



Genobiostoic

11111100111

Repaso, verificación de calidad y escogencia del
genoma de referencia: Día3



Elaboración de scripts.



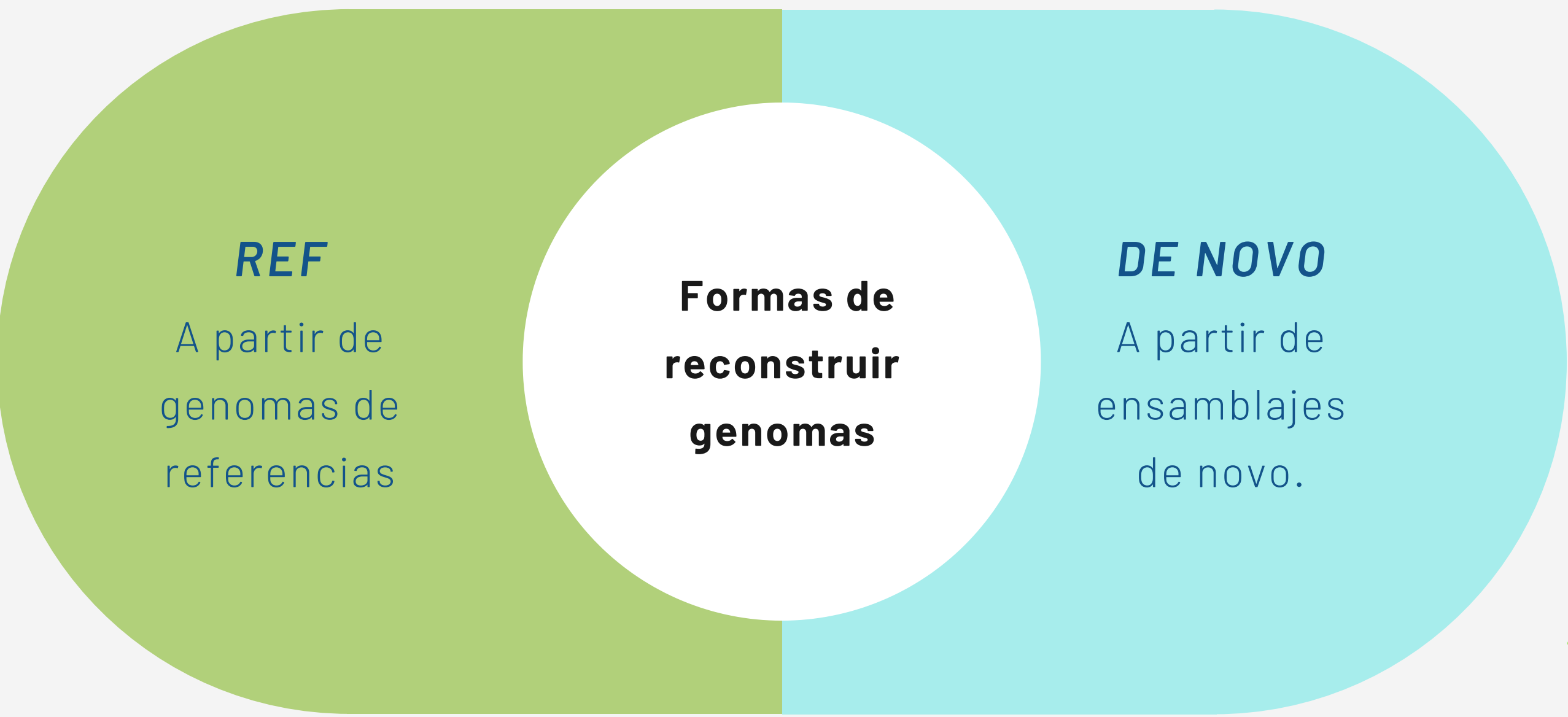
Frontend

nohup ./

```
source activate samtools

samtools view -q 20 -f 0x2 -bSh -@ 20 /data6/testacc/fabiola/10ADI_Enovo.sam > /data6/testacc/fabiola/10ADI_Enovo.bam

samtools view -q 20 -f 0x2 -bSh -@ 20 /data6/testacc/fabiola/10ADYI_Enovo.sam > /data6/testacc/fabiola/10ADYI_Enovo.bam
```



NCBI

- National Center for Biotechnology Information

<https://www.ncbi.nlm.nih.gov/>



Interfaz
amigable



GeneBank

Genomas de referencia para especies no modelo

Atributos

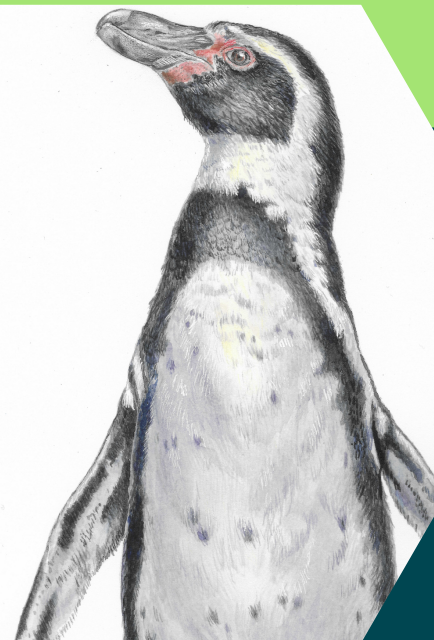


Pudú

Pudu puda

Tamaño del genoma?

Anotación



Pinguino de Humboldt

Spheniscus humboldti

Cobertura de secuenciación?

Tecnología



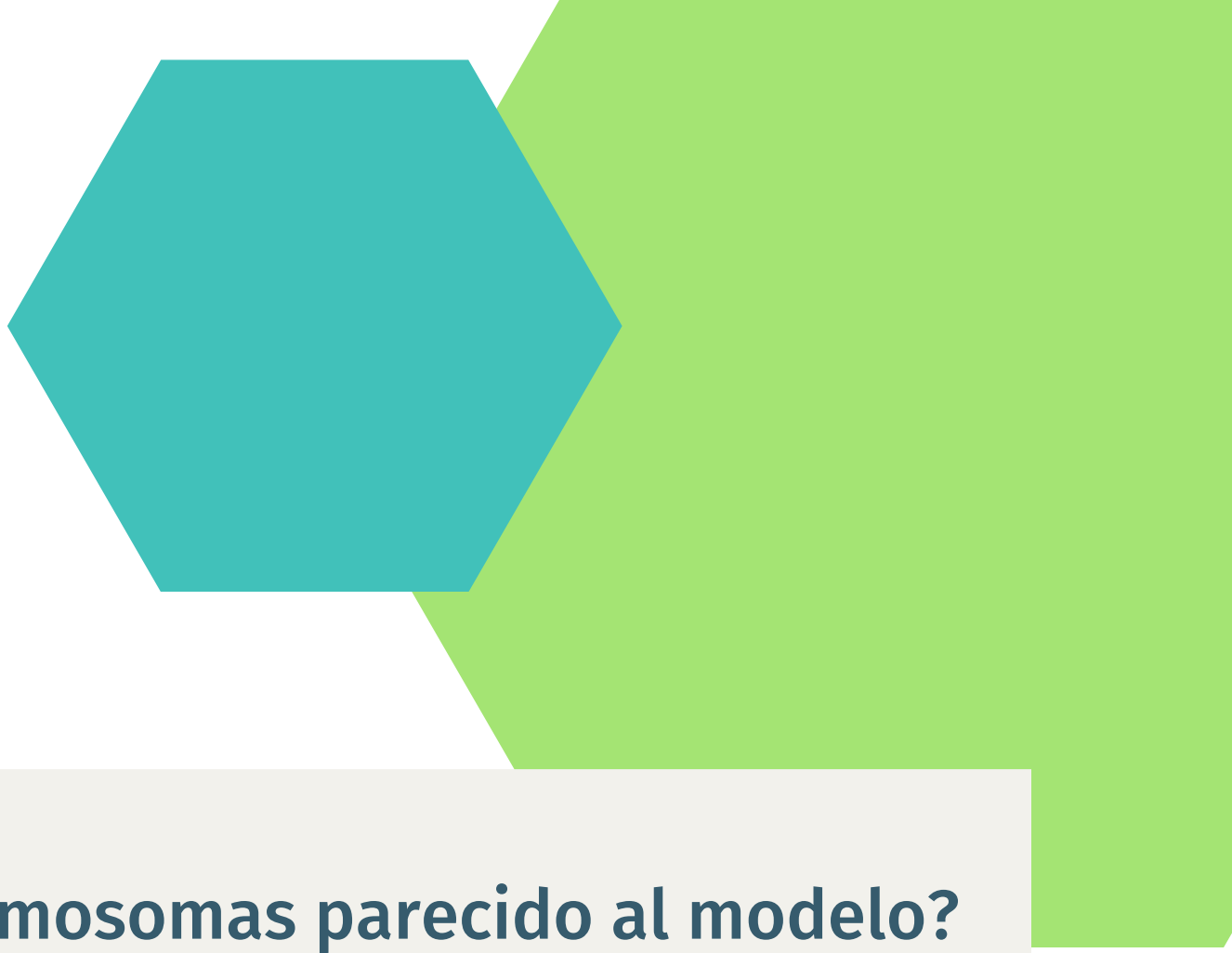
Ranita de Darwin

Rhinoderma darwini

Número de Scaffolds o

Chromosomas?

Consideraciones para escoger Genoma de referencia



Scaffolds vs Cromosomas	Número de Cromosomas parecido al modelo?
Cobertura de Secuenciación	Qué tan lejanos son de nuestro modelo (filogenia)
Formato GenBank vs RefSeq	Tecnología de secuenciación

Backend

qsub

```
#!/bin/bash
#PBS -V
#PBS -N sorted_prueba_nuevaREF_Enovo
#PBS -k eo
#PBS -l nodes=1:ppn=40
#PBS -l walltime=02:00:00
#####
### Switch to the working directory;
cd $PBS_O_WORKDIR
### Run:
date +"%T"

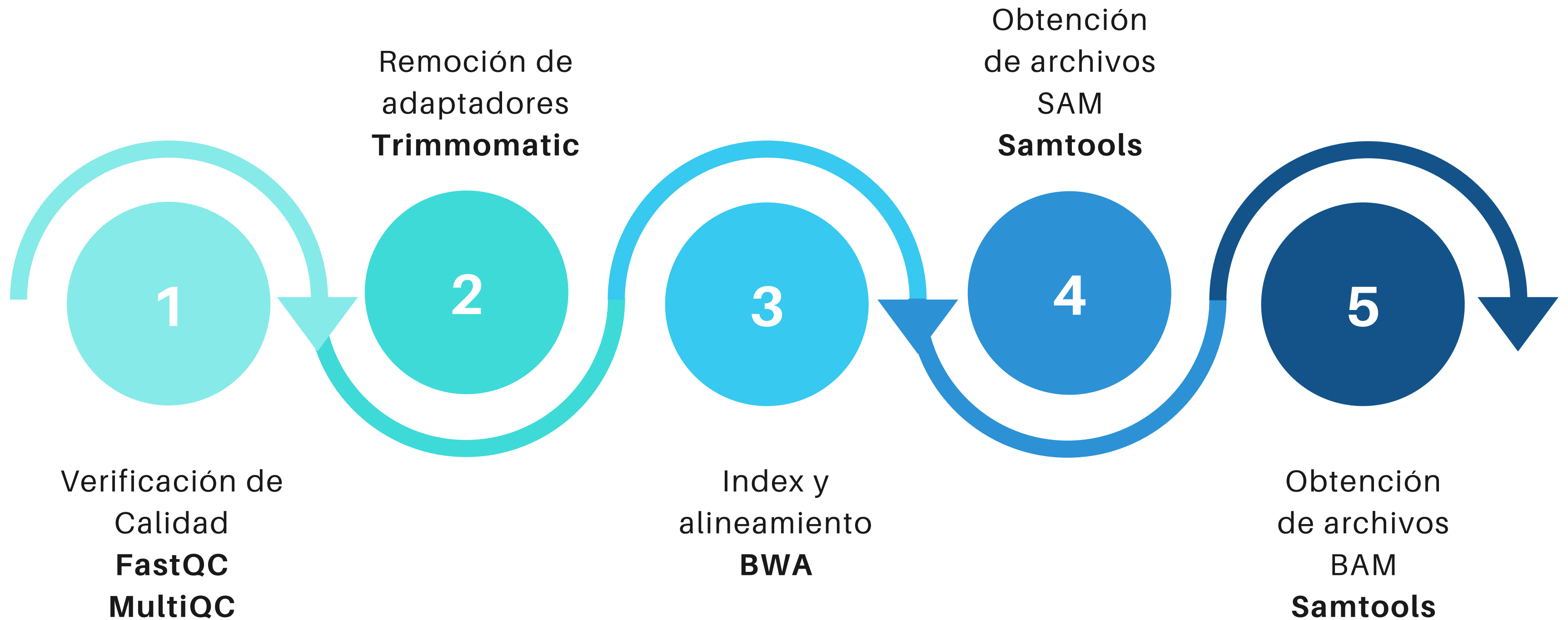
source activate samtools

samtools sort -o /data6/testacc/fabiola/10ADI_sorted_Enovo.bam -@ 40 /data6/testacc/fabiola/10ADI_Enovo.bam
samtools sort -o /data6/testacc/fabiola/10ADYI_sorted_Enovo.bam -@ 40 /data6/testacc/fabiola/10ADYI_Enovo.bam

qstat -f $PBS_JOBID
```


GENÓMICA

Proceso de resecuenciación contra genoma de referencia



Conoce y ajusta a tu proyecto

- El trabajo de ensamblaje debe ajustarse a las preguntas y particularidades de cada genoma.
- Es muy importante **NO** replicar scripts a la ligera.
- Se recomienda estudiar las funciones de cada uno de los programas que se utilizan y ajustar los parámetros según el proyecto

FASTQC y MULTIQC

- Es una herramienta de control de calidad para datos de secuencias NGS
- Posee interfaz grafica amigable o terminal
- Proporciona un conjunto modular de análisis para saber si los datos tienen algún problema que deba tener en cuenta antes de continuar con el ensamblaje
- Entrega archivos html de fácil visualización



[Volver a la página de la agenda](#)

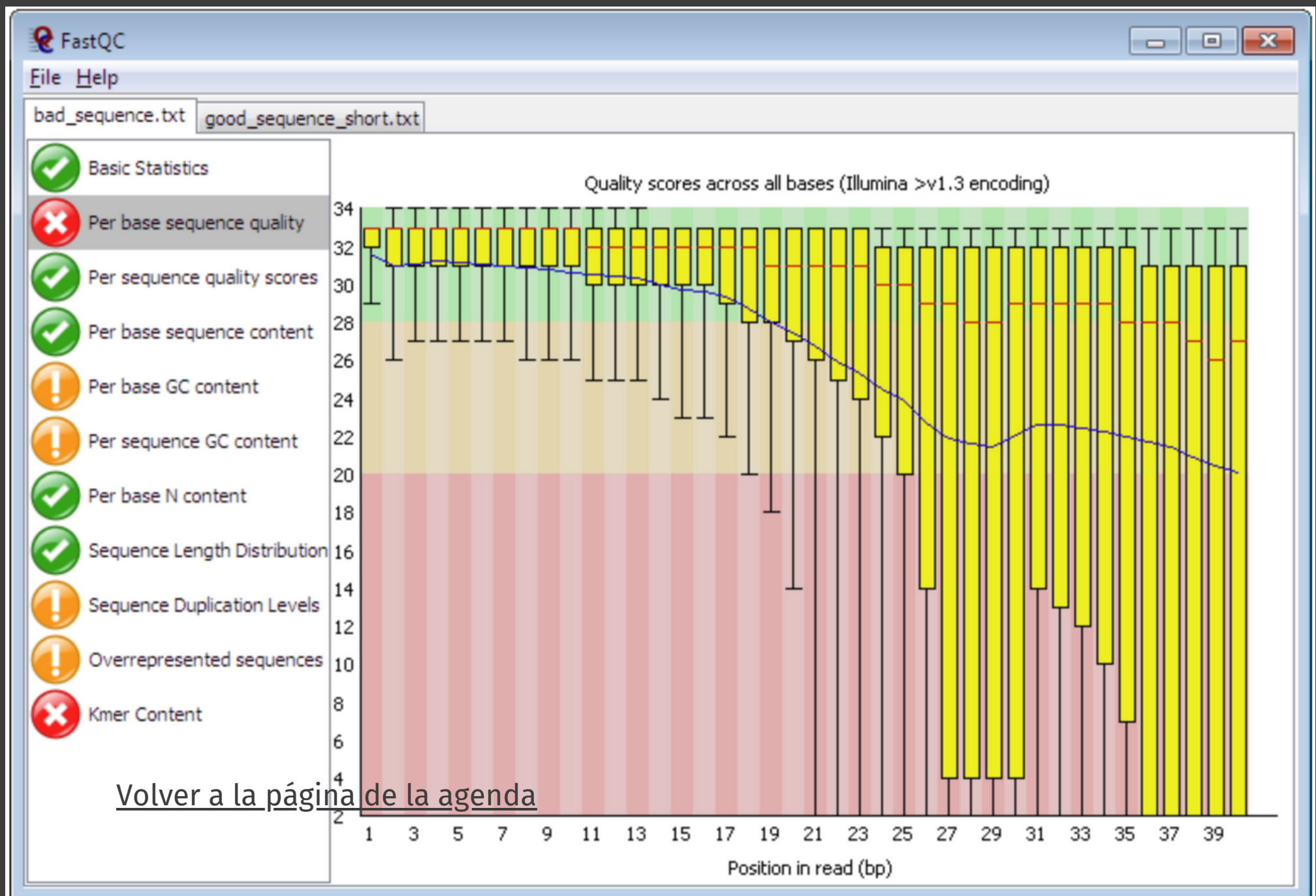


MultiQC

Phil Ewels
phil.ewels@scilifelab.se

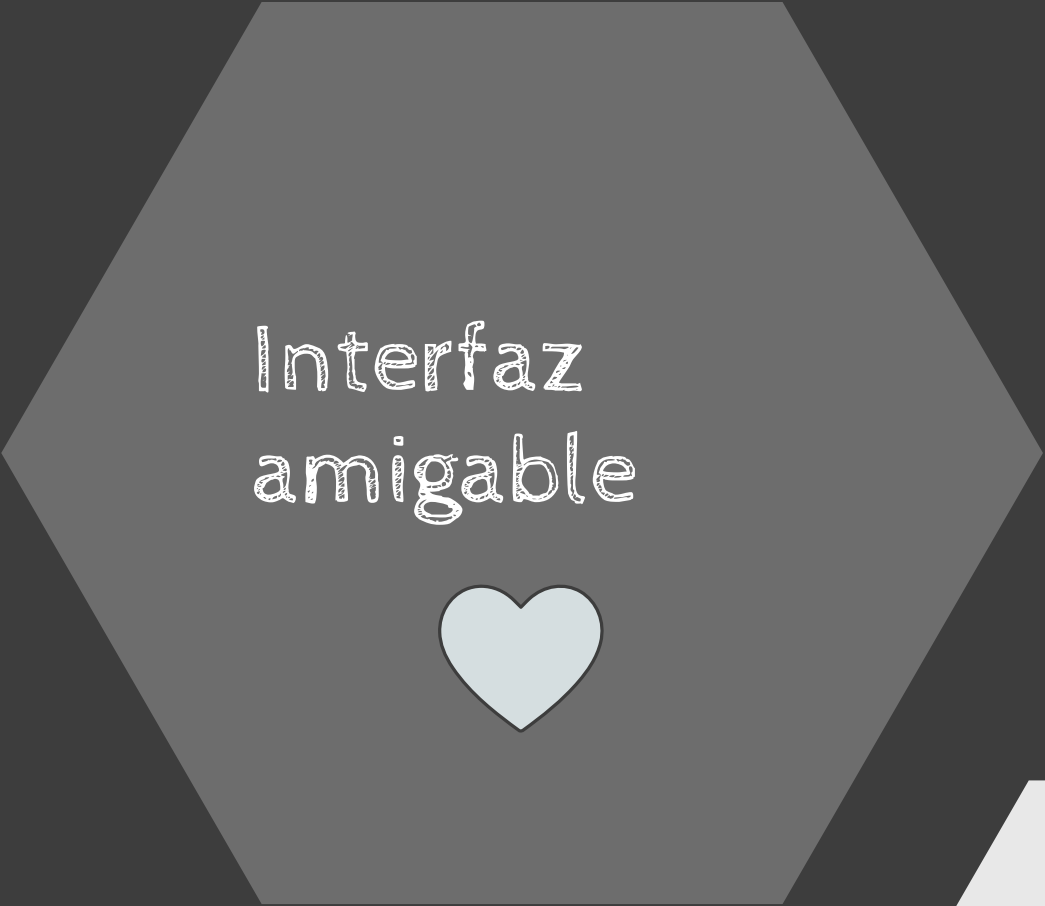
FASTQC

Interfaz amigable 



Entrega un resumen de un sólo genoma

MULTIQC



General Statistics

Copy table

Configure Columns

Sort by highlight

Plot

Showing 8/8 rows and 9/11 columns.

Sample Name	5'-3' bias	M Aligned	% Aligned	M Aligned	% Aligned	M Aligned	% Dups	% GC	M Seqs
Irrel_kd_1	1.18	35.6	86.4%	31.2	92.1%	33.2	55.9%	47%	36.1
Irrel_kd_2	1.14	30.4	86.0%	26.5	92.2%	28.4	53.6%	47%	30.8
Irrel_kd_3	1.19	23.6	85.7%	20.5	92.0%	22.0	50.1%	48%	23.9
Mov10_kd_2	1.13	51.9	86.0%	45.3	91.6%	48.3	60.5%	48%	52.7
Mov10_kd_3	1.13	30.7	86.0%	26.8	91.6%	28.5	54.6%	47%	31.1
Mov10_oe_1	1.09	38.1	80.2%	32.1	88.9%	35.5	56.5%	47%	40.0
Mov10_oe_2	1.18	35.4	81.0%	30.0	88.8%	33.0	55.9%	48%	37.1
Mov10_oe_3		20.3	81.5%	17.3	90.0%	19.1	50.1%	47%	21.2



[Volver a la página de la agenda](#)

Burrows Wheeler Aligner (BWA)



Es un programa para alinear las lecturas o reads de secuenciación contra un genoma de referencia

Qué es un alineamiento?



Indexar REF

BWA index

Crea un índice de genoma de REF



.amb
.ann
.pac
.bwt
.sa

Alinear

BWA MEM

Alineamientos de short reads contra genoma de REF



.sam



[Volver a la página de la agenda](#)

