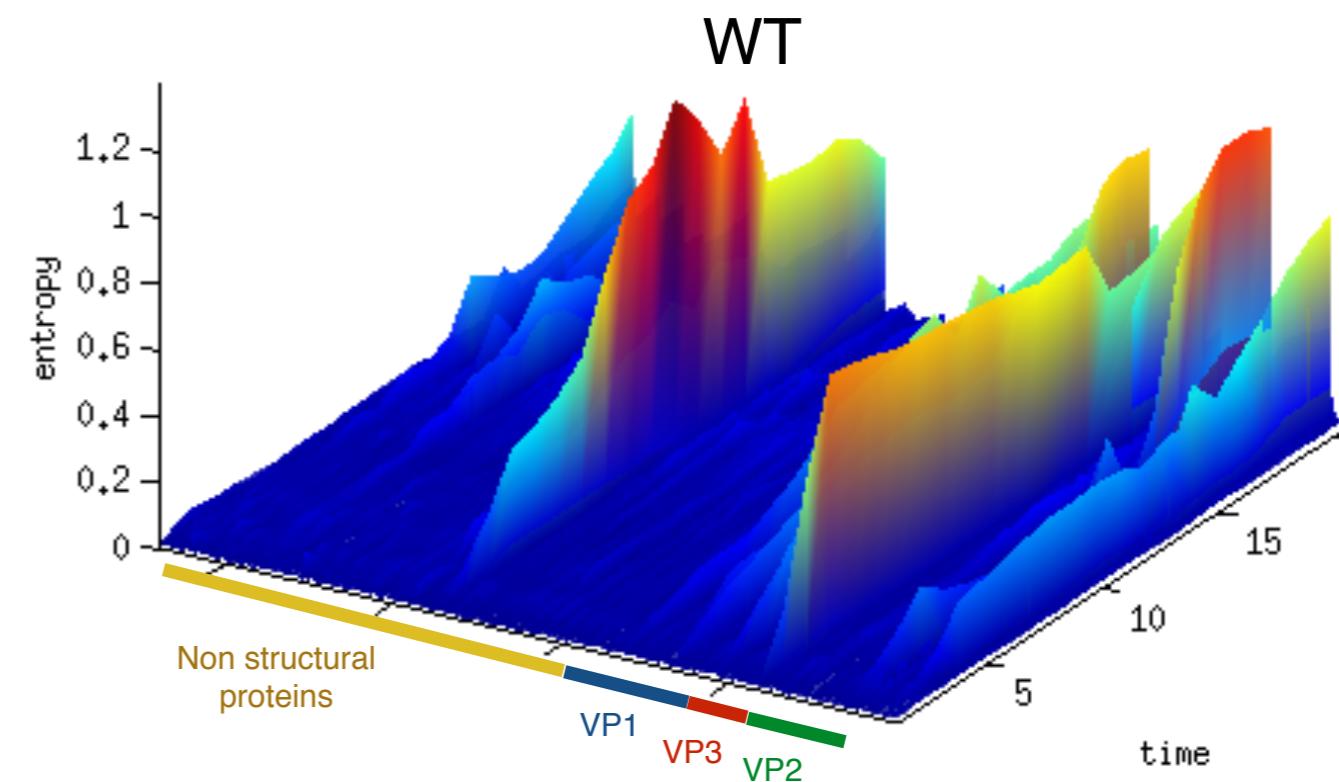


Cambios de frecuencia significativos

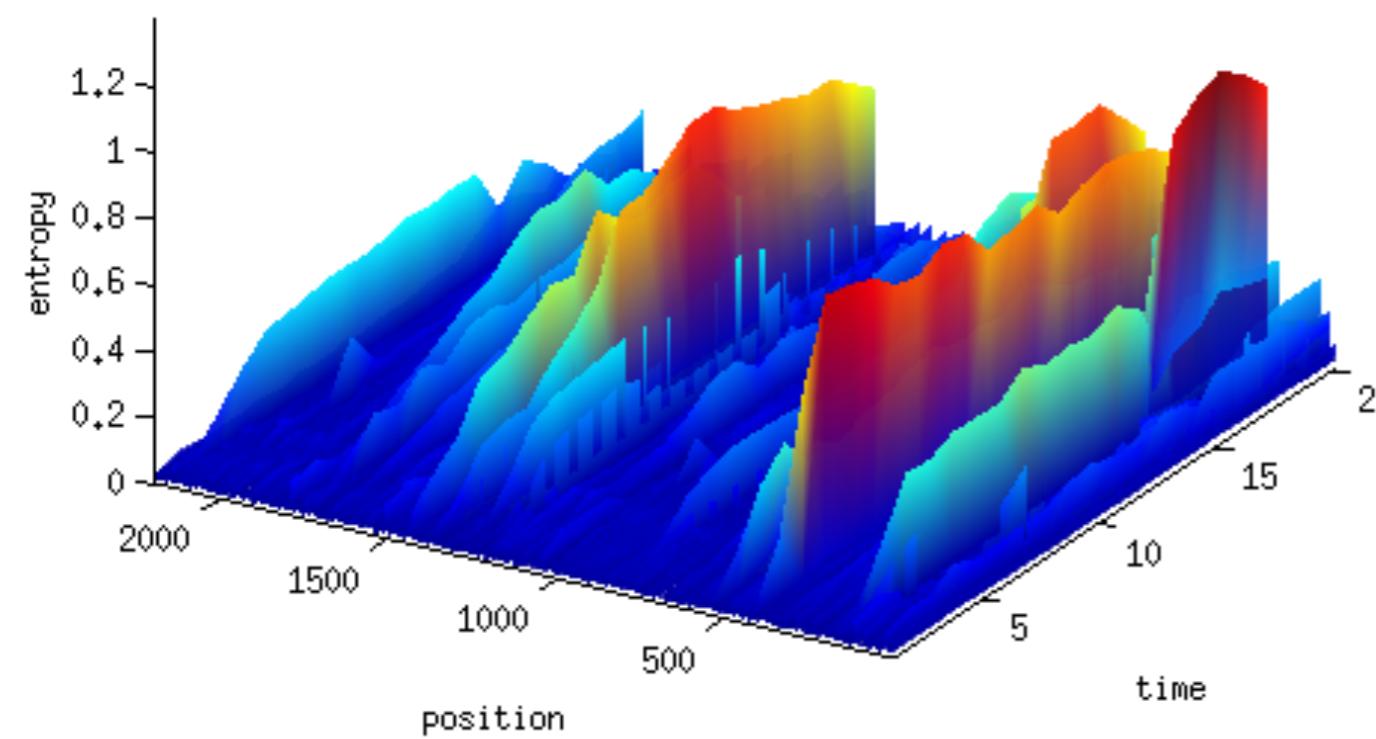


Entropía de las mutaciones de los codones a través de los pasajes

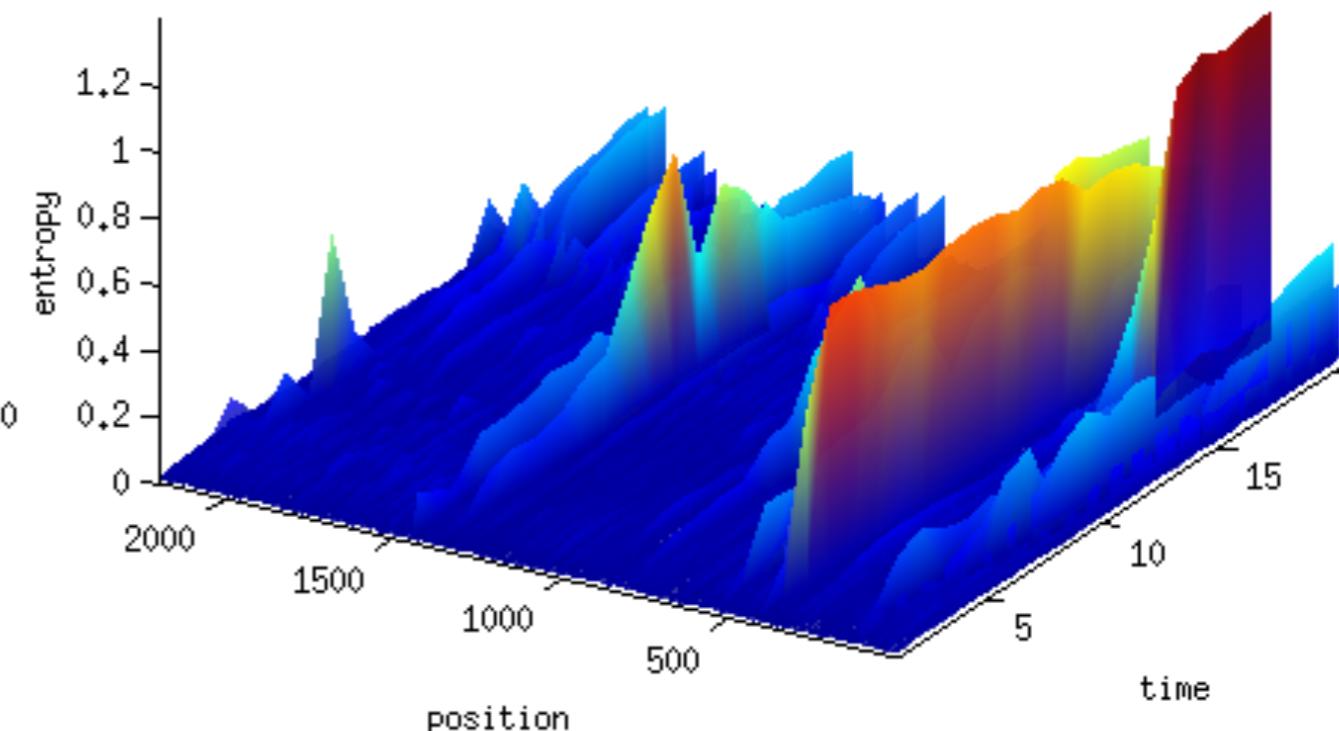
$$\text{Entropy} = - \sum_{i=1}^{64} p_i \log p_i$$



