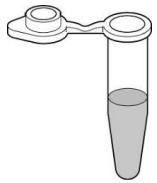
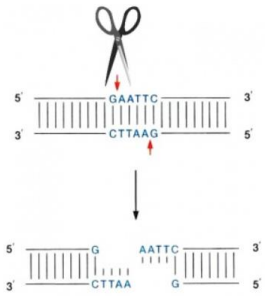
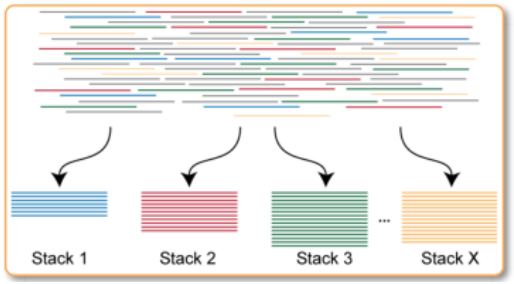
¿Cómo se preparan los datos genómicos?





Secuenciación illumina



GB o TB de información

Alineación al genoma de referencia

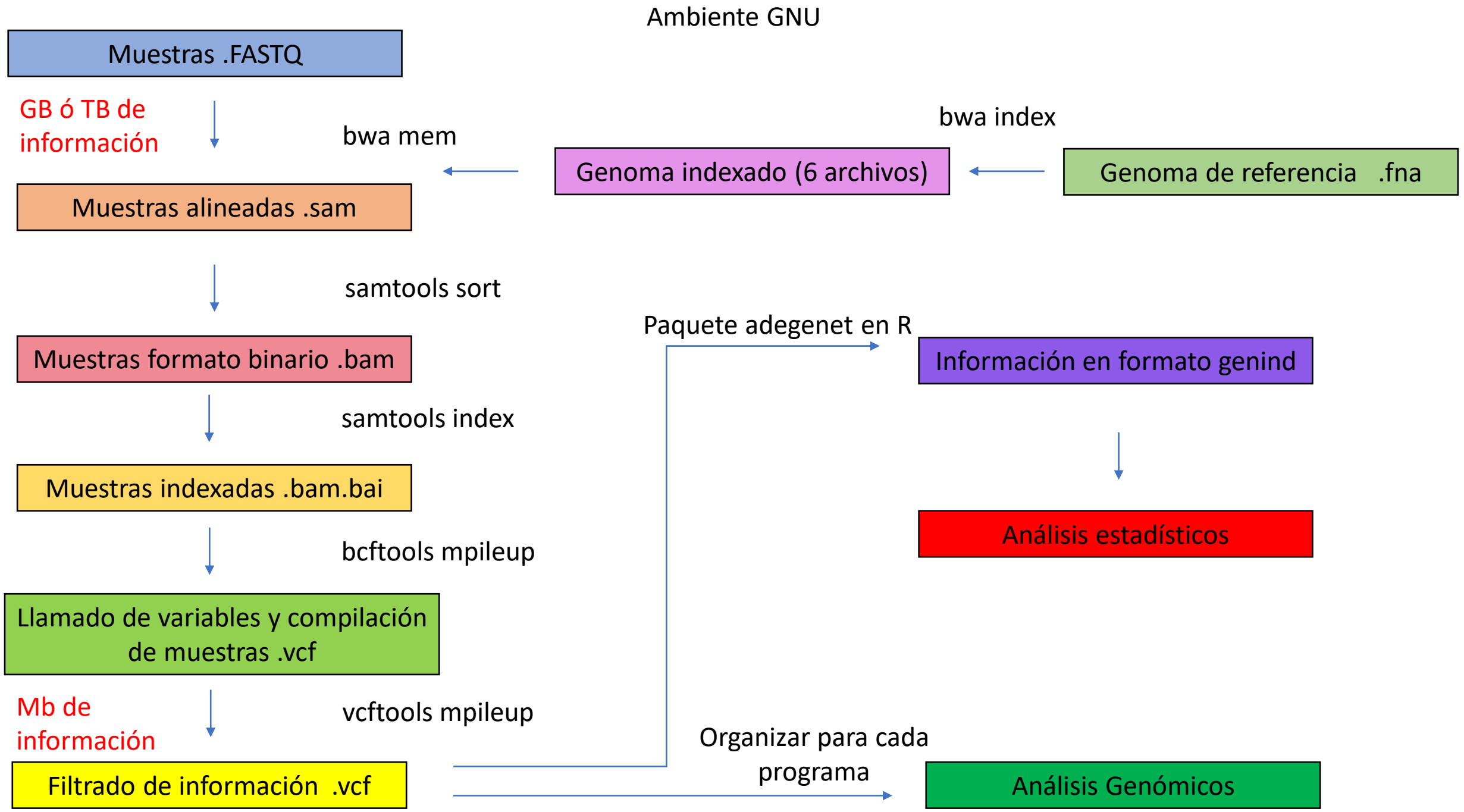
Llamado de variables (SNPs)

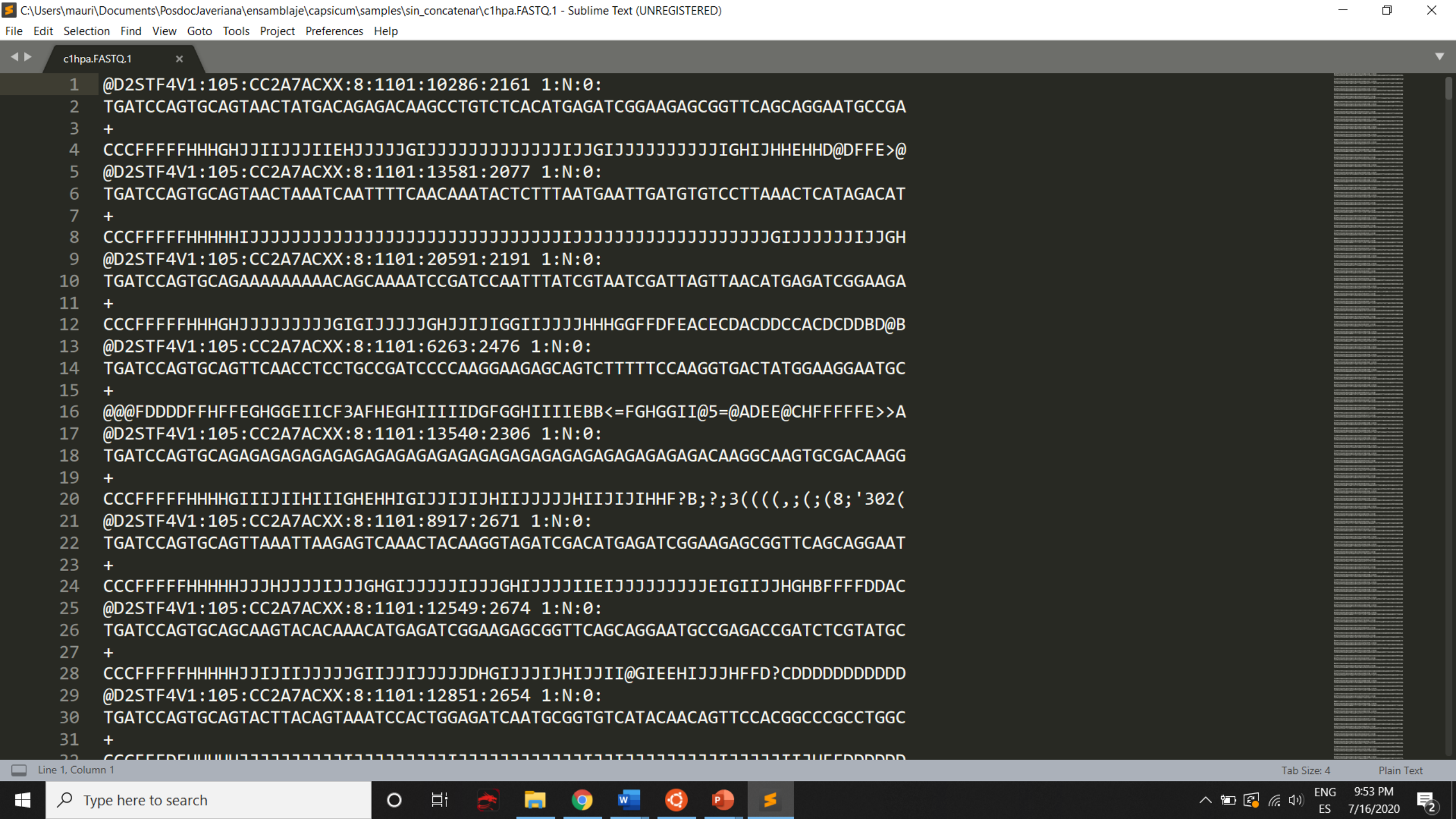
TGACGGTCA
TGACGGACA

Codificación de la informacion (MB)

TGACGGTCA TGACGGTCA	0	-1
TGACGGTCA TGACGGACA	1	0
TGACGGACA TGACGGACA	2	1

Análisis de la información y aplicación de modelos


Ambiente GNU
Muestras .FASTQ
GB ó TB de información
bwa mem
bwa index
Genoma indexado (6 archivos)
Genoma de referencia .fna
Muestras alineadas .sam
samtools sort
Paquete adegenet en R
Muestras formato binario .bam
Información en formato genind
samtools index
Muestras indexadas .bam.bai
Análisis estadísticos
bcftools mpileup
Llamado de variables y compilación de muestras .vcf
Mb de información
vcftools mpileup
Organizar para cada programa
Filtrado de información .vcf
Análisis Genómicos


C:\Users\mauri\Documents\PosdocJaveriana\ensamblaje\capsicum\samples\sin_concatenar\c1hpa.FASTQ.1 - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help
c1hpa.FASTQ.1
@D2STF4V1:105:CC2A7ACXX:8:1101:10286:2161 1:N:0:
TGATCCAGTGCAGTAACTATGACAGAGACAAGCCTGTCTCACATGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
+
CCCFFFFFHHHGHJJIIJJJIIEHJJJJJGIJJJJJJJJJJJJJJIJJGIJJJJJJJJJJIGHIJHHEHHD@DFFE>@
@D2STF4V1:105:CC2A7ACXX:8:1101:13581:2077 1:N:0:
TGATCCAGTGCAGTAACTAAATCAATTTTCAACAAATACTCTTTAATGAATTGATGTGTCCTTAAACTCATAGACAT
+
CCCFFFFFHHHHHIJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJGIJJJJJJIJJGH
@D2STF4V1:105:CC2A7ACXX:8:1101:20591:2191 1:N:0:
TGATCCAGTGCAGAAAAAAAAACAGCAAAATCCGATCCAATTTATCGTAATCGATTAGTTAACATGAGATCGGAAGA
+
CCCFFFFFHHHGHJJJJJJJJJGIGIJJJJJGHJJIJIGGIIJJJJHHHGGFFDFEACECDACDDCCACDCDDBD@B
@D2STF4V1:105:CC2A7ACXX:8:1101:6263:2476 1:N:0:
TGATCCAGTGCAGTTCAACCTCCTGCCGATCCCCAAGGAAGAGCAGTCTTTTTCCAAGGTGACTATGGAAGGAATGC
+
@@@FDDDDFFHFFEGHGGEIICF3AFHEGHIIIIIDGFGGHIIIIEBB<=FGHGGII@5=@ADEE@CHFFFFFE>>A
@D2STF4V1:105:CC2A7ACXX:8:1101:13540:2306 1:N:0:
TGATCCAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACAAGGCAAGTGCGACAAGG
+
CCCFFFFFHHHHGIIIJIIHIIIGHEHHIGIJJIJIJHIIJJJJJHIIJIJIHHF?B;?;3((((,;(;(8;'302(
@D2STF4V1:105:CC2A7ACXX:8:1101:8917:2671 1:N:0:
TGATCCAGTGCAGTTAAATTAAGAGTCAAACTACAAGGTAGATCGACATGAGATCGGAAGAGCGGTTCAGCAGGAAT
+
CCCFFFFFHHHHHJJJHJJJJIJJJGHGIJJJJJIJJJGHIJJJJIIEIJJJJJJJJJEIGIIJJHGHBFFFFDDAC
@D2STF4V1:105:CC2A7ACXX:8:1101:12549:2674 1:N:0:
TGATCCAGTGCAGCAAGTACACAAACATGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGC
+
CCCFFFFFHHHHHJJIJIIJJJJJGIIJJIJJJJJDHGIJJJIJHIJJII@GIEEHIJJJHFFD?CDDDDDDDDDD
@D2STF4V1:105:CC2A7ACXX:8:1101:12851:2654 1:N:0:
TGATCCAGTGCAGTACTTACAGTAAATCCACTGGAGATCAATGCGGTGTCATACAACAGTTCCACGGCCCGCCTGGC
+
Line 1, Column 1
Tab Size: 4
Plain Text
Type here to search
ENG
ES
9:53 PM
7/16/2020

C:\Users\mauri\Documents\PosdocJaveriana\ensamblaje\capsicum\samples\vcf\study_filtered.vcf - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

c1hpa.FASTQ.1 | capsicumsamtools.vcf | study_filtered.vcf

```
81406  ##FILTER=<ID=LOWQUAL,Description="Set if not true: %QUAL>10">
81407  ##bcftools_filterVersion=1.10.2+htslib-1.10.2-3
81408  ##bcftools_filterCommand=filter -O z -o study_filtered..vcf.gz -s LOWQUAL -i%QUAL>10 capsicumsamtools.vcf.gz; Date=Wed Jul 15 15:22:55 2020
81409  #CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  ./samples/bam/c1a.bam   ./samples/bam/c2a.bam   ./samples/bam/h1a.bam   ./samples/bam/h2a.bam   ./sa
81410  CM009386.1  165260  .   G   T   19.266  PASS    DP=6;VDB=0.000411984;SGB=1.38429;MQ0F=0;AC=4;AN=4;DP4=0,0,6,0;MQ=9  GT:PL   ./.:0,0,0   ./.:0,0,0   1/1:22,9
81411  CM009386.1  165297  .   CTT CTTT    19.2658 PASS    INDEL;IDV=3;IMF=1;DP=6;VDB=0.000411984;SGB=1.38429;MQ0F=0;AC=4;AN=4;DP4=0,0,6,0;MQ=9    GT:PL   ./.:0,0,
81412  CM009386.1  213156  .   T   C   390 PASS    DP=145;VDB=1.26117e-44;SGB=63.9545;RPB=8.16392e-10;MQB=1;BQB=6.62383e-06;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;A
81413  CM009386.1  213203  .   A   G   999 PASS    DP=143;VDB=0;SGB=8.15606;RPB=1;MQB=1;BQB=1;MQ0F=0;AC=12;AN=12;DP4=0,1,0,142;MQ=60   GT:PL   1/1:223,81,0    1/1:
81414  CM009386.1  217520  .   G   C   391 PASS    DP=166;VDB=0;SGB=58.6008;RPB=0.995314;MQB=1;BQB=0.211265;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;AN=12;DP4=109,0,5
81415  CM009386.1  286590  .   CTTTT   CTT 5.83016 LOWQUAL INDEL;IDV=1;IMF=1;DP=4;VDB=0.0058656;SGB=-0.471632;MQ0F=0;ICB=0.07;HOB=0.03125;AC=7;AN=8;DP4=1,0,3,0;MQ=
81416  CM009386.1  287149  .   C   G   354 PASS    DP=139;VDB=1.30483e-15;SGB=16.1321;RPB=0.997538;MQB=3.7147e-14;BQB=0.452973;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=
81417  CM009386.1  287150  .   T   G   394 PASS    DP=139;VDB=0;SGB=67.8441;RPB=0.989494;MQB=2.16297e-20;BQB=0.995715;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;AN=12;D
81418  CM009386.1  325424  .   T   C   30.9464 PASS    DP=8;VDB=0.02;SGB=0.940308;RPB=1;MQB=1;BQB=0.166667;MQ0F=0;ICB=0.07;HOB=0.03125;AC=1;AN=8;DP4=0,6,0,2;MQ=60
81419  CM009386.1  325483  .   A   G   371 PASS    DP=72;VDB=1.62445e-19;SGB=37.6728;RPB=0.108112;MQB=1;BQB=0.940696;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;AN=12;DP
81420  CM009386.1  381506  .   C   A   375 PASS    DP=69;VDB=1.95767e-21;SGB=36.617;RPB=0.00054252;MQB=1.60206e-11;BQB=0.106763;MQ0F=0;ICB=0.104167;HOB=0.444444;AC
81421  CM009386.1  396583  .   C   T   10.6014 PASS    DP=2;SGB=-0.045183;RPB=1;MQB=1;BQB=1;MQ0F=0;ICB=0.5;HOB=0.5;AC=2;AN=4;DP4=0,1,0,1;MQ=60 GT:PL   ./.:0,0,0
81422  CM009386.1  422715  .   A   G   7.91056 LOWQUAL DP=4;SGB=-0.045183;RPB=1;MQB=1;BQB=1;MQ0F=0;ICB=0.128205;HOB=0.0555556;AC=1;AN=6;DP4=3,0,1,0;MQ=60  GT:PL
81423  CM009386.1  422719  .   G   A   38.1784 PASS    DP=4;VDB=0.02;SGB=0.334254;RPB=1;MQB=1;BQB=0.5;MQ0F=0;ICB=0.128205;HOB=0.0555556;AC=1;AN=6;DP4=2,0,2,0;MQ=60
81424  CM009386.1  451394  .   AT  ATT 278 PASS    INDEL;IDV=9;IMF=1;DP=31;VDB=9.78562e-19;SGB=11.6757;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;AN=12;DP4=0,16,0,15;MQ
81425  CM009386.1  460048  .   G   A   182 PASS    DP=6;VDB=0.000527421;SGB=0.410784;MQ0F=0;AC=6;AN=6;DP4=0,0,6,0;MQ=60    GT:PL   1/1:92,9,0  1/1:41,3,0  ./.:0,0,
81426  CM009386.1  461446  .   T   G   44.2145 PASS    DP=2;VDB=0.02;SGB=0.334254;MQ0F=0;AC=2;AN=2;DP4=0,0,2,0;MQ=60   GT:PL   ./.:0,0,0   1/1:69,6,0  ./.:0,0,0
81427  CM009386.1  464138  .   G   A   379 PASS    DP=62;VDB=2.19388e-19;SGB=33.3255;RPB=0.0016044;MQB=1;BQB=0.997917;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;AN=12;D
81428  CM009386.1  471265  .   C   A   109 PASS    DP=5;VDB=0.0058656;SGB=0.580869;RPB=1;MQB=1;BQB=1;MQ0F=0;ICB=0.128205;HOB=0.0555556;AC=5;AN=6;DP4=1,0,4,0;MQ=60
81429  CM009386.1  490048  .   G   A   16.317  PASS    DP=1;SGB=-0.045183;MQ0F=0;AC=2;AN=2;DP4=0,0,0,1;MQ=60   GT:PL   ./.:0,0,0   ./.:0,0,0   1/1:41,3,0  ./.:0,0,
81430  CM009386.1  495510  .   G   A   7.91052 LOWQUAL DP=4;SGB=-0.045183;RPB=1;MQB=1;BQB=1;MQ0F=0;ICB=0.128205;HOB=0.0555556;AC=1;AN=6;DP4=3,0,1,0;MQ=60  GT:PL
81431  CM009386.1  508536  .   C   T   202 PASS    DP=67;VDB=6.99472e-07;SGB=-8.24061;RPB=0.996137;MQB=1;BQB=0.770885;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=4;AN=12;D
81432  CM009386.1  508602  .   A   G   35.0107 PASS    DP=9;VDB=0.02;SGB=0.940308;RPB=0;MQB=1;BQB=0.785714;MQ0F=0;ICB=0.5;HOB=0.5;AC=2;AN=4;DP4=7,0,2,0;MQ=60  GT:P
81433  CM009386.1  576326  .   G   A   731 PASS    DP=107;VDB=0;SGB=25.8539;RPB=1;MQB=1;BQB=0.81786;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=8;AN=12;DP4=0,35,0,72;MQ=60
81434  CM009386.1  576882  .   A   T   77  PASS    DP=5;VDB=0.0221621;SGB=0.129598;RPB=1;MQB=1;BQB=0.666667;MQ0F=0;ICB=0.778547;HOB=0.28125;AC=5;AN=8;DP4=0,2,0,3;M
81435  CM009386.1  579098  .   C   CTACAATG    604 PASS    INDEL;IDV=11;IMF=1;DP=52;VDB=7.80621e-41;SGB=4.00432;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=8;AN=12;DP4=27,
81436  CM009386.1  579106  .   T   A   495 PASS    DP=52;VDB=1.30483e-15;SGB=4.00432;RPB=5.05053e-09;MQB=5.05053e-09;BQB=0.124298;MQ0F=0;ICB=0.104167;HOB=0.444444;
81437  CM009386.1  579120  .   T   C   178 PASS    DP=52;VDB=6.99472e-07;SGB=3.46416;RPB=0.00735482;MQB=6.88068e-06;BQB=0.017513;MQ0F=0;ICB=0.104167;HOB=0.444444;A
81438  CM009386.1  595432  .   A   C   16.317  PASS    DP=1;SGB=-0.045183;MQ0F=0;AC=2;AN=2;DP4=0,0,1,0;MQ=60   GT:PL   ./.:0,0,0   ./.:0,0,0   ./.:0,0,0   ./.:0,0,
81439  CM009386.1  642141  .   C   T   624 PASS    DP=75;VDB=7.12062e-27;SGB=16.4087;RPB=1;MQB=1;BQB=0.226003;MQ0F=0;ICB=0.104167;HOB=0.444444;AC=8;AN=12;DP4=32,0,
81440  CM009386.1  680076  .   G   A   350 PASS    DP=11;VDB=2.26006e-08;SGB=-0.917646;MQ0F=0;AC=10;AN=10;DP4=0,0,0,11;MQ=60   GT:PL   1/1:64,6,0  1/1:122,12,0
81441  CM009386.1  680114  .   G   C   157 PASS    DP=11;VDB=0.000411984;SGB=1.38429;RPB=0.950952;MQB=0.950952;BQB=0.8618;MQ0F=0;ICB=1;HOB=0.21875;AC=5;AN=8;DP4=5,
81442  CM009386.1  680117  .   A   C   159 PASS    DP=11;VDB=0.000411984;SGB=1.38429;RPB=0.950952;MQB=0.950952;BQB=0.950952;MQ0F=0;ICB=1;HOB=0.21875;AC=5;AN=8;DP4=
81443  CM009386.1  680979  .   A   T   33.7833 PASS    DP=12;VDB=0.02;SGB=0.981224;RPB=1;MQB=1;BQB=0.9;MQ0F=0;ICB=0.07;HOB=0.03125;AC=1;AN=8;DP4=0,10,0,2;MQ=60
81444  CM009386.1  681037  .   G   A   13.3568 PASS    DP=41;VDB=0.02;SGB=0.981224;RPB=1;MQB=1;BQB=0.641026;MQ0F=0;ICB=0.3;HOB=0.125;AC=1;AN=4;DP4=39,0,2,0;MQ=60
```

Line 1, Column 1 | Tab Size: 4 | Plain Text

Type here to search

ENG ES 9:55 PM 7/16/2020

C:\Users\mauri\Documents\PosdocJaveriana\ensamblaje\capsicum\samples\vcf\R.vcf - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

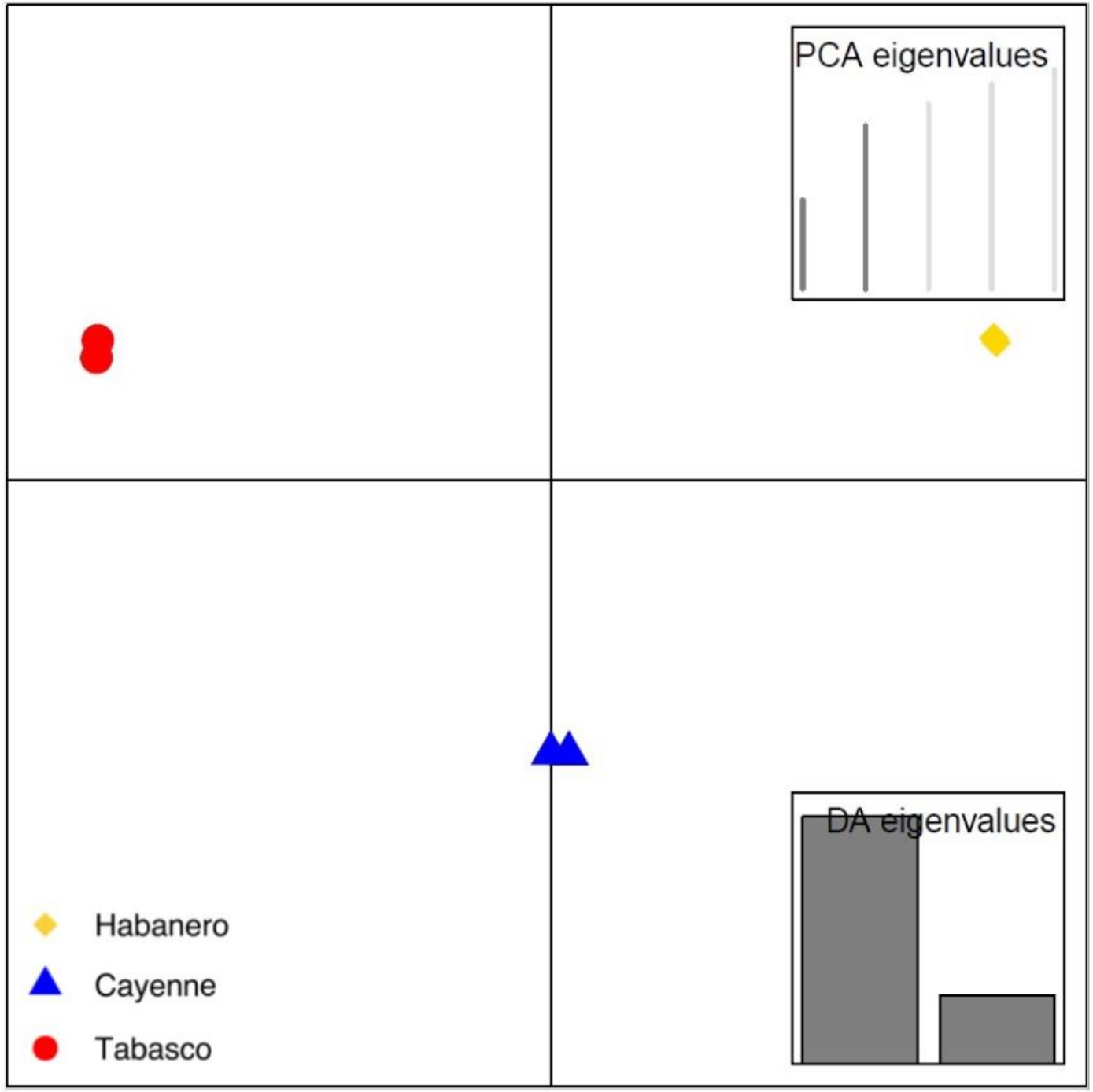
c1hpa.FASTQ.1 | R.vcf

```
#   c1  c2  h1  h2  t1  t2
    TT  TT  TT  TT  CC  CC
    GG  GG  GG  GG  CC  CC
    CC  CC  GG  GG  CC  CC
    TT  TT  TT  TT  GG  GG
    AA  AA  AA  AA  GG  GG
    CC  CC  CC  CC  AA  AA
    GG  GG  GG  GG  AA  AA
    CC  CC  TT  TT  CC  CC
    GG  GG  AA  AA  AA  AA
    TT  TT  AA  AA  AA  AA
    TT  TT  TT  TT  CC  CC
    CC  CC  TT  TT  TT  TT
    AA  AA  AA  AA  AC  CC
    GG  GG  AA  AA  GG  GG
    CC  CC  GG  GG  CC  CC
    CC  CC  TT  TT  CC  CC
    GG  GG  GG  GG  CC  CC
    CC  CC  CC  CC  GG  GG
    AA  AA  GG  GG  GG  GG
    GG  GG  AA  AA  AA  AA
    TT  TT  TA  AA  TT  TT
    AA  AA  TT  TT  AA  AA
    AA  AA  AT  AT  AA  AA
    TT  TT  TC  TC  TT  TT
    AA  AA  TT  AT  AA  AA
    AA  AA  GG  GG  GG  GG
    TC  TC  CC  CC  CC  CC
    TG  TG  GG  GG  GG  GG
    AA  AA  AA  AA  AG  GG
    TT  TT  CC  CC  TT  TT
    CC  CC  CC  CC  GG  GG
    AA  AA  CC  CC  CC  CC
    CC  CC  GG  GG  CC  CC
    AA  AA  GG  GG  GG  GG
    TT  TT  CC  CC  CC  CC
    CC  CC  TT  TT  TT  TT
    TT  TT  CC  CC  TT  TT
    GC  CC  GG  GG  GG  GG
    TT  TT  TC  CC  TT  TT
```

Line 1, Column 1 | Tab Size: 4 | Plain Text

Type here to search

ENG ES 9:56 PM 7/16/2020


PCA eigenvalues
DA eigenvalues
Habanero
Cayenne
Tabasco

Para esta semana:

- Etiquetar variables por cromosoma y posición
- Refinar el filtrado de variables por profundidad
- Organizar guía
- Descargar datos de arroz y ejecutar este procedimiento para obtener los datos necesarios para realizar los análisis de selección genómica