

# NON METAS A GAMBA! ANÁLISES DE DATOS XENÓMICOS CON R

[Adrián Casanova](#)<sup>1</sup>, Miguel Hermida<sup>1</sup>, Paulino Martínez<sup>1</sup>, Inmaculada Carrasco<sup>2</sup>, Francisca Robles<sup>3</sup>, Rafael Navajas<sup>3</sup>, Roberto de la Herrán<sup>3</sup>, Carmelo Ruiz<sup>3</sup>

<sup>1</sup> Departamento de Zooloxía, Xenética e Antropoloxía Física (ACUIGEN GI-1251), Facultade de Veterinaria, Campus Terra, Universidade de Santiago de Compostela, Lugo, España

<sup>2</sup> Asociación Organización de Productores Pesqueros de Motril OPP85, Motril, Granada, España

<sup>3</sup> Departamento de Genética, Facultad de Ciencias, Universidad de Granada, Granada, España

# Índice

**Introdución**

**Metodoloxía | Paquetes de R**

**Resultados**

**Primeiras conclusións**

# INTRODUCCIÓN

# Camarón soldado, quisquilla... (*Plesionika edwardsii*)



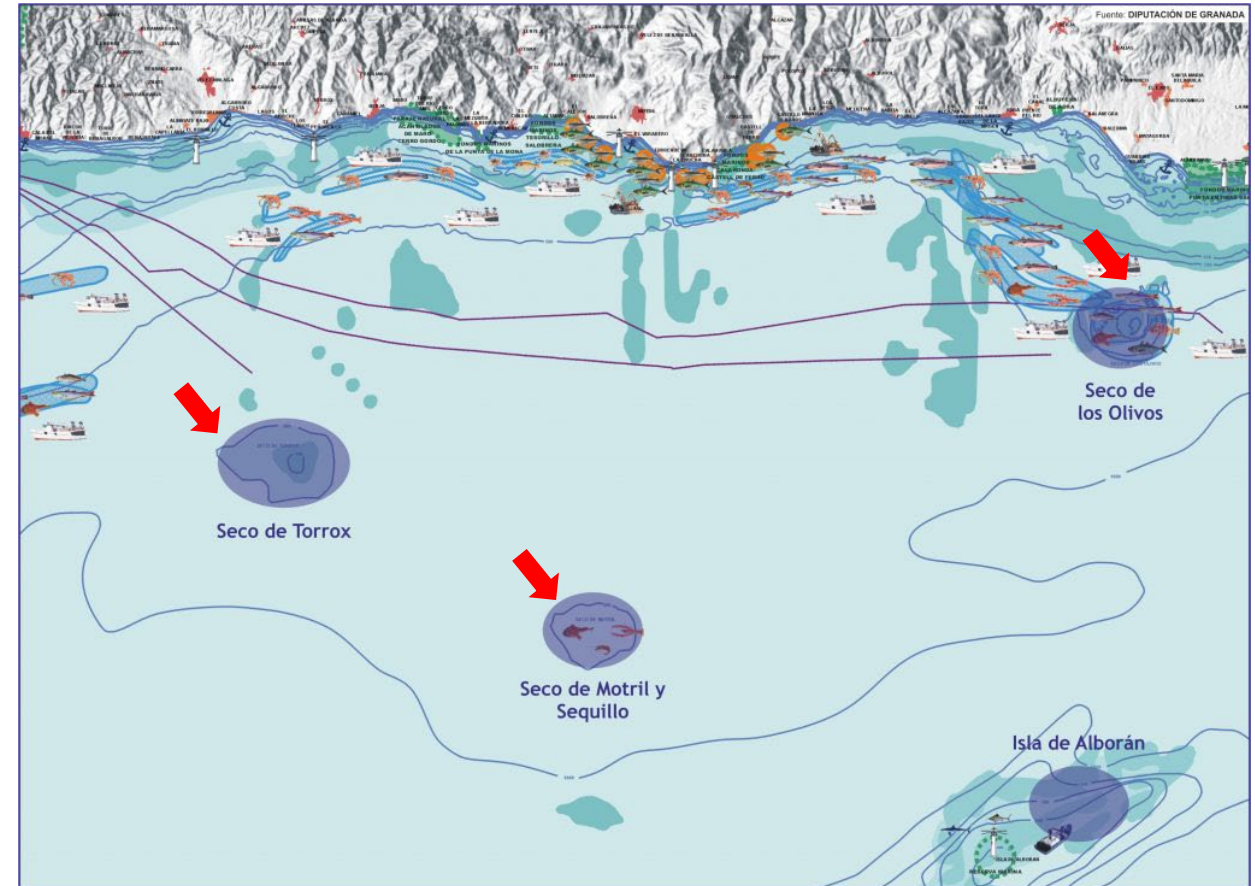
Imaxe de Luis Sánchez Tocino (Universidad de Granada). Fonte: <https://litoraldegranada.ugr.es/el-litoral/el-litoral-sumergido/fauna/artropodos/crustaceos/decapodos/carideos-2/plesionika-edwardsii/>

- ❑ Crustáceo decápodo de até uns 17 centímetros de lonxitude.
- ❑ Habita nun rango de profundidade entre 200-500 metros. Ligada ao fondo mariño pero con migracións verticais a traveso da columna de auga.
- ❑ Importante recurso pesqueiro no Mediterráneo occidental. O total de capturas en España ascendeu a ~400 toneladas de peso vivo (2022), cun valor medio nas lonxas duns 17 euros/kg.



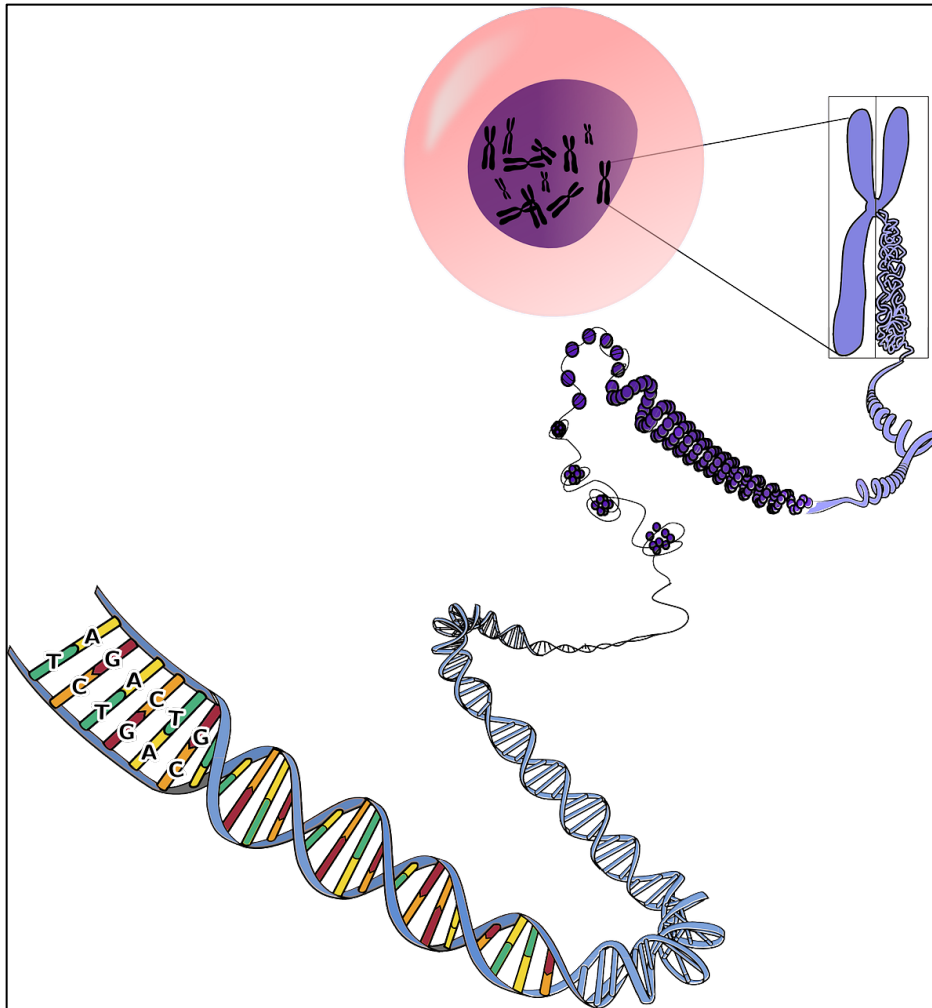
Imaxe de Luis Sánchez Tocino (Universidad de Granada). Fonte: <https://litoraldegranada.ugr.es/el-litoral/el-litoral-sumergido/fauna/artropodos/crustaceos/decapodos/carideos-2/plesionika-edwardsii/>

- ❑ **Secos:** montes submarinos que se agrupan en extensas mesetas, hábitats para unha multitude de especies de elevado valor ecolóxico e económico (caladoiros de pesca).
- ❑ Debido á localización en **diferentes secos**, aos **diferentes réximes hidrográficos** e as **correntes mariñas** existentes as poboacións do mar de Alborán poderían presentar un certo grao de illamento xenético.



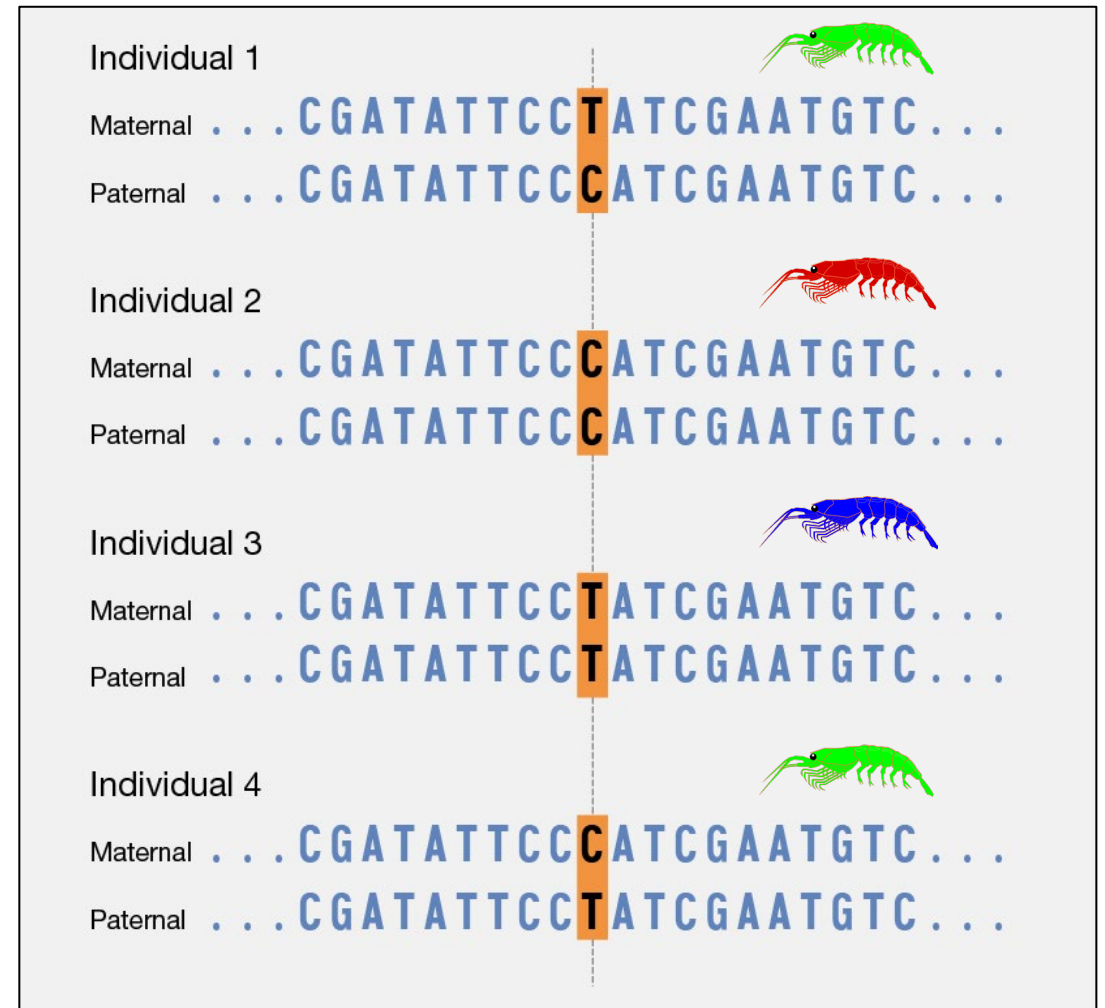
Fonte: Diputación de Granada

## ADN: A linguaxe da vida



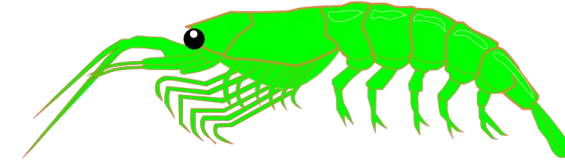
Fonte da imaxe (OpenClipart-Vectors; Pixabay):  
<https://pixabay.com/vectors/genetics-chromosomes-rna-dna-156404/>

## SNP: Polimorfismo de nucleótido único



Fonte da imaxe: <https://www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms>

**CA-056 | Xenoma completo (~17 xigabases)**



...AAAGCTGGCATCGATTGGATTGGGAAGCACCCACTCG**GCA**GGGATT**TGC**ATCTCAGTTTAAGGGGCTAGTTACTGATGA...



**AlfI** encima de restrición  
(tesoira molecular)



**Stacks**

SNP 154\_26 (CA-056)