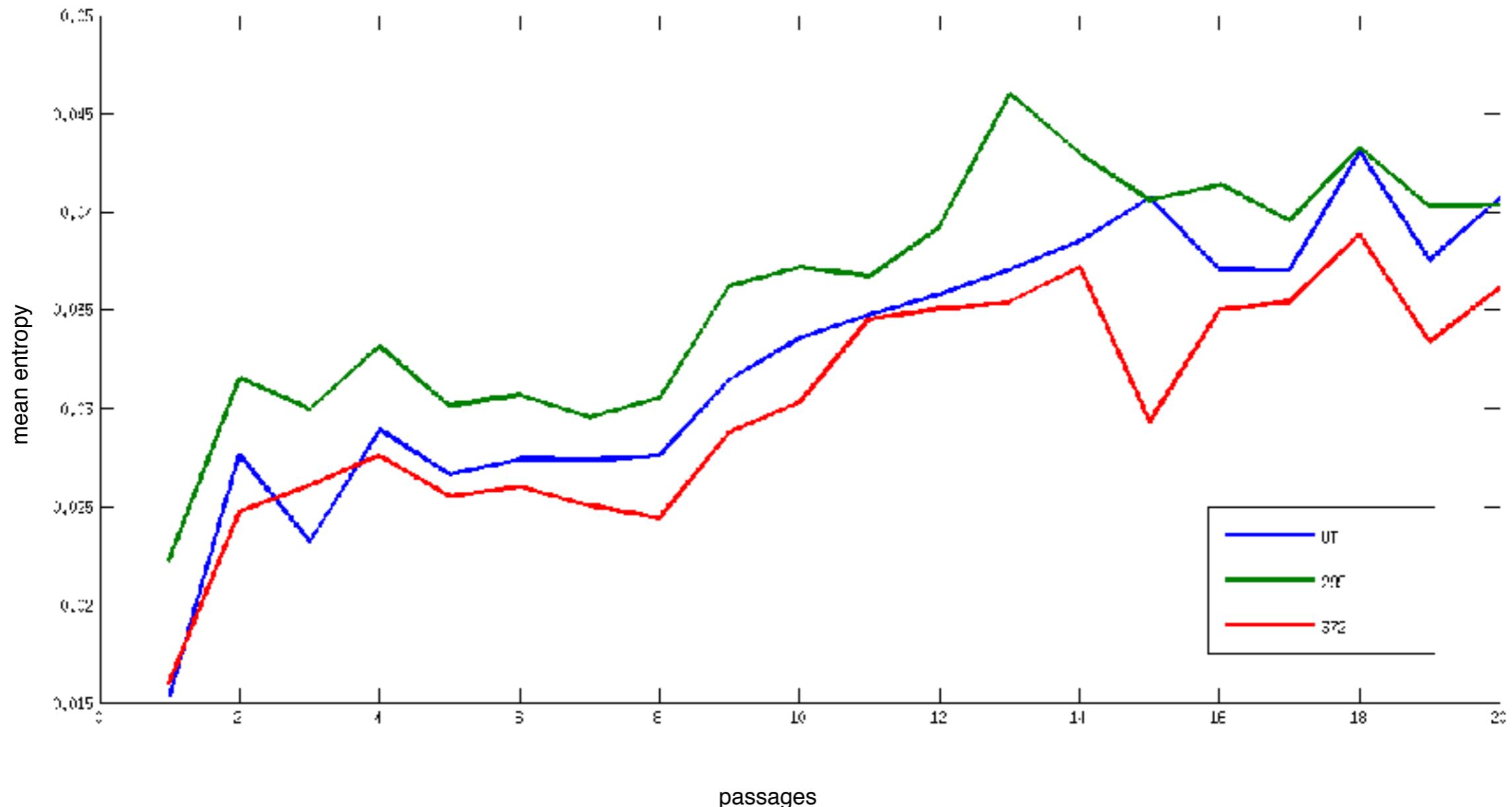
Low Fidelity



High Fidelity

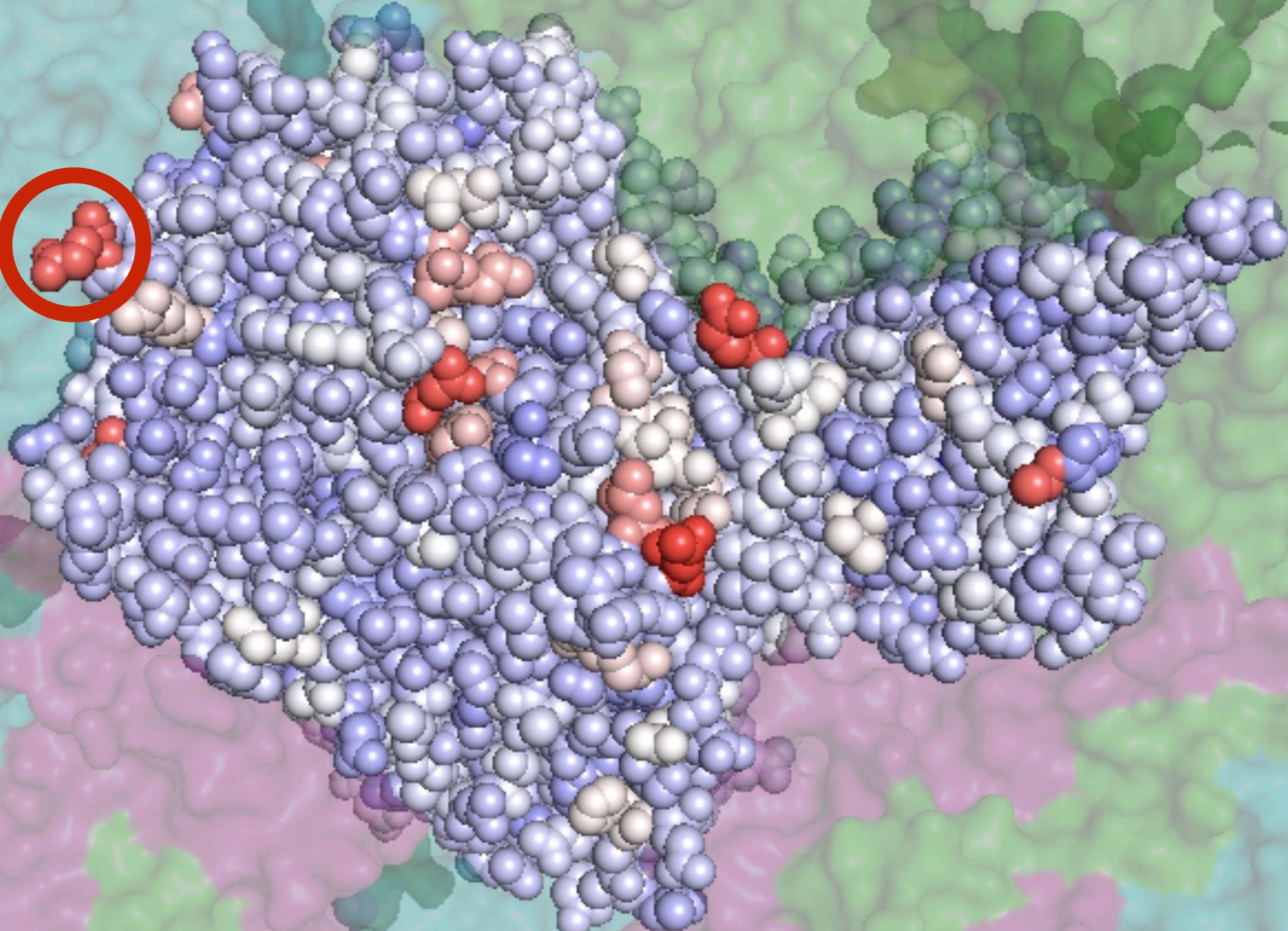
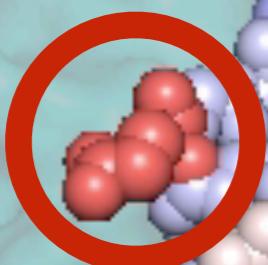


Evolución de la entropía media



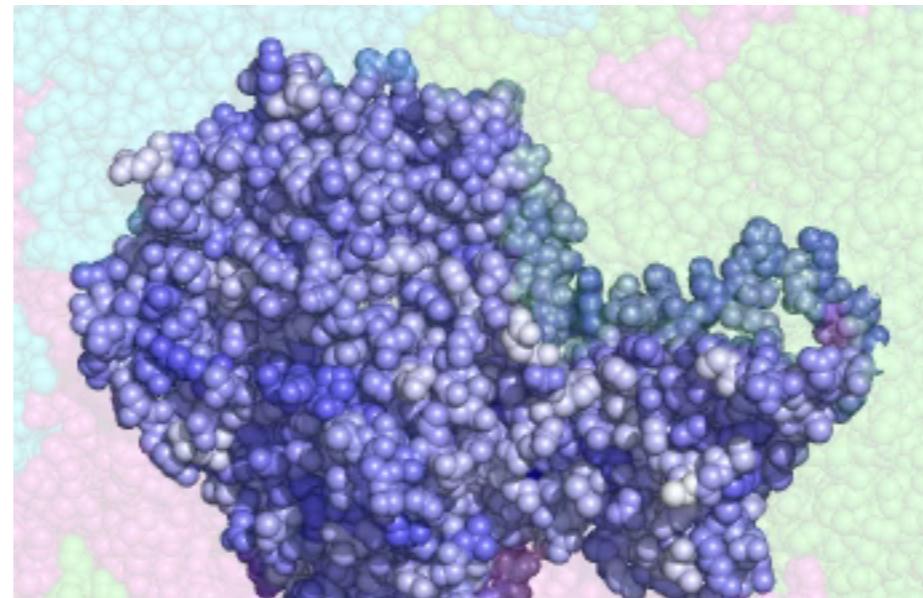
Entropía de las mutaciones de los codones mareadas en la estructura del protómero

Amino-acid
coded by
one codon

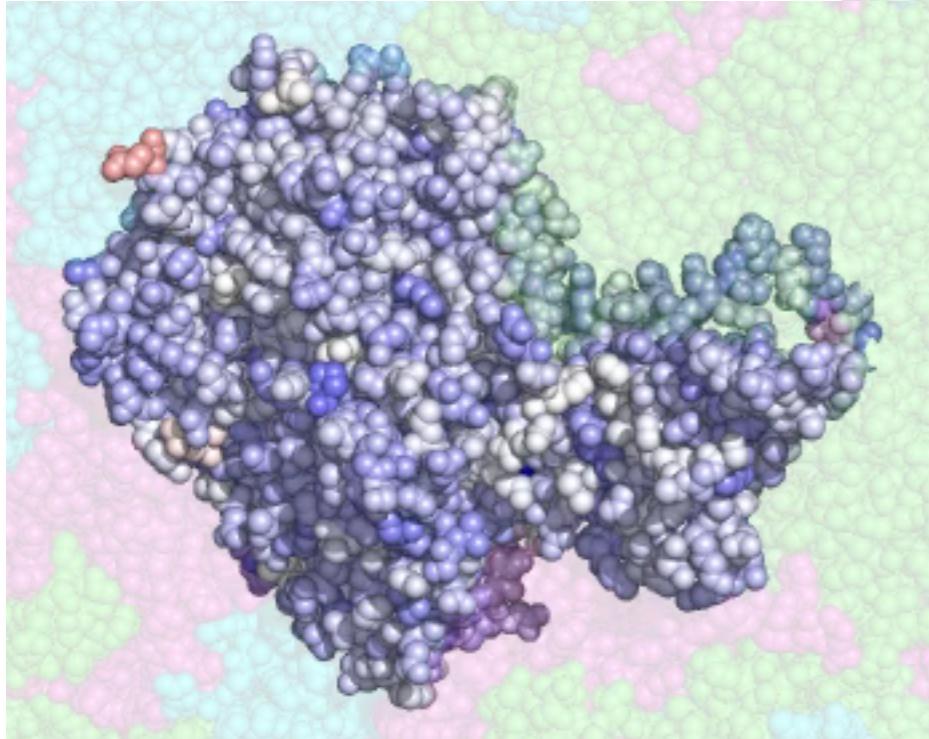


Entropía de las mutaciones de los codones mareadas en la estructura del protómero

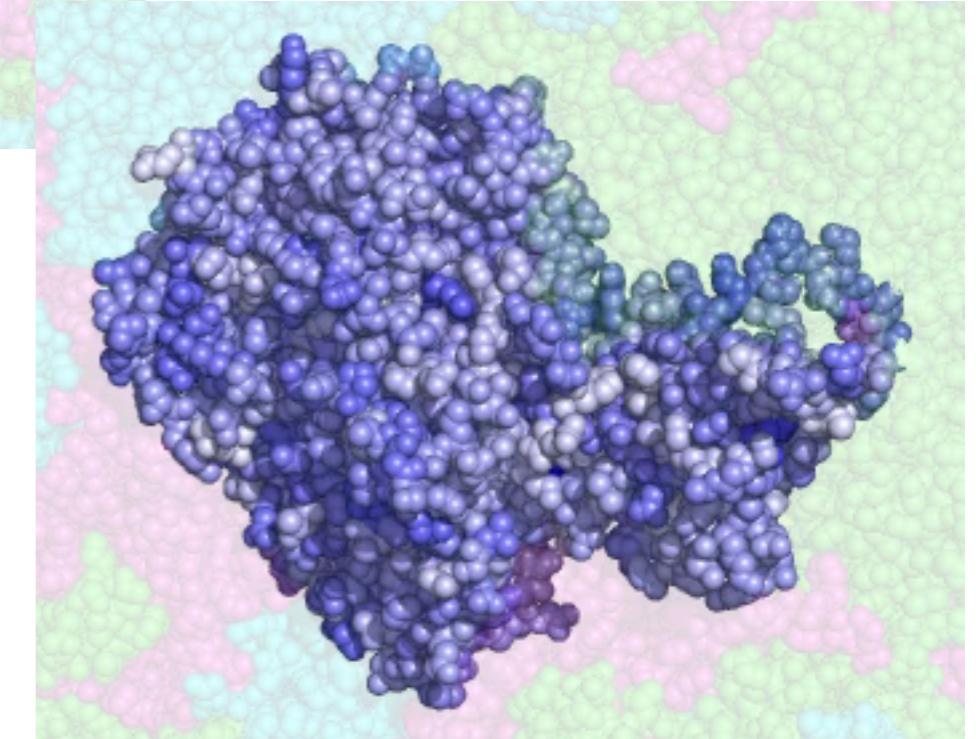
WT



Low Fidelity



High Fidelity



Entropy

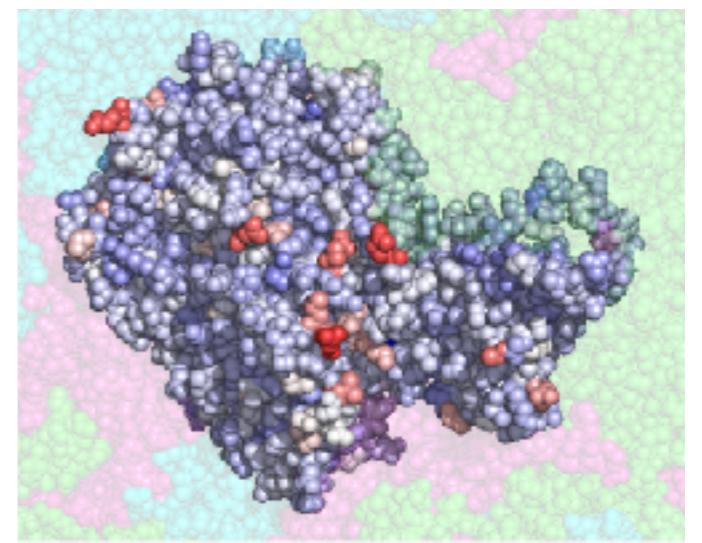
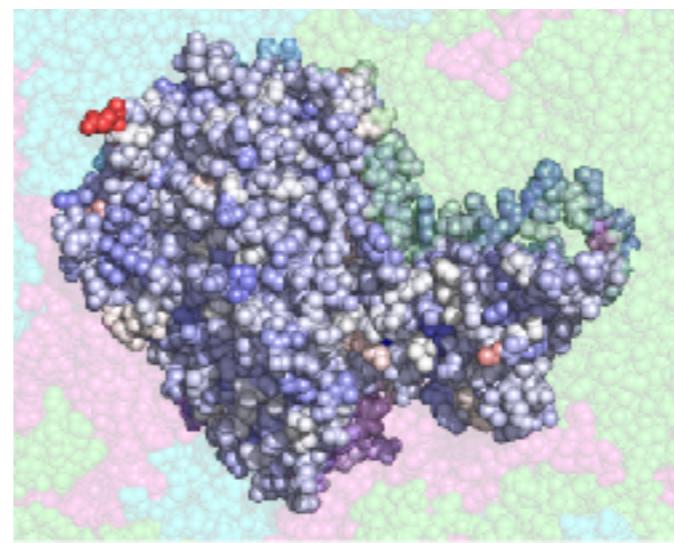
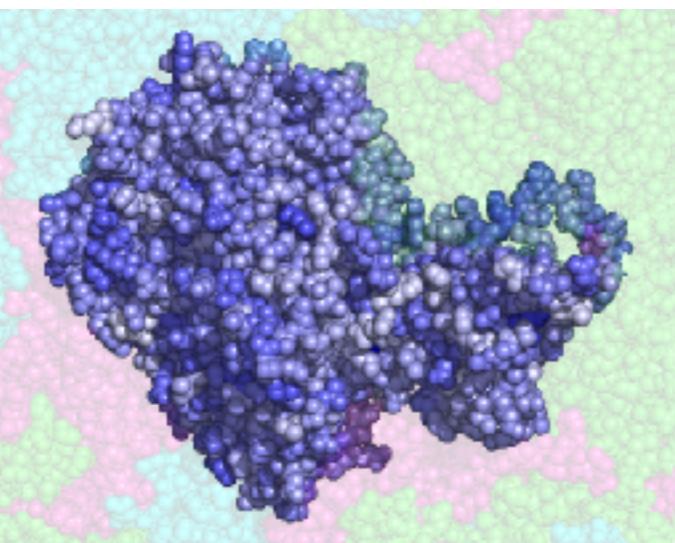


Entropía de las mutaciones de los codones mareadas en la estructura del protómero

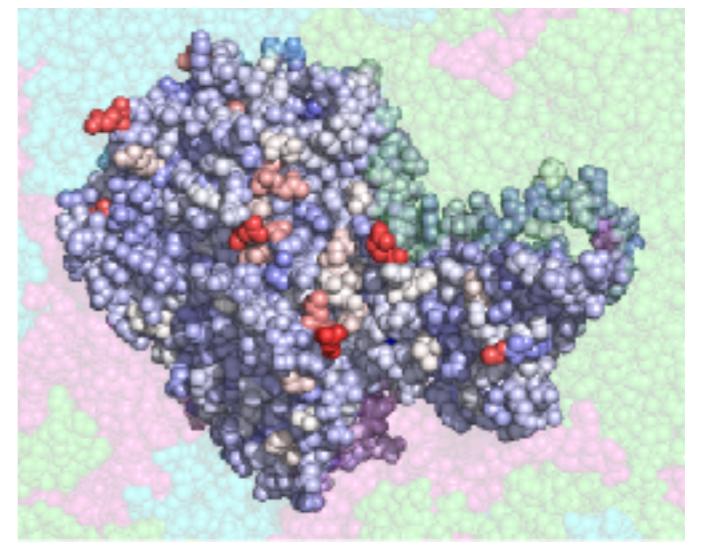
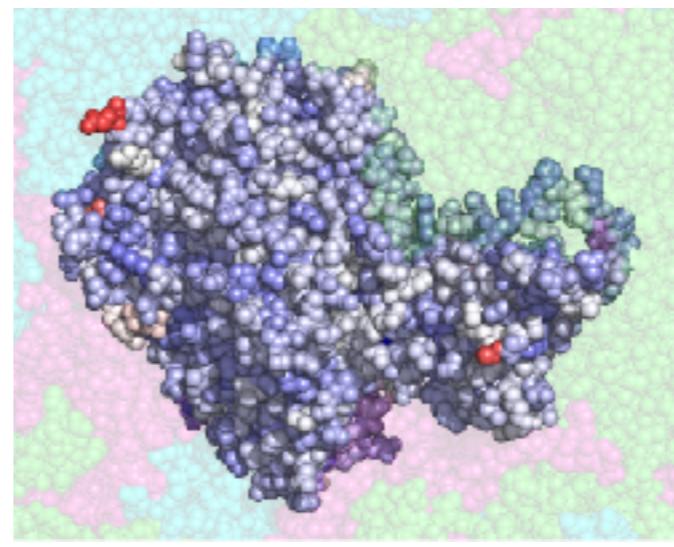
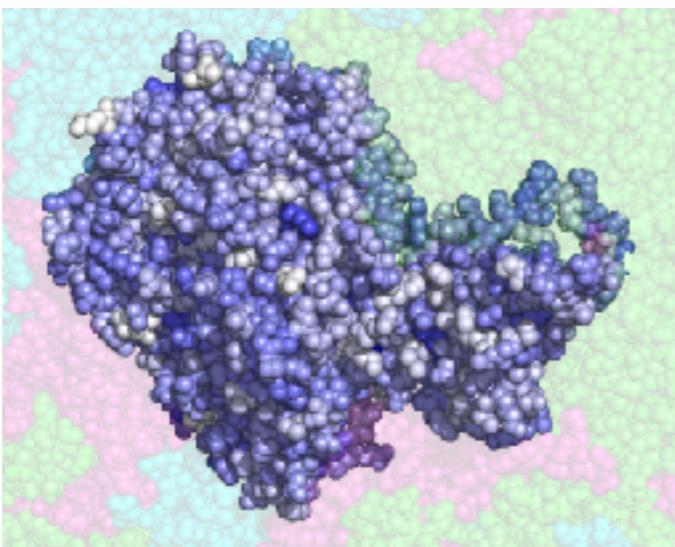


Entropy

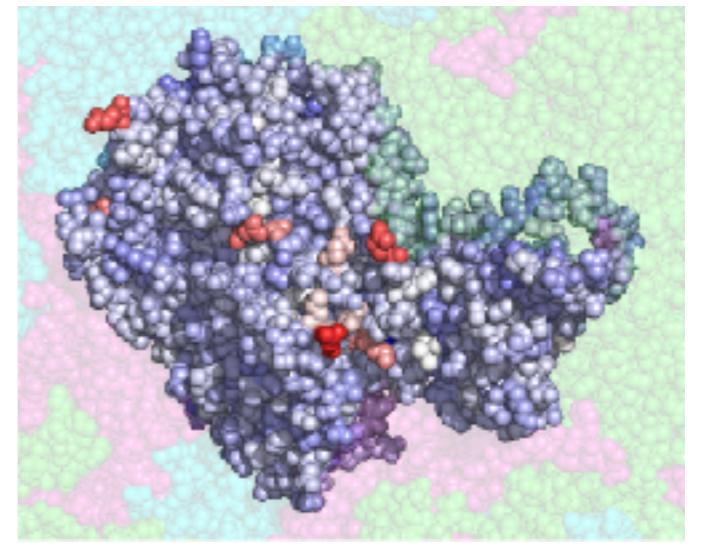
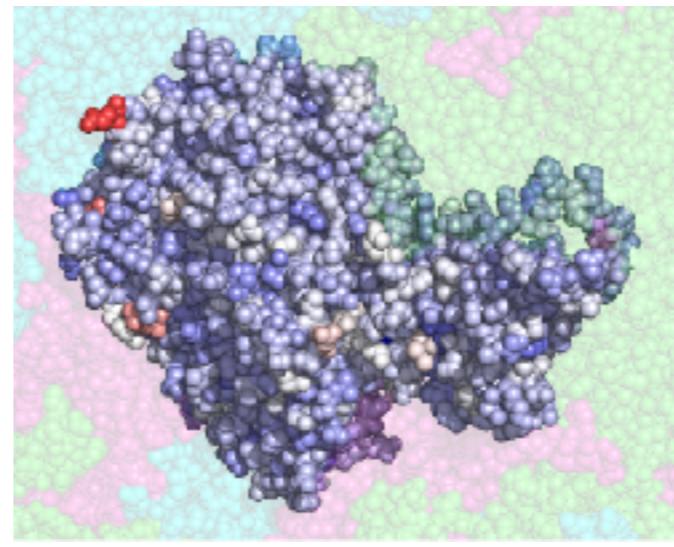
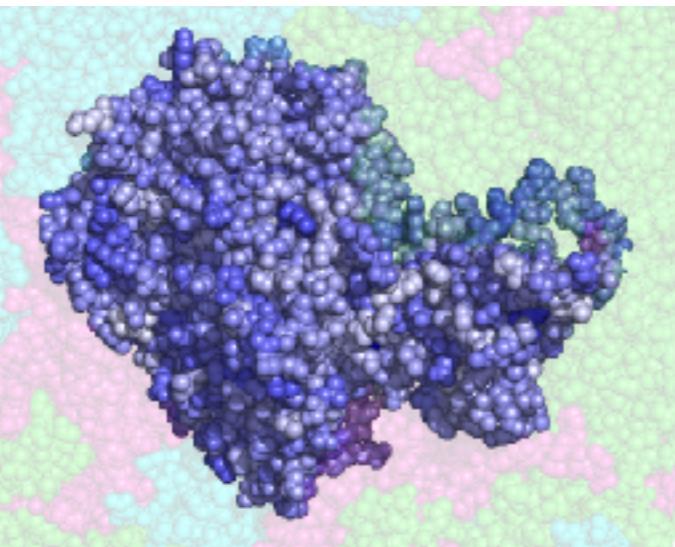
LF



WT



HF



Passages

