

ómica

A stylized graphic element on the right side of the word 'ómica'. It features a white DNA double helix structure. The top strand of the helix is composed of a series of white circles, each containing a binary digit (0 or 1). The bottom strand is a solid white line. The helix curves upwards and to the right, ending in a small cluster of circles.

El futuro
es de todos

Gobierno
de Colombia



Week report

Camila Riccio Rengifo

August 28th, 2020

In-silico



El futuro
es de todos

Gobierno
de Colombia



Salt-stress reads assembly

Index reference genome

```
bwa index IRGSP-1.0_genome.fasta
```

Download genotype files (control x2 and stress x2)

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR549/002/SRR5495129/SRR5495129.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR549/002/SRR5495130/SRR5495130.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR549/002/SRR5495131/SRR5495131.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR549/002/SRR5495132/SRR5495132.fastq.gz
```

Concat files from the same genotype

```
zcat SRR5495129.fastq.gz SRR5495130.fastq.gz SRR5495131.fastq.gz SRR5495132.fastq.gz | gzip -c > 101.fastq.gz
rm SRR5495129.fastq.gz SRR5495130.fastq.gz SRR5495131.fastq.gz SRR5495132.fastq.gz
```

Align fastq file to the reference genome

```
bwa mem -t 8 ../reference/IRGSP-1.0_genome.fasta 101.fastq.gz > 101.sam
rm 101.fastq.gz
```

Transform your .sam file to binary format, sort it and generate a new file with extension .bam

```
samtools view -b 101.sam | samtools sort -t 8 -O bam -T tmp_ -o ../aligned/101.bam
rm 101.sam
```

Index bam file

```
samtools index ../aligned/101.bam
```

.fastq.gz X4
~600 MB c/u



.fastq.gz
~5 GB



.sam
~25 GB



.bam
~4 GB



.bai
~600 kB

~6 h
Camila

~25 min
Mauricio

~1 h
Camila

~2.5 h
Mauricio

In-silico



El futuro
es de todos

Gobierno
de Colombia



Salt-stress reads assembly

.bam x92
~4 GB c/u



.bai x92
~600 kB c/u

call variants from aligned files

```
bcftools mpileup --per-sample-mF -annotate FORMAT/AD,FORMAT/ADF,FORMAT/ADR,FORMAT/SP,INFO/AD,INFO/ADF,INFO/ADR FORMAT/DP  
-Ou -f ../reference/IRGSP-1.0_genome.fasta ../aligned/*.bam | bcftools call -vm0 z -o ../vcf/results.vcf.gz
```

unzip vcf file

```
gunzip ../vcf/results.vcf.gz
```

Remove indels and missing data, call maximum two alleles per locus with a Phred score quality of 30, a deep of 6, and a minor allele frequency of 0.05.

```
vcftools --vcf ../vcf/results.vcf --remove-indels --max-missing 1 --max-alleles 2 --minQ 30 --minDP 6 --maf 0.05 --recode  
--recode-INFO-all --out ../vcf/snpsq30dp6.vcf
```

compute number of homozygous and heterozygous markers

```
vcftools --vcf ../vcf/snpsq30dp6.vcf -het
```

sort the information for analysis in R or another programs