T/C

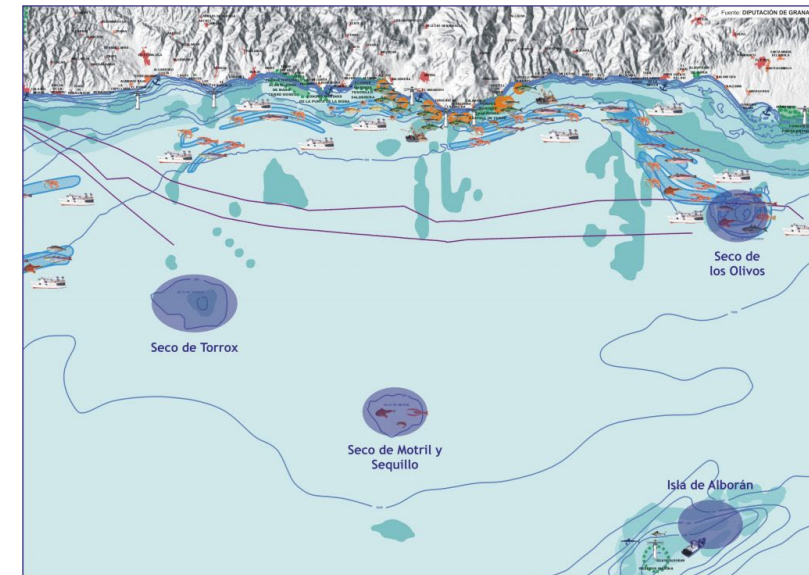
Profundidade de  
cobertura = 7x

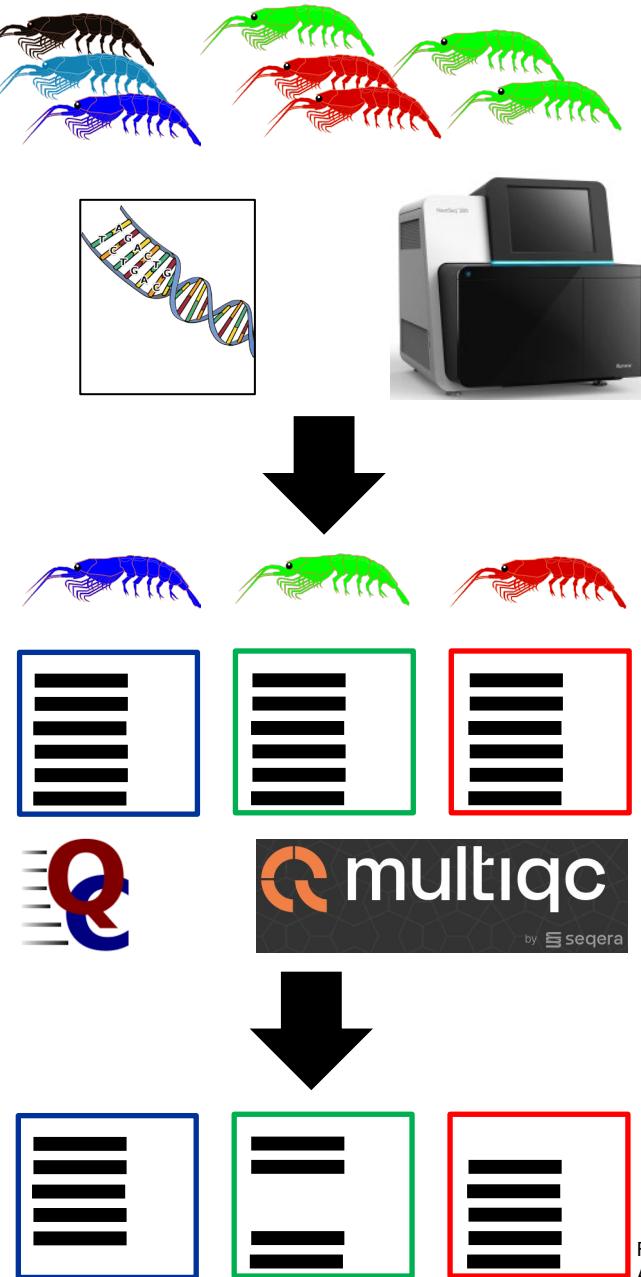
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ATCTCAGTTTAA	}	<b>T</b> 5x
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ATCTCAGTTTAA		
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ATCTCAGTTTAA		
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ATCTCAGTTTAA		
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ATCTCAGTTTAA		
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ACCTCAGTTTAA	}	<b>C</b> 2x
AGCACCCACTCG <b>GCA</b> GGGATT <b>TGC</b> ACCTCAGTTTAA		



# METODOLOXÍA | PAQUETES DE R

Localidade	Código	Océano/mar	N mostras
Illas Canarias	CA	Atlántico	28
Huelva	HU	Atlántico	24
Seco de Motril	SM	Mediterráneo	24
Seco de los Olivos	SO	Mediterráneo	28
Seco de Torrox	ST	Mediterráneo	24

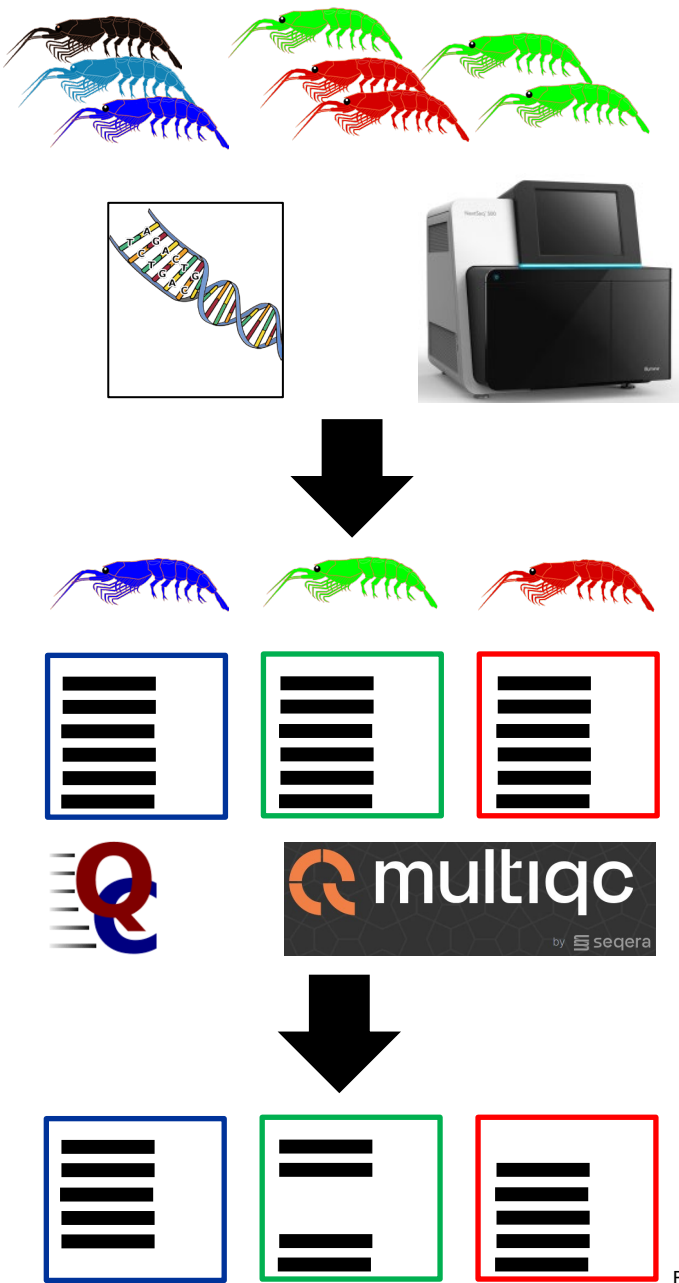




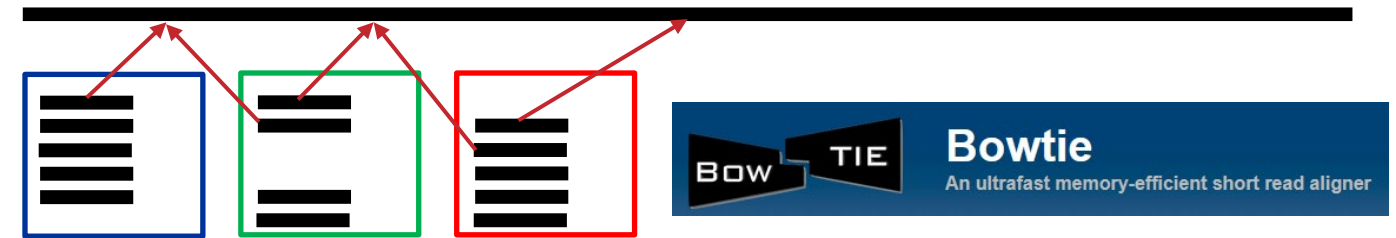
FastQC e multiQC:

Andrews S. (2010). FastQC a Quality Control Tool for High Throughput Sequence Data. [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Ewels P, Magnusson M, Lundin S, Källér M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*; **32**: 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>



## Xenoma de referencia



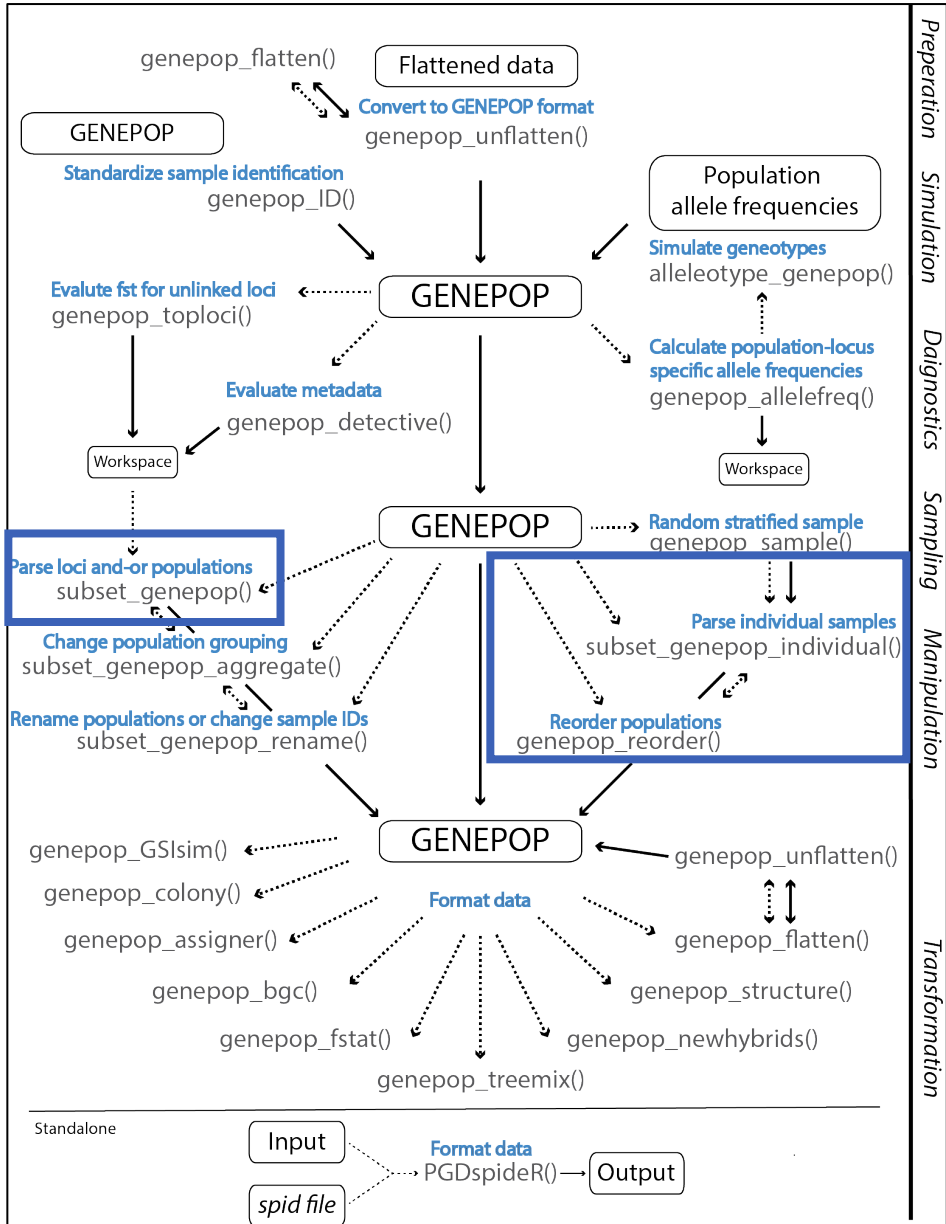
## Stacks

SNP1	AA	AC	CC
SNP2	AG	GG	AA
SNP3	CT	CT	TT
SNP4	TT	AT	TT

Panel bruto de SNPs  
**VCF: 4.76 xigabytes!**  
**GENEPOP: 1.19 xigabytes!**



# METODOLOXÍA | PAQUETES DE R



Fonte imaxe (GitHub): <https://github.com/rystanley/genepopedit>

# Crear un obxecto vectorial en R coa lista de SNPs a eliminar, co formato adecuado

```
blacklist<-read.table(file = "blacklist.txt", header = FALSE)
```

```
blacklist_T<-t(blacklist)
```

```
blacklist_V<-as.vector(blacklist_T)
```

# Eliminar os SNPs de baixa calidade usando o paquete **{genepopedit}**

```
subset_genepop(genepop= "1.Plesionika_N128_XXX_SNPs.gen", keep = FALSE,
subs = blacklist_V, path = "2.Plesionika_N128_XXX_SNPs.gen")
```

# Ler o novo arquivo genepop usando o paquete **{adegenet}**


```
read.genepop("2.Plesionika_N128_XXX_SNPs.gen",ncode = 3L,quiet = FALSE)
```

As referencias do software citado no Resumo pódense consultar usando este QR



## genepop: Population Genetic Data Analysis Using Genepop

Makes the Genepop software available in R. This software implements a mixture of traditional population genetic methods and some more focused developments: it computes exact tests for Hardy-Weinberg equilibrium, for population differentiation and for genotypic disequilibrium among pairs of loci; it computes estimates of F-statistics, null allele frequencies, allele size-based statistics for microsatellites, etc.; and it performs analyses of isolation by distance from pairwise comparisons of individuals or population samples.

Version: 1.2.2  
Imports: [Rcpp](#) ( $\geq 0.12.10$ ), [stringr](#)  
LinkingTo: [Rcpp](#), [RcppProgress](#)  
Suggests: [testthat](#), [knitr](#), [shiny](#)  
Published: 2023-01-19  
DOI: [10.32614/CRAN.package.genepop](https://doi.org/10.32614/CRAN.package.genepop)  
Author: François Rousset  [aut, cre, cph], Jimmy Lopez [ctb], Alexandre Genin [ctb], Khalid Belkhir [ctb]  
Maintainer: François Rousset <francois.rousset at umontpellier.fr>  
License: [CeCILL-2](#)  
URL: <https://www.r-project.org>, <https://kimura.univ-montp2.fr/~rousset/Genepop.htm>  
NeedsCompilation: yes  
Citation: [genepop citation info](#)  
Materials: [NEWS](#)  
CRAN checks: [genepop results](#)

Fonte (CRAN): <https://cran.r-project.org/web/packages/genepop/index.html>

## # Obtención da $F_{ST}$ global e outros estatísticos F

```
genepop::Fst("N128b_6x_360_DEF.recode.txt", sizes = FALSE, pairs = FALSE, outputFile = "N128b_6x_360.FST", dataType = "Diploid")
```

## # Test Hardy-Weinberg

```
genepop::test_HW("N128b_6x_360_DEF.recode.txt", which = "Proba", outputFile = "N128b_6x_360.P", enumeration = TRUE, verbose = interactive())
```

## # Estimacións da diversidade xenética

```
genepop::genedivFis("N128b_6x_360_DEF.recode.txt", sizes = FALSE, outputFile = "N128b_6x_360.DIV", dataType = "Diploid", verbose = interactive())
```




## MOLECULAR ECOLOGY RESOURCES

Resource Article

### StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations

Luke W. Pembleton, Noel O. I. Cogan, John W. Forster

First published: 06 June 2013 | <https://doi.org/10.1111/1755-0998.12129> | Citations: 479

 PDF  TOOLS  SHARE

### Statistical Analysis of Mixed Ploidy Populations



Documentation for package 'StAMPP' version 1.6.3

- [DESCRIPTION file](#).
- [Package NEWS](#).

