

1. Descripción del catálogo de genes.

Primero descargamos la tabla refGene.sql y los archivos de los 4 genomas del navegador genómico de UCSC:

```
student@ubuntuM0151: ~  
File Edit View Search Terminal Help  
ase/refGene.sql  
--2021-05-18 19:15:43-- http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database  
/refGene.sql  
Resolving hgdownload.soe.ucsc.edu (hgdownload.soe.ucsc.edu)... 128.114.119.163  
Connecting to hgdownload.soe.ucsc.edu (hgdownload.soe.ucsc.edu)|128.114.119.163|  
:80... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 1964 (1.9K) [application/sql]  
Saving to: 'refGene.sql'  
  
refGene.sql      100%[=====] 1.92K --.-KB/s   in 0s  
  
2021-05-18 19:15:44 (959 KB/s) - 'refGene.sql' saved [1964/1964]  
  
student@ubuntuM0151:~$ http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/r  
efGene.txt.gz  
bash: http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/refGene.txt.gz: No  
such file or directory  
student@ubuntuM0151:~$ wget http://hgdownload.soe.ucsc.edu/goldenPath/hg38/datab  
ase/refGene.txt.gz  
--2021-05-18 19:17:27-- http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database  
/refGene.txt.gz  
Resolving hgdownload.soe.ucsc.edu (hgdownload.soe.ucsc.edu)... 128.114.119.163  
Connecting to hgdownload.soe.ucsc.edu (hgdownload.soe.ucsc.edu)|128.114.119.163|
```

Entramos en mysql y cargamos la tabla en el database “genomas”. Comprobamos:

```
mysql> SHOW TABLES;  
+-----+  
| Tables_in_genoma |  
+-----+  
| refGene           |  
+-----+  
1 row in set (0.00 sec)  
  
mysql> DESCRIBE refGene;  
+-----+-----+-----+-----+-----+  
| Field      | Type                | Null | Key | Default |  
+-----+-----+-----+-----+-----+  
| bin        | smallint(5) unsigned | NO   |     | NULL    |  
| name       | varchar(255)         | NO   | MUL | NULL    |  
+-----+-----+-----+-----+-----+  
txEnd      | int(10) unsigned     | NO   |     | NULL    |  
cdsStart   | int(10) unsigned     | NO   |     | NULL    |  
cdsEnd     | int(10) unsigned     | NO   |     | NULL    |  
exonCount  | int(10) unsigned     | NO   |     | NULL    |  
exonStarts | longblob             | NO   |     | NULL    |  
exonEnds   | longblob             | NO   |     | NULL    |  
score      | int(11)              | YES  |     | NULL    |  
name2      | varchar(255)         | NO   | MUL | NULL    |  
cdsStartStat | enum('none','unk','incompl','cpl') | NO   |     | NULL    |  
cdsEndStat  | enum('none','unk','incompl','cpl') | NO   |     | NULL    |  
exonFrames | longblob             | NO   |     | NULL    |
```

Cargamos el genoma humano en la tabla refGene:

EINES INFORMÀTIQUES PER A LA BIOINFORMÀTICA. PAC2
Amelia Martínez Sequera.

```
mysql> LOAD DATA LOCAL INFILE 'hg38.txt' INTO TABLE refGene;
Query OK, 172767 rows affected, 10 warnings (1.94 sec)
Records: 172767 Deleted: 0 Skipped: 0 Warnings: 10

mysql> SELECT DISTINCT chrom FROM refGene WHERE chrom NOT LIKE '%\_%';
+-----+
| chrom |
+-----+
| chr1  |
| chr10 |
| chr11 |
| chr12 |
| chr13 |
| chr14 |
| chr15 |
| chr16 |
| chr17 |
| chr18 |
| chr19 |
| chr2  |
| chr20 |
| chr21 |
| chr22 |
| chr3  |
| chr4  |
| chr5  |
```

Para la levadura:

```
student@ubuntuM0151: ~
File Edit View Search Terminal Help

mysql> LOAD DATA LOCAL INFILE 'cer3.txt' INTO TABLE refGene;
Query OK, 6126 rows affected, 10 warnings (0.06 sec)
Records: 6126 Deleted: 0 Skipped: 0 Warnings: 10

mysql> SELECT name2, name, chrom, txStart, txEnd FROM refGene ORDER BY name2 LIMIT 10;
+-----+-----+-----+-----+-----+
| name2 | name          | chrom | txStart | txEnd |
+-----+-----+-----+-----+-----+
| AAC1  | NM_001182554.1 | chrXIII | 387314 | 388244 |
| AAC3  | NM_001178433.3 | chrII  | 415982 | 416906 |
| AAD10 | NM_001181813.1 | chrX   | 727404 | 728271 |
| AAD14 | NM_001183169.1 | chrXIV | 16117  | 17248  |
| AAD15 | NM_001183418.1 | chrXV  | 1646   | 2078   |
| AAD3  | NM_001178814.1 | chrIII | 313889 | 314981 |
| AAD4  | NM_001180303.1 | chrIV  | 17576  | 18566  |
| AAH1  | NM_001182979.1 | chrXIV | 359595 | 360639 |
| AAP1  | NM_001179177.1 | chrVIII | 198739 | 201310 |
| AAR2  | NM_001178314.1 | chrII  | 86721  | 87789  |
+-----+-----+-----+-----+-----+
10 rows in set (0.00 sec)

mysql> SELECT DISTINCT chrom FROM refