



CESGA

GALICIA SUPERCOMPUTING CENTER

XORNADA USUARIOS CESGA 2023
Santiago de Compostela

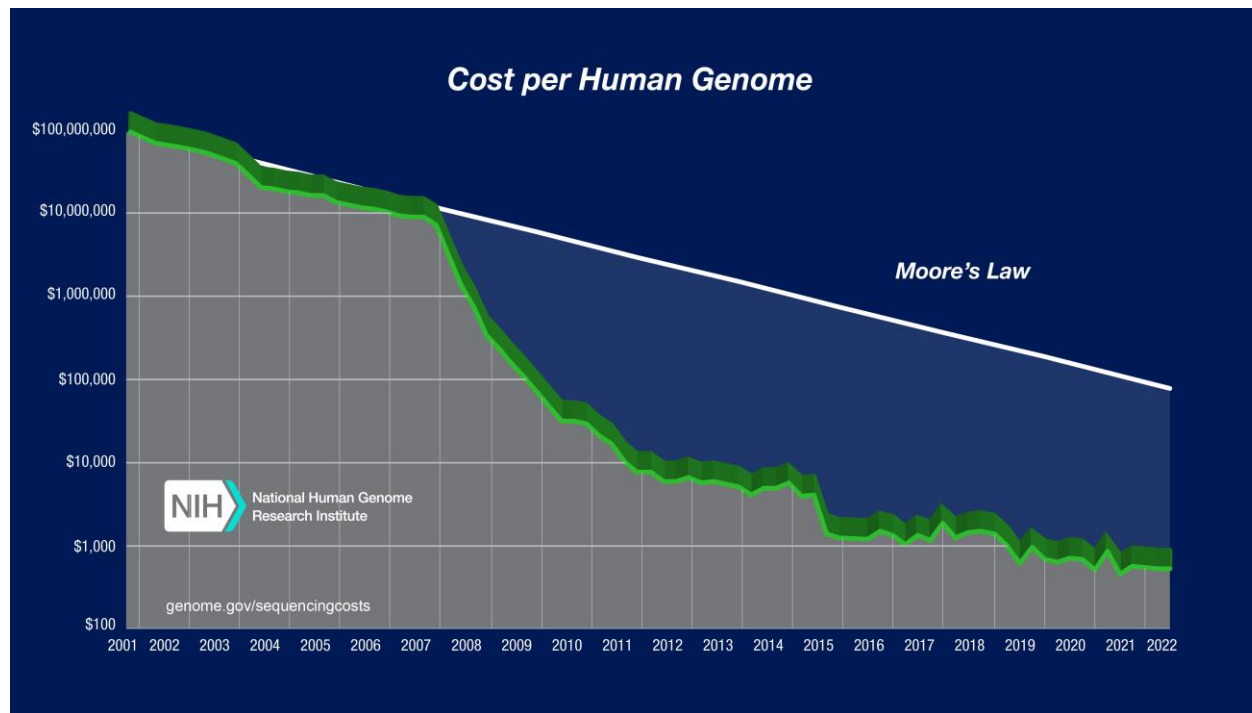
The multi-omics revolution in the big data age: a computational challenge and a scientific opportunity in Genetics

ACUIGEN research group: Casanova, A; Aramburu, O; Arana, AJ; Carreira, M; Fernández, M; Fernández, S; Mejuto, N; Pampín, M; Quelle-Regaldie, A; Sambade, IM; Torres, D; Valín, R; Veiga-Rúa, S; Villamayor, PR; Blanco, A; Cabezas, P; Fernández, C; Hermida, M; Robledo, D; Rubiolo, J; Bouza, C; Pardo, BG; Sánchez, L; Vera, M; Vilas, R; Viñas, A; Martínez, P.

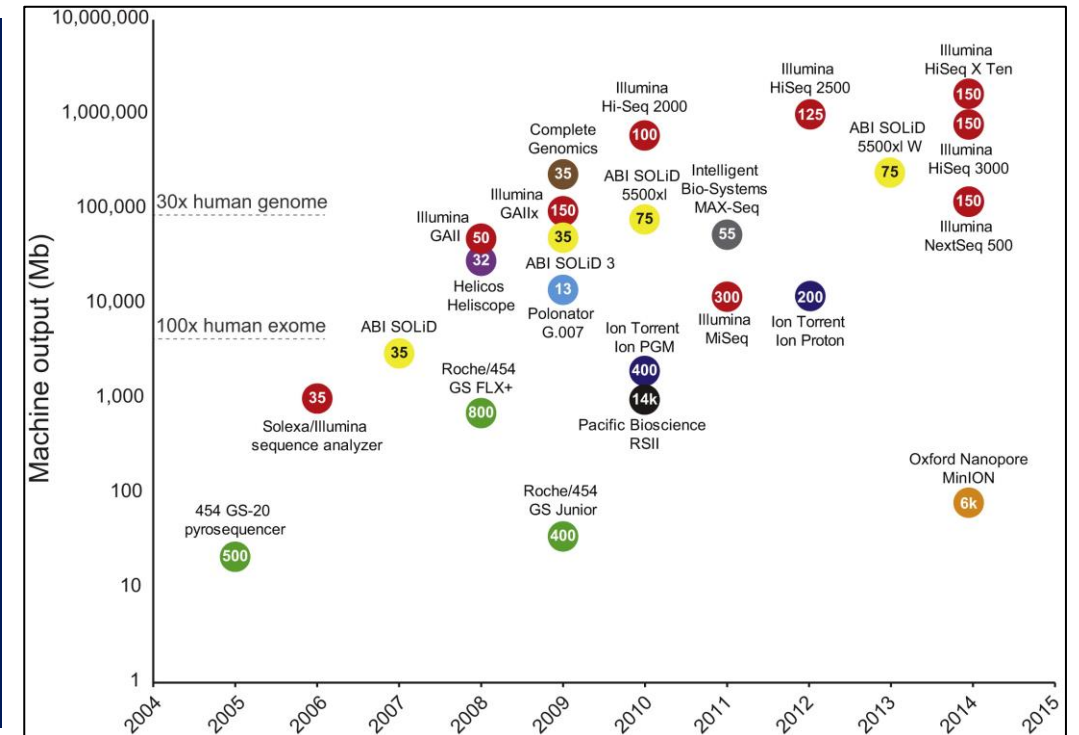
INDEX

1. Introduction
2. Research group objective
3. Methodology
4. Results
5. Take-home messages

High Throughput Sequencing (HTS) technologies have represented a **breakthrough** for genetic studies due to the **sharp reduction of sequencing costs**, implying the **democratization** of massive sequencing. Various techniques for different “-omics” approaches have been developed in the last few years.

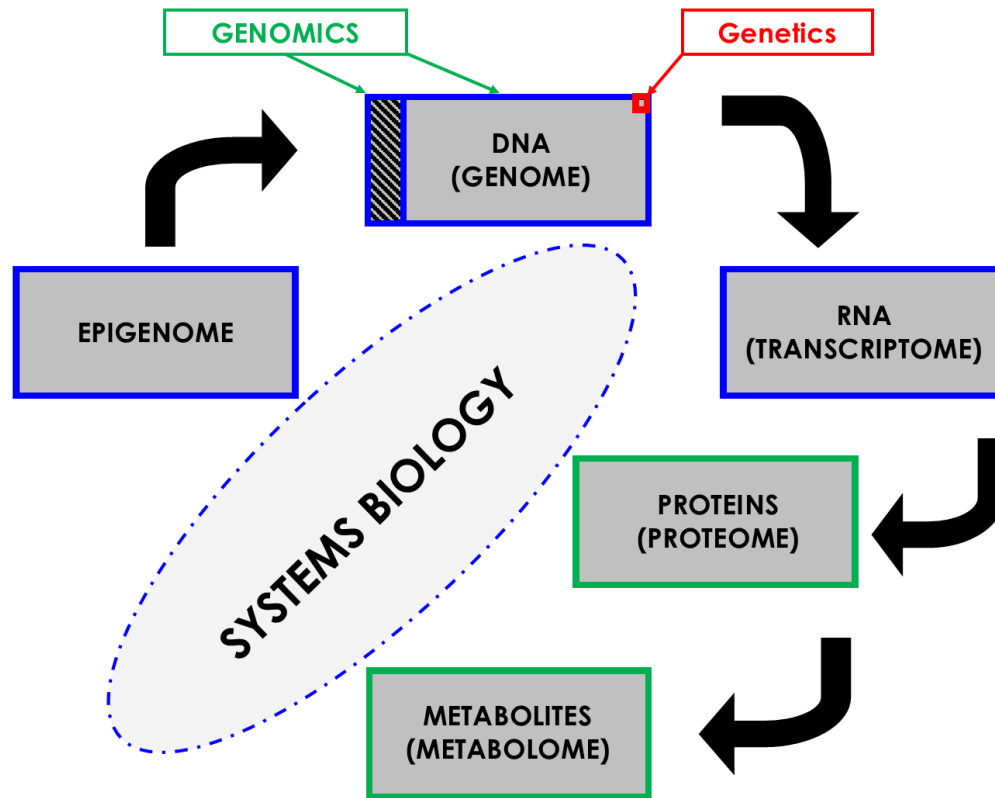


Wetterstrand KA. DNA Sequencing Costs: Data | NHGRI. 2022. Accessed 27 November 2023: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>.



Reuter, J. A., Spacek, D. V., & Snyder, M. P. High-throughput sequencing technologies. *Molecular cell* **58**(4), 586–597 (2015). <https://doi.org/10.1016/j.molcel.2015.05.004>

And what are the -omics data?



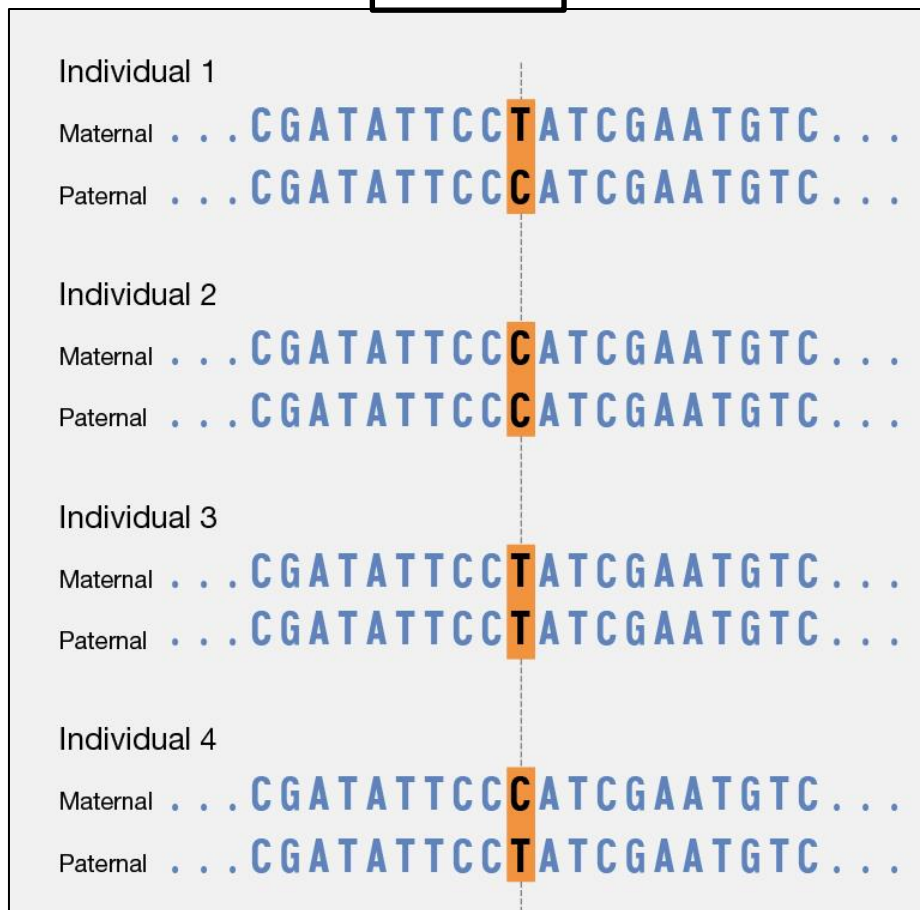
This suffix is used to represent fields of study in life sciences focusing on **large-scale data** of different **biological entities**, such as

- Genome → Genomics
- Transcriptome → Transcriptomics

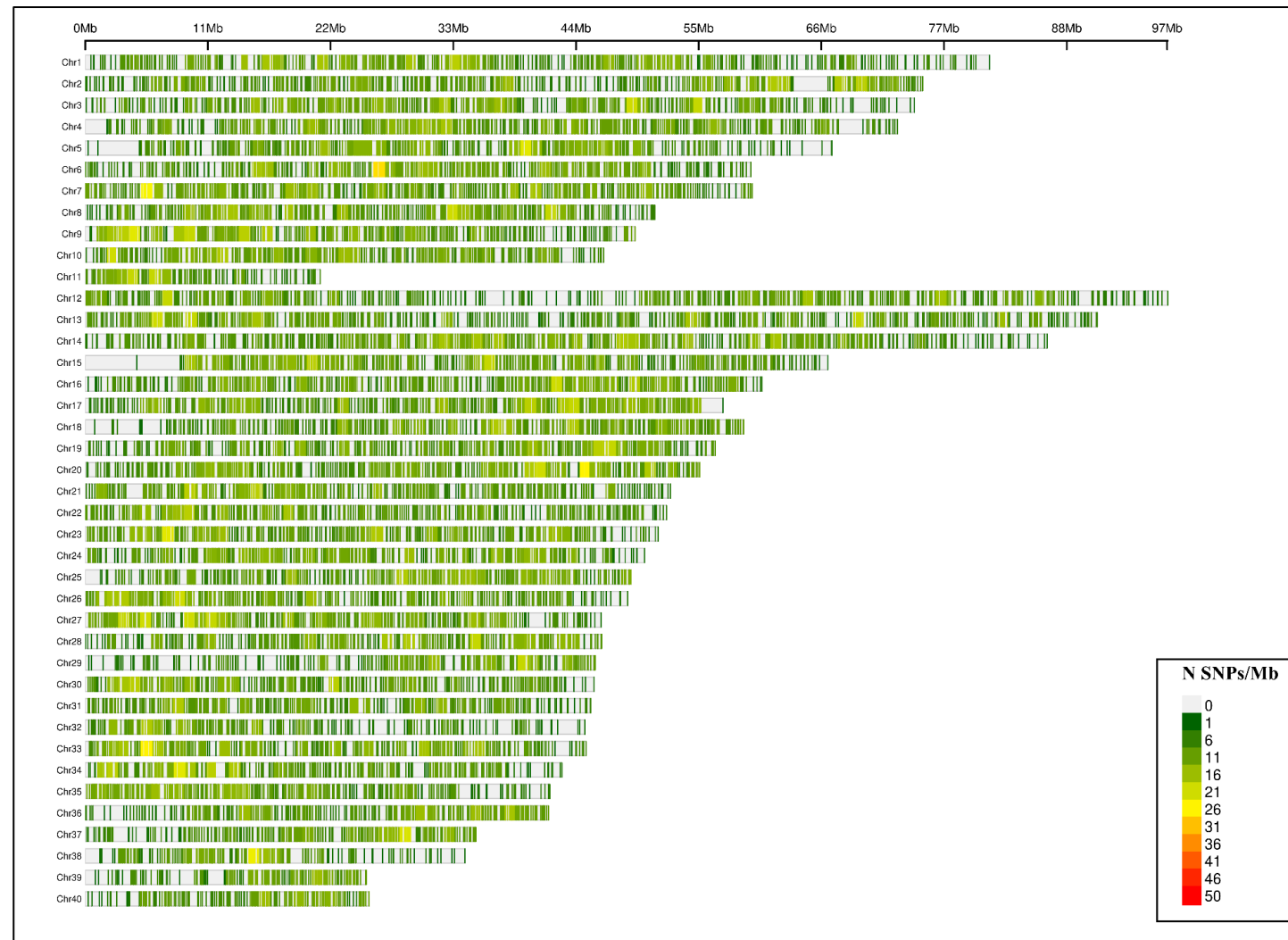
...

And what are the -omics data?

SNPs



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



The scientific activity of ACUIGEN is centered on “-omics” approaches and their applications to **aquaculture, biomedicine,** and the **conservation** of biological resources. This is projected towards the Technological Transfer to Companies and Administrations related to aquaculture and environmental health.

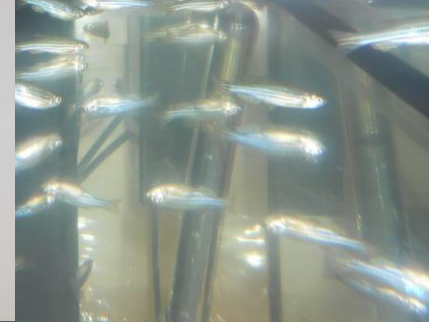
Introduction

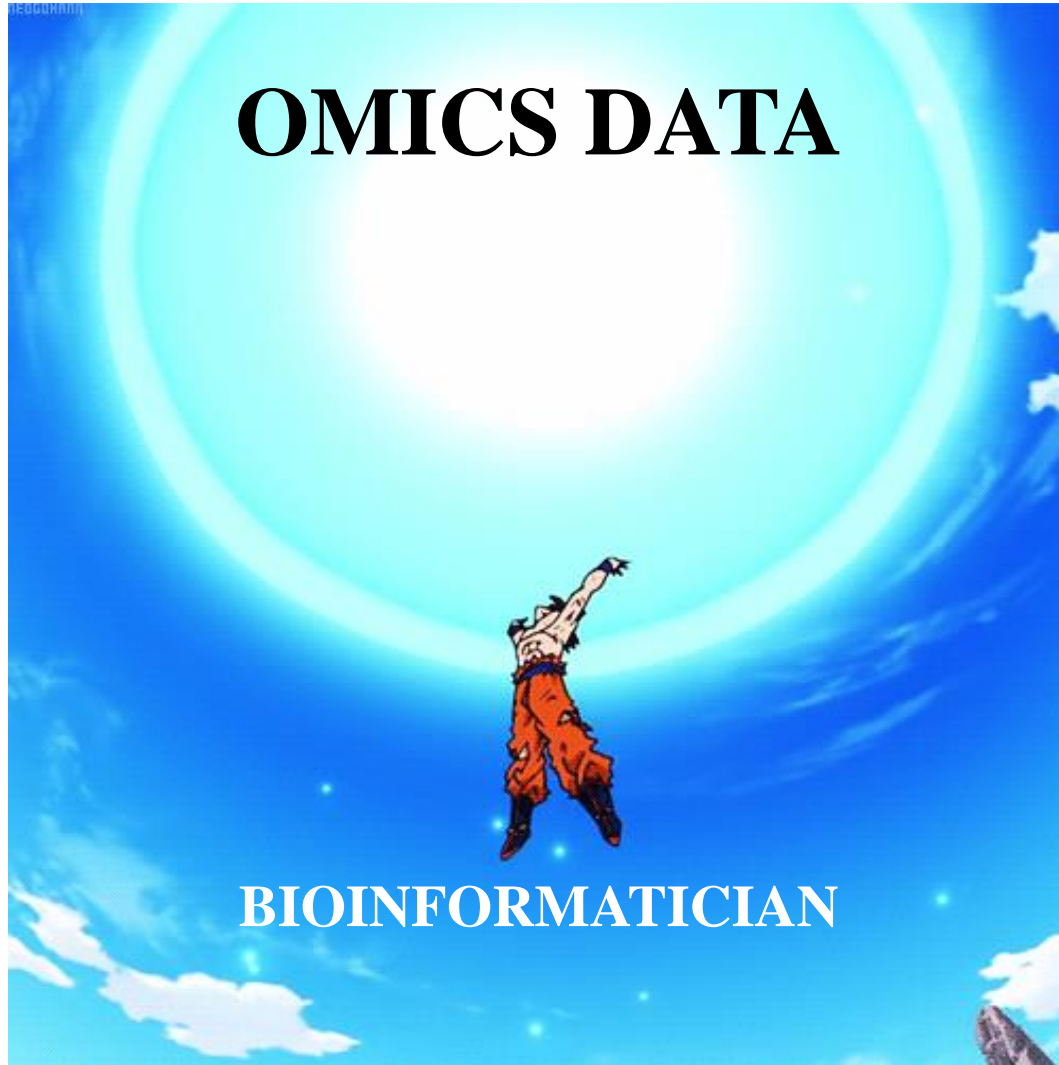
Objective

Methodology

Results

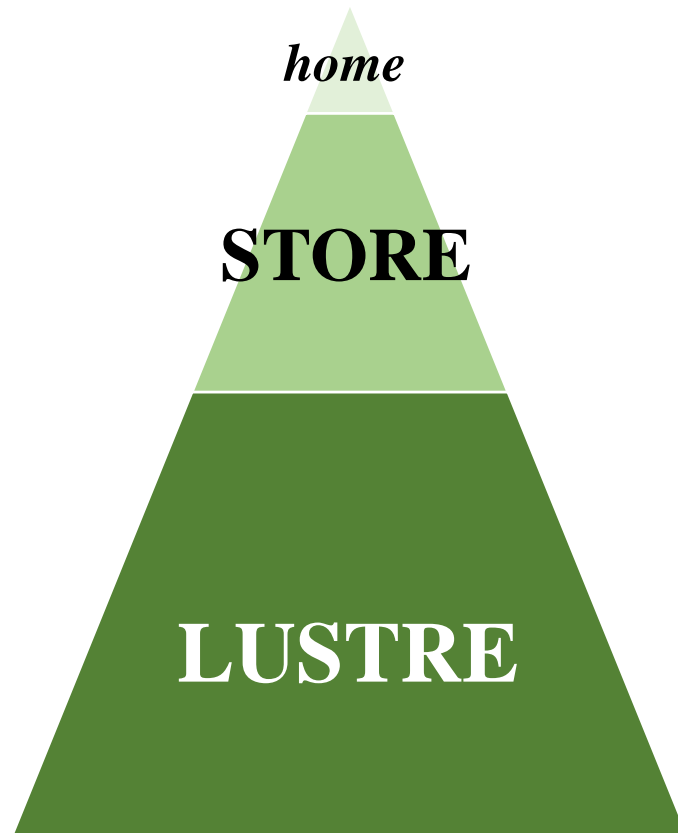
Take-home messages





- ✓ **Hardware**
- ✓ **Software**
- ✓ **Technical support**

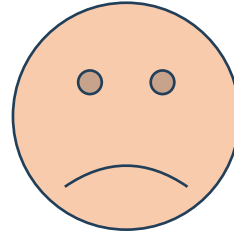




User storage: ~3.5 TB

FINISTERRAE III

	Thin nodes	GPU nodes	Vis nodes	Transfer nodes	Fat nodes	OLM node
CANTIDADE	256	66	16	2	17	Simulador de circuitos cuánticos Atos 30 qbits
PROC.	2x Intel Xeon Ice Lake 8352Y 32 cores 2.2 GHz					
MEMORIA	256 GB				2 TB (16)	
					8 TB (1)	
DISCO LOCAL	960 GB SSD NVMe				1920 GB SSD NVMe	
GPU	2x Nvidia A100 (64)		Nvidia T4			
	5x Nvidia A100 (1) + 256 GB RAM					
	8x Nvidia A100 (1) + 256 GB RAM					
REDE	Infiniband HDR 100					



```

cstacks is done.

*****
*
*          JOB EFFICIENCY REPORT (seff 4981898)
*
*****

Job ID: 4981898
Cluster: finisterrae3
User/Group: uscgeacc/usc
State: COMPLETED (exit code 0)
Nodes: 1
Cores per node: 10
CPU Utilized: 03:40:53
CPU Efficiency: 38.14% of 09:39:10 core-walltime
Job Wall-clock time: 00:57:55
Memory Utilized: 3.69 GB
Memory Efficiency: 18.46% of 20.00 GB

+++++
++  Memory Efficiency is too small. Please review the requested memory.  ++
++  It seems that you do not need that much memory so we recommend    ++
++  requesting less memory in other similar jobs.                      ++
+++++

*****

```

```

*****
*
*          JOB EFFICIENCY REPORT (seff 4987227)
*
*****

Job ID: 4987227
Cluster: finisterrae3
User/Group: uscgeacc/usc
State: COMPLETED (exit code 0)
Nodes: 1
Cores per node: 5
CPU Utilized: 00:28:06
CPU Efficiency: 88.74% of 00:31:40 core-walltime
Job Wall-clock time: 00:06:20
Memory Utilized: 4.55 GB
Memory Efficiency: 90.99% of 5.00 GB

*****
(END)

```

**CESGA IS A PUBLIC RESOURCE
WITH GREAT POWER COMES GREAT RESPONSIBILITY!**



Stacks



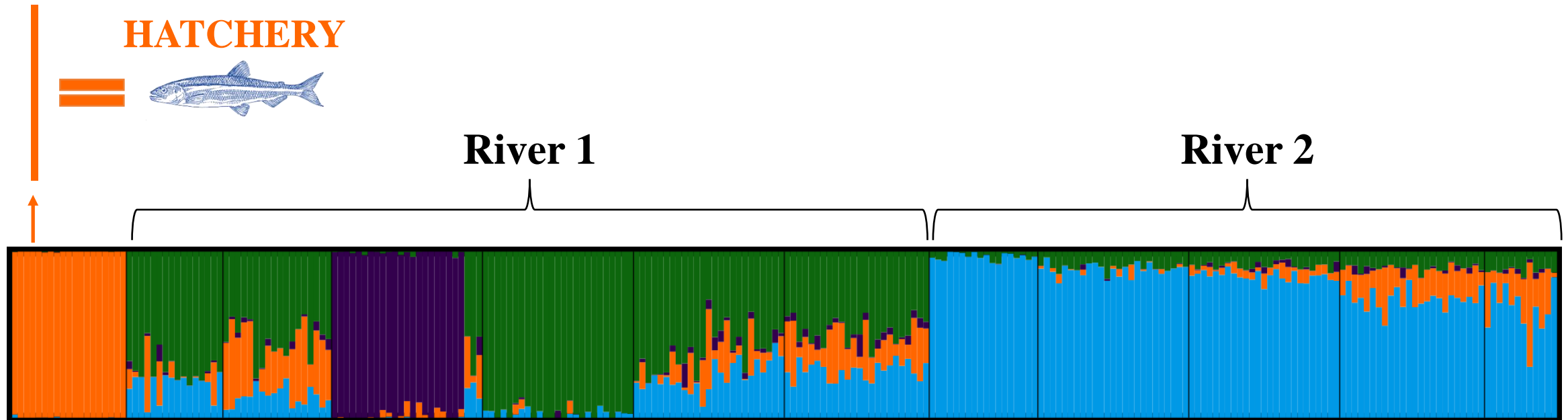
AB₂SS v₂

nf-core

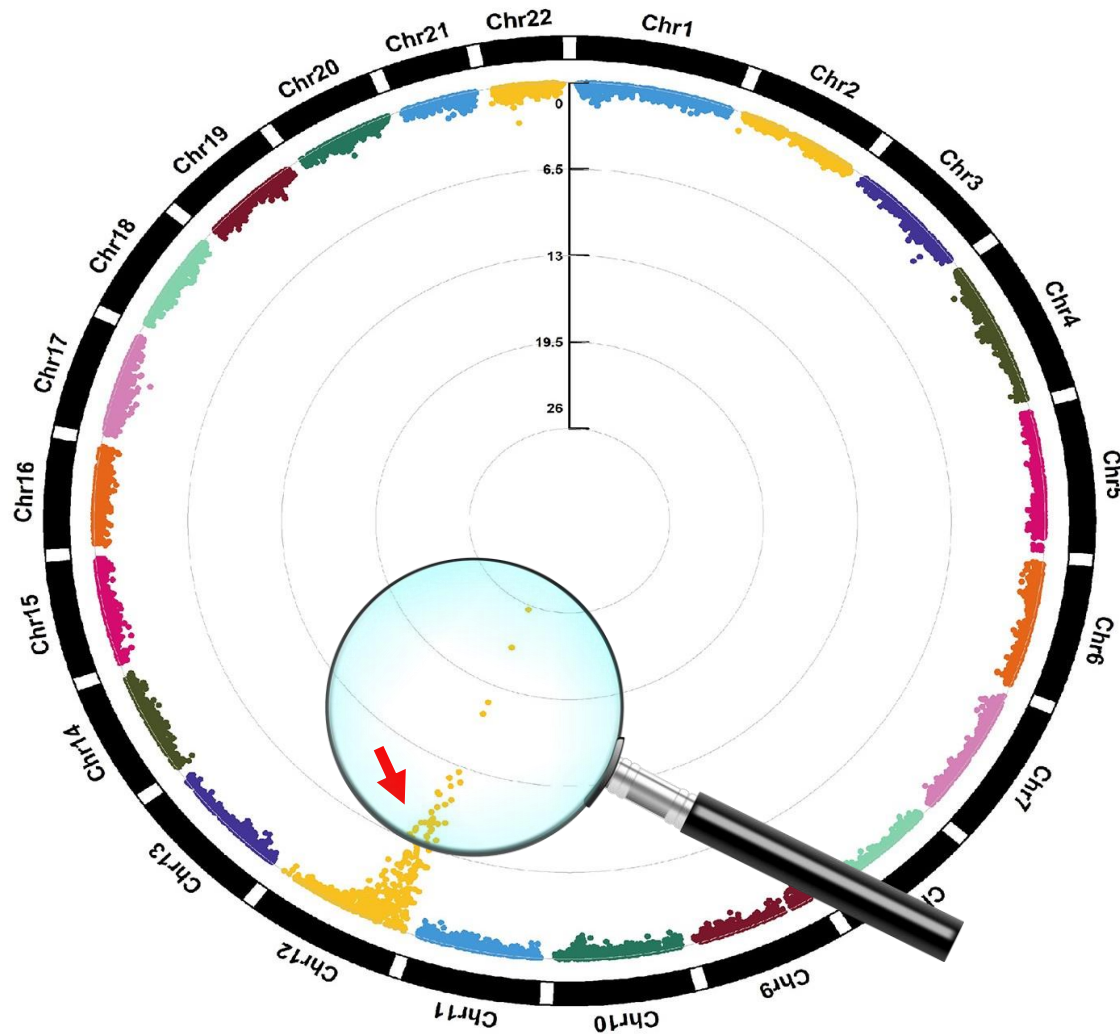


VCFtools

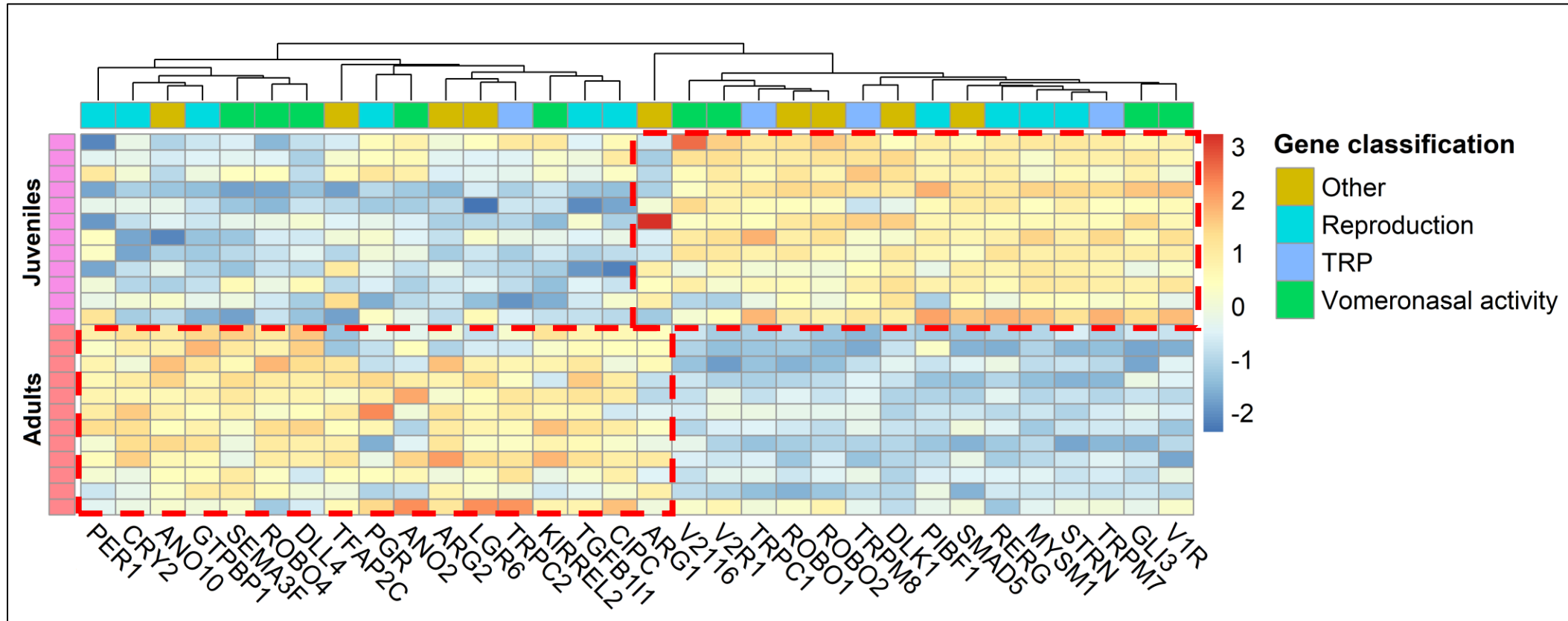




- ❑ It is essential to know the structure and the genetic diversity of wild populations to design adequate conservation management strategies.
- ❑ Brown trout (*Salmo trutta*)
- ❑ R package: {**ParallelStructure**} | Multi-core analysis at CESGA



- ❑ SNPs significantly associated with the sex of the turbot. ZZ/ZW system.
- ❑ Candidate sex determination gene: *sox2*
- ❑ Turbot (*Scophthalmus maximus*)
- ❑ R package: {**GenABEL**}



- Differential expression patterns (RNA) of genes of interest between juveniles and adults.
- Rabbit (*Oryctolagus cuniculus*)
- R package: {**Pheatmap**}

- With **OMICS data**, we can address a **broader range of biological questions**.
- The human, hardware and software resources provided by CESGA are **indispensable** for the bioinformatic analyses of this OMICS data.
- **The use of CESGA in OMICS analyses is fully consolidated in our group.** Following current trends, new tools (hardware + software) will emerge in line with new data and challenges in this dynamic field.