```
bcftools query -H -f '%CHROM %POS [%t%TGT]\n' ../vcf/snpsq30dp6.vcf > R.vcf
```

In-silico



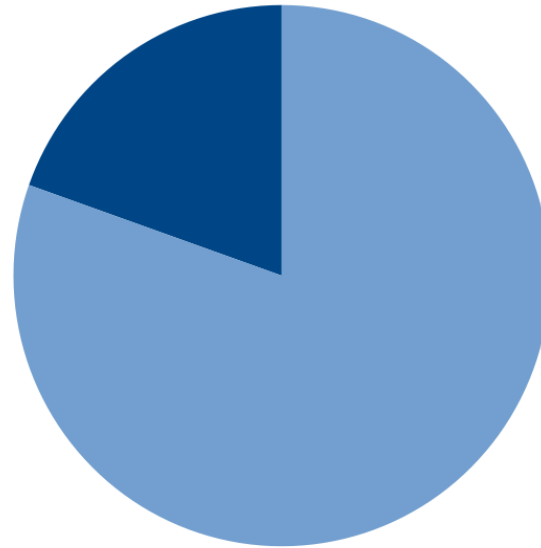
El futuro
es de todos

Gobierno
de Colombia



Progress

Complete
18 genotypes
20%



Incomplete
74 genotypes
80%

In-silico



El futuro
es de todos

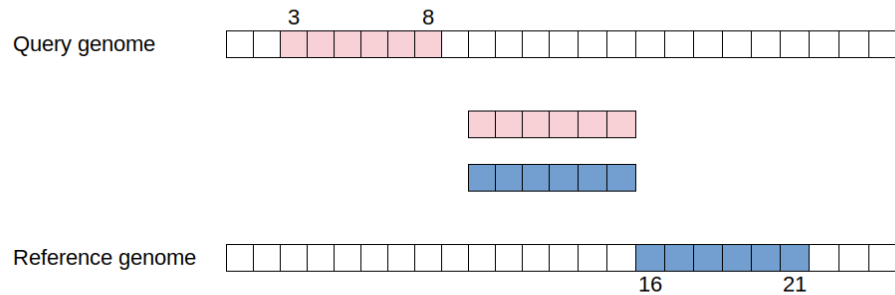
Gobierno
de Colombia



Genomes comparison

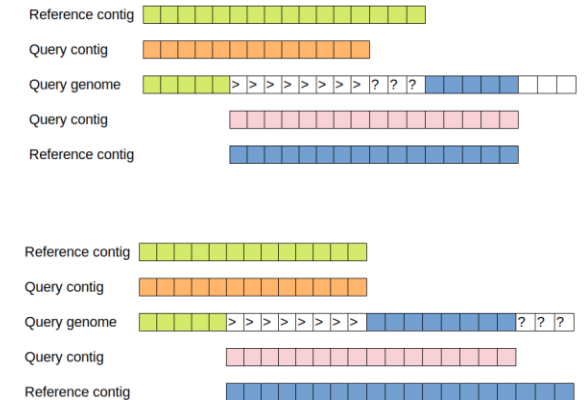
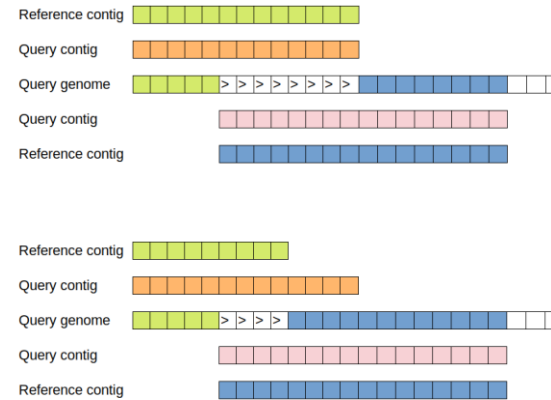
	Sr	Er	Sq	Eq	len_r	len_q	%IDY
0	8701	184119	1	175451	175419	175451	99.88
1	184133	193002	176105	184984	8870	8880	98.93
2	193099	218454	184979	210400	25356	25422	98.49
3	218635	246229	225148	252840	27595	27693	98.36
4	247275	249168	252834	254721	1894	1888	98.94

5.208 contigs



	qpos	qbase	rpos	rbase	inversion
0	1	a	8701.0	a	False
1	2	a	8702.0	a	False
2	3	a	8703.0	c	False
3	4	c	8704.0	c	False
4	5	c	8705.0	c	False

44.011.168 Mb



In-silico





Aliados



Apoyan



El futuro
es de todos

Gobierno
de Colombia