#### Help Pages

<a href="#">potato</a>	Example genotype input format
<a href="#">potato.mini</a>	Smaller example genotype input format
<a href="#">stampp2genlight</a>	Convert StAMPP genotype data to genlight object
<a href="#">stamppAmove</a>	Analysis of Molecular Variance
<a href="#">stamppConvert</a>	Import and Convert
<a href="#">stamppFst</a>	Fst Computation
<a href="#">stamppGmatrix</a>	Genomic Relationship Calculation
<a href="#">stamppNeisD</a>	Genetic Distance Calculation
<a href="#">stamppPhylip</a>	Export to Phylip Format

```
# Ler o arquivo cos xenotipados en formato genepop usando o paquete {adegenet}
genind<-read.genepop("N128b_6x_360_DEF.recode.gen",ncode = 3L,quiet =
FALSE)

# Conversión dun obxecto genind nun obxecto genlight {dartR}
genlight<-gi2gl(gi = genind,parallel = TRUE,verbose = NULL)

# Estimación da FST por pares de localidades {StAMPP}
output<-stampFst(genlight,nboots = 10000,percent = 95,nclusters = 20)

# Exportación dos resultados {base}
write.matrix(output$Fsts,file = "matrix_Fst_B10000_n128_Plesionika.txt")
write.matrix(output$Pvalues,file = "matrix_Pvalues_B10000_n128_Plesionika.txt")
out_Boots<-output$Bootstraps
out_Boots_t<-t(out_Boots)
write.table(out_Boots_t,file = "boots_10000_n128_Plesionika.txt")
```



## Exploratory Analysis of Genetic and Genomic Data



Documentation for package 'adegenet' version 2.1.10

• [DESCRIPTION file](#).

### Help Pages

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [misc](#)

-- A --

<a href="#">a_score</a>	Compute and optimize a-score for Discriminant Analysis of Principal Components (DAPC)
<a href="#">addStrata</a>	Access and manipulate the population strata for genind or genlight objects.
<a href="#">addStrata-method</a>	Access and manipulate the population strata for genind or genlight objects.
<a href="#">addStrata&lt;-</a>	Access and manipulate the population strata for genind or genlight objects.
<a href="#">addStrata&lt;-method</a>	Access and manipulate the population strata for genind or genlight objects.
<a href="#">adegenet</a>	The adegenet package
<a href="#">adegenet.package</a>	The adegenet package
<a href="#">adegenetIssues</a>	Functions to access online resources for adegenet
<a href="#">adegenetServer</a>	Web servers for adegenet
<a href="#">adegenetTutorial</a>	Functions to access online resources for adegenet

```
# Obtención do número máis probable de clústeres empregando o BIC (Bayesian Information Criterion) {adegenet}
```

```
Clusters_Plesionika<- find.clusters(genind_Plesionika_N128, max.n.clust=20, n.iter = 1000000, n.start = 100)
```

```
# DAPC {adegenet}
```

```
Plesionika_N128_DAPC<-dapc(genind_Plesionika_N128,genind_Plesionika_N128$pop,n.pca = 30,var.contrib = TRUE, var.loadings=TRUE)
```

```
# Gráfico de dispersion cos resultados do DAPC
```

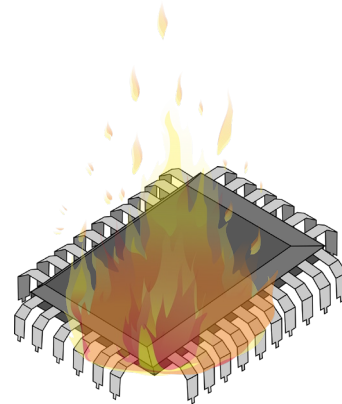
```
myCol <- c("dodgerblue","blue","darkolivegreen","darkolivegreen4","darkolivegreen2")
```

```
## Creación dun arquivo SVG (Scalable Vector Graphics) {grDevices}
```

```
svg("20240514_Plesionika_N128_30p_DAPC.svg")
```

```
scatter(genind_Plesionika_N128_DAPC,legend = TRUE, scree.da=TRUE, scree.pca = TRUE,cstar = TRUE, col=myCol)
```

```
dev.off()
```



*Eu estiven alí Gandalf, fai 3.000 anos, cando Isildur executou aquela análise..., que aínda dura.*


## Structure Software

Pritchard Lab, Stanford University

Home	Research	Publications	Software	Data	Contact	Lab Members	Join Us
------	----------	--------------	----------	------	---------	-------------	---------

The program *structure* is a free software package for using multi-locus genotype data to investigate population structure. Its uses include inferring the presence of distinct populations, assigning individuals to populations, studying hybrid zones, identifying migrants and admixed individuals, and estimating population allele frequencies in situations where many individuals are migrants or admixed. It can be applied to most of the commonly-used genetic markers, including SNPs, microsatellites, RFLPs and AFLPs.

In 2016 John Novembre wrote a short [historical perspective](#) of Structure.



Download [Structure 2.3.4](#).

**fastSTRUCTURE for large SNP datasets** is out now! Links to the [preprint](#) and [software](#) (beta release) by Anil, Matthew and Jonathan.

**What to cite:** The basic algorithm was described by Pritchard, Stephens & Donnelly (2000). Extensions to the method were published by Falush, Stephens and Pritchard (2003), and (2007) and Hubisz, Falush, Stephens and Pritchard (2009).

Fonte: <https://web.stanford.edu/group/pritchardlab/structure.html>

## ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers

Francois Besnier<sup>1</sup>, Kevin A. Glover<sup>1,2</sup>

<sup>1</sup> Department of Population Genetics, Institute of Marine Research, Bergen, Norway, <sup>2</sup> Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Bergen, Bergen, Norway

### Structure Software

Pritchard Lab, Stanford University

[Home](#) [Research](#) [Publications](#) [Software](#) [Data](#) [Contact](#) [Lab Members](#) [Join Us](#)

The program *structure* is a free software package for using multi-locus genotype data to investigate population structure. Its uses include inferring the presence of distinct populations, assigning individuals to populations, studying hybrid zones, identifying migrants and admixed individuals, and estimating population allele frequencies in situations where many individuals are migrants or admixed. It can be applied to most of the commonly-used genetic markers, including SNPs, microsatellites, RFLPs and AFLPs.

In 2016 John Novembre wrote a short [historical perspective](#) of Structure.



Download [Structure 2.3.4](#).

fastSTRUCTURE for large SNP datasets is out now! Links to the [preprint](#) and [software](#) (beta release) by Anil, Matthew and Jonathan.

What to cite: The basic algorithm was described by Pritchard, Stephens & Donnelly (2000). Extensions to the method were published by Falush, Stephens and Pritchard (2003), and (2007) and Hubisz, Falush, Stephens and Pritchard (2009).

Fonte: <https://web.stanford.edu/group/pritchardlab/structure.html>



```
17416_Plesionika_N128.sh
1 #!/bin/bash
2 # this file is 17416_Plesionika_N128.sh
3
4 #SBATCH -t 28:00:00
5 #SBATCH -c 30
6 #SBATCH --mem-per-cpu 1G
7 #SBATCH --mail-type=all
8 #SBATCH --mail-user=adrian.casanova@usc.es
9
10 Rscript 17416_Plesionika_N128.R
11
```



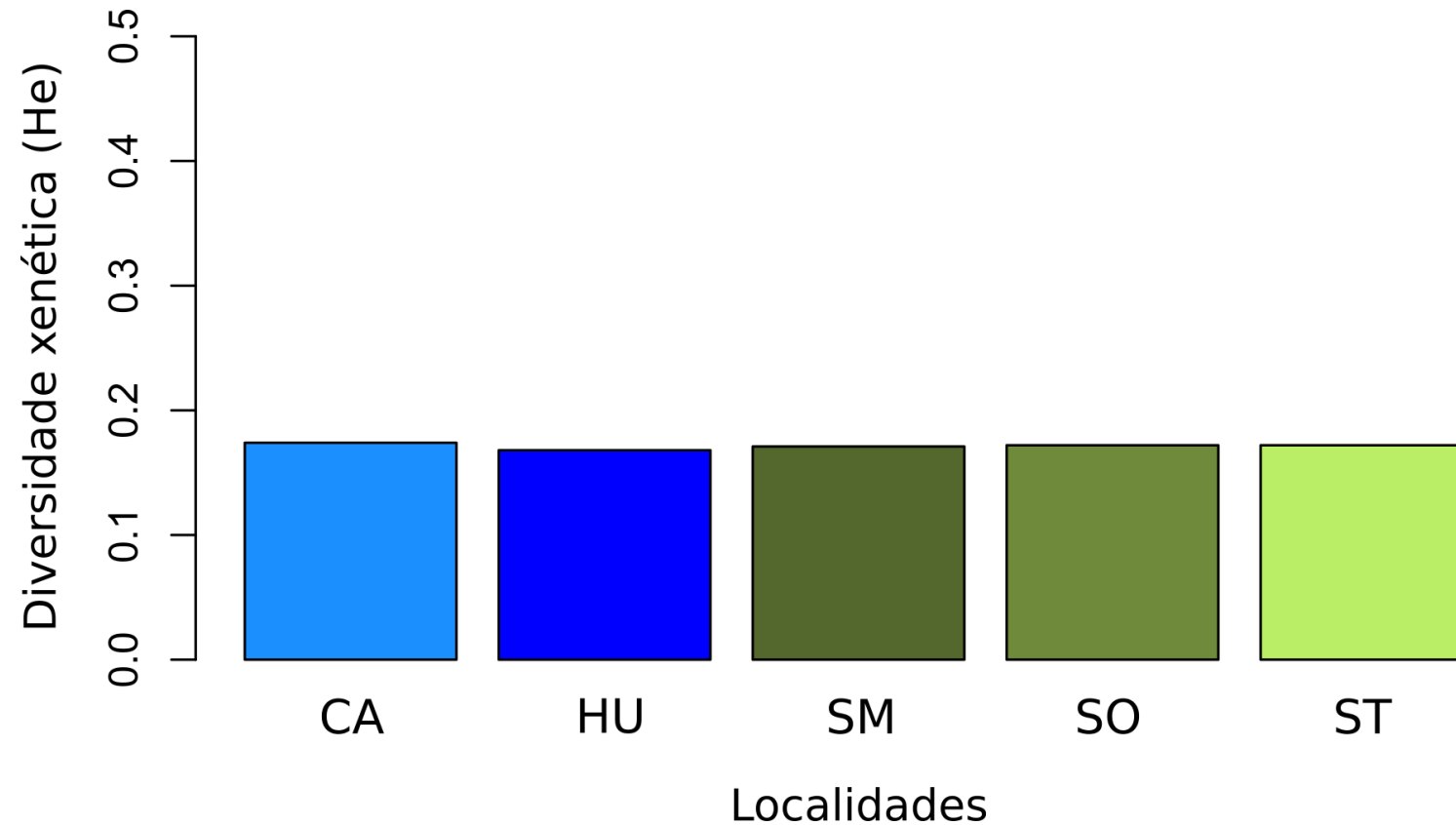
```
# Análise de ParallelStructure no CESGA (Centro de Supercomputación de Galicia)
```

```
library("ParallelStructure")
```

```
parallel_structure(structure_path="/home/usc/ge/acc/R/x86_64-pc-linux-gnu-library/", joblist="17416_Plesionika_N128_job", n_cpu=30,
infile="./N128b_6x_360_DEF.recode.stru", numinds=128, numloci=17416, printqhat=1, plot_output=0, onerowperind=0, locprior=0, noadmix =
0, popflag = 0, usepopinfo = 0, freqscorr = 1, markernames=1)
```

# RESULTADOS

*A diversidade xenética é a materia prima da evolución*

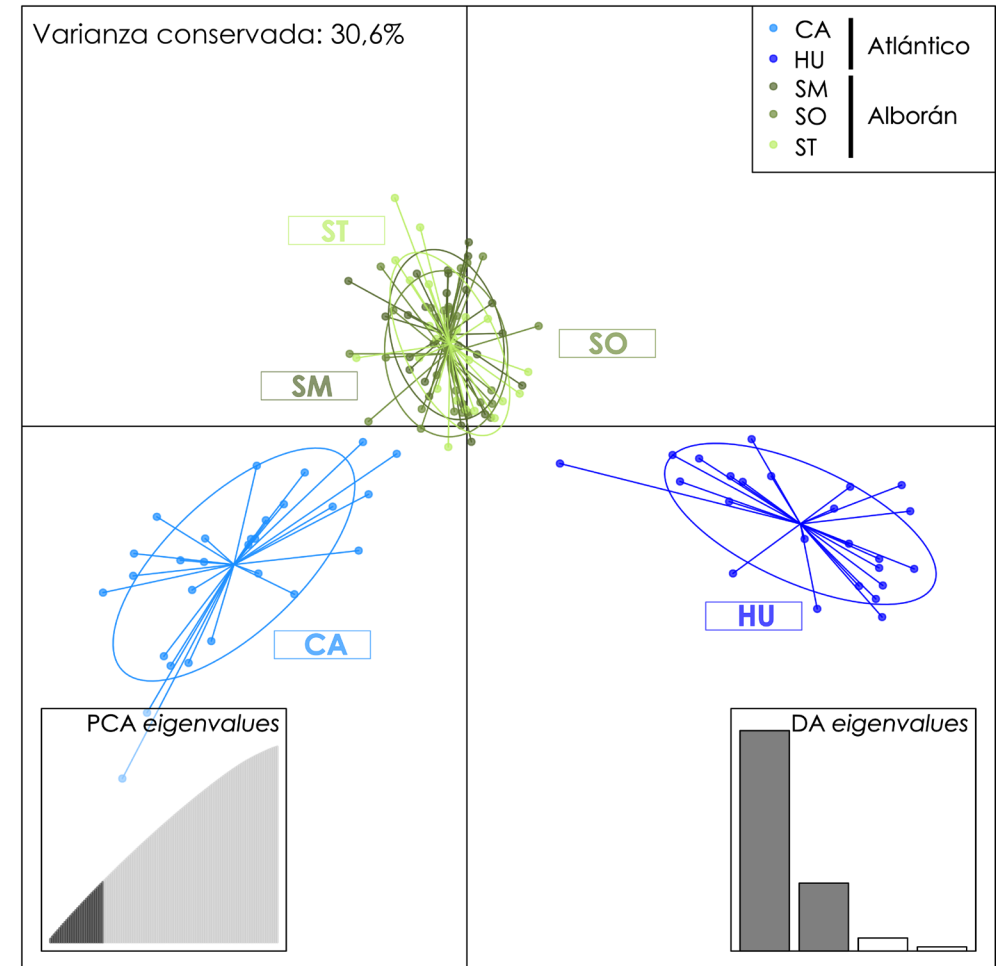
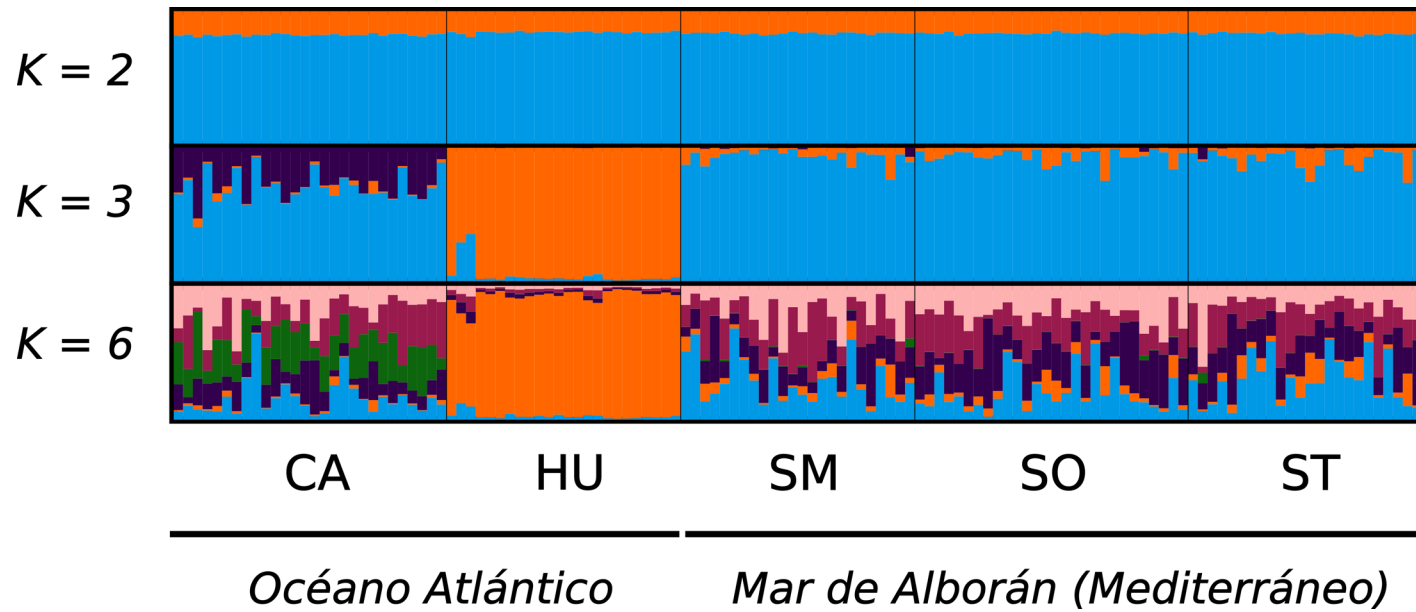


## $F_{ST}$ (Índice de fixación): Diferenciación xenética entre localidades

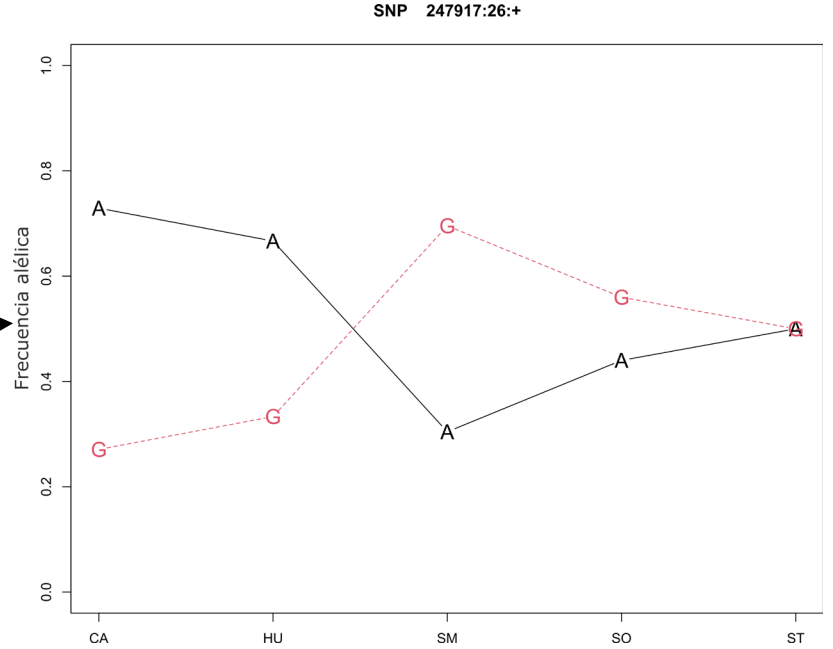
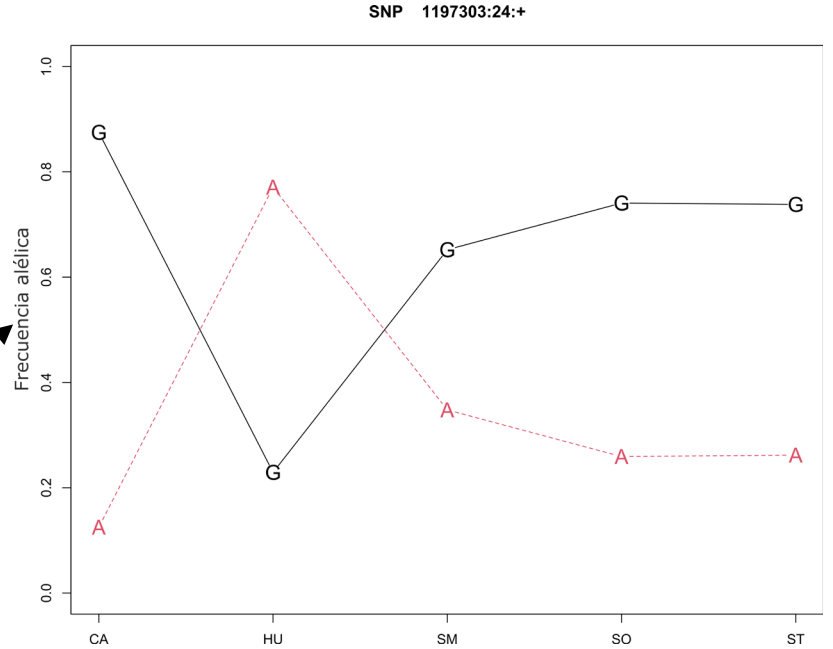
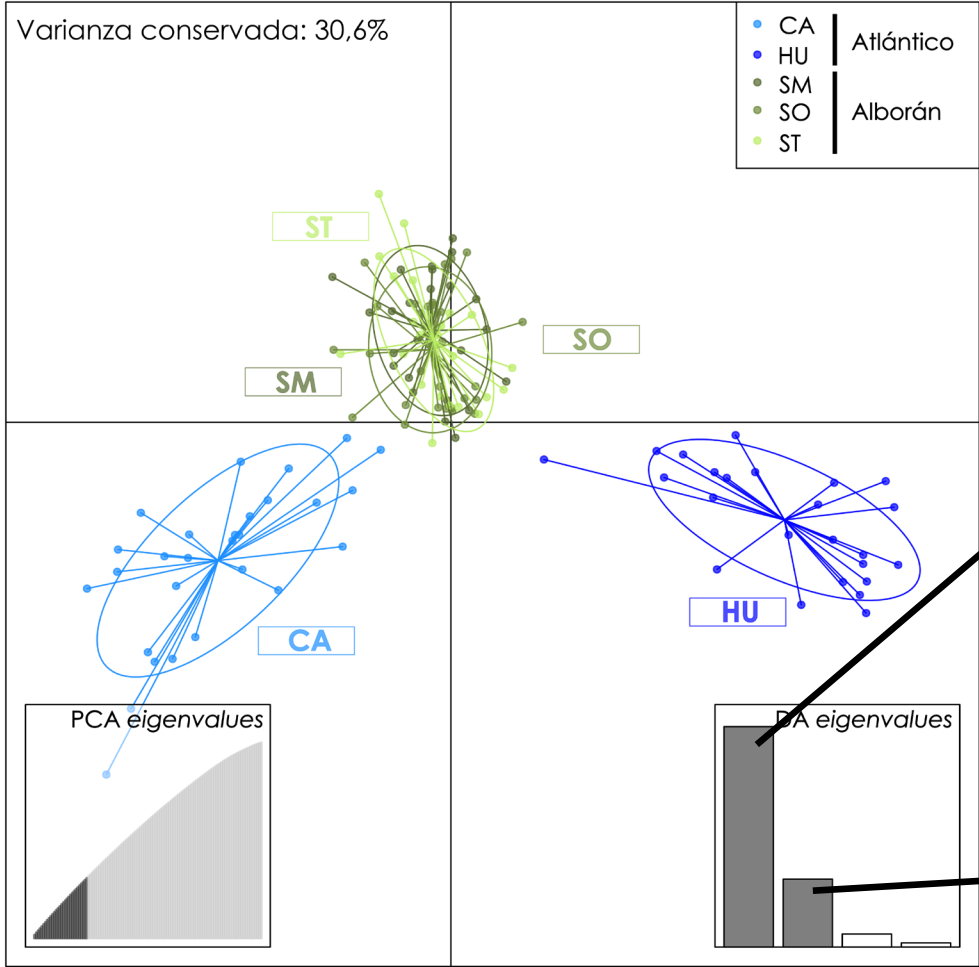
	CA	HU	SM	SO	ST
CA	—	***	***	***	***
HU	0.016	—	***	***	***
SM	0.008	0.006	—	*	NS
SO	0.008	0.006	0.001	—	NS
ST	0.008	0.006	0.000	0.000	—

\*\*\*  $p$ -valor < 0.001, \*\*  $p$ -valor < 0.01, \*  $p$ -valor < 0.05, NS Non Significativo

$$F_{ST} \text{ global} = 0.0062$$



# XI XORNADA DE USUARIOS DE R EN GALICIA





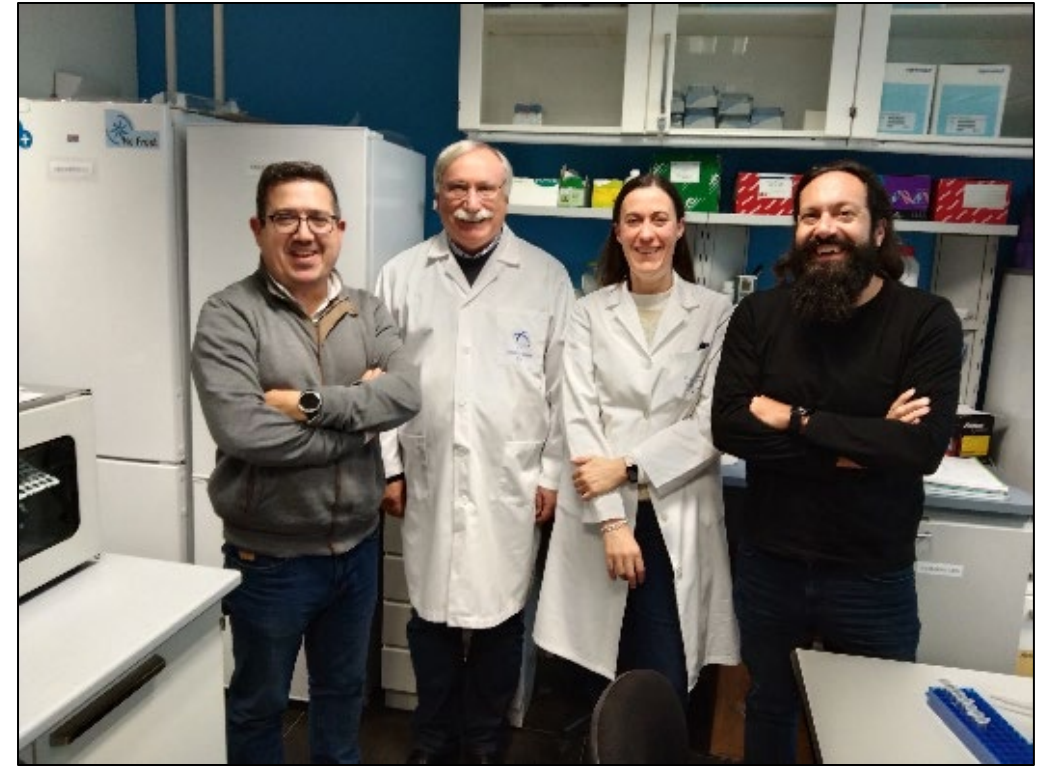
# Xerando máis resultados...



# PRIMEIRAS CONCLUSIÓN

- As localidades de quisquilla do Mar de Alborán poderían ser incluídas na mesma Unidade de Xestión (MU, *Management Unit*).
- A diferenciación xenética coas localidades atlánticas foi moi baixa, especialmente coa das Illas Canarias a pesares de estar a máis de 1.000 quilómetros de distancia.
- É posible que nas diferentes características organolépticas que a quisquilla poida presentar haxa un maior peso das variables ambientais respecto ás variables xenómicas.

## O melloR traballo, en equipo ( $p$ -valor $< 0.0000000000000001$ )



Este traballo realizouse no marco do proxecto "Análisis genético de *Plesionika edwardsii* en poblaciones del Mar de Alborán (PLESIGEN)" B-BIO-678-UGR20, financiado polos Proyectos I+D+i del Programa Operativo FEDER Andalucía 2020 de la Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía. Adrián Casanova é financiado por un contrato posdoutoral Xunta de Galicia-Campus Terra (2022). Agradecemos o apoio bioinformático do CESGA.



Junta de Andalucía

Consejería de Transformación Económica, Industria, Conocimiento y Universidades



UNIVERSIDAD  
DE GRANADA



XUNTA  
DE GALICIA

CONSELLERÍA DE EDUCACIÓN,  
CIENCIA, UNIVERSIDADES E  
FORMACIÓN PROFESIONAL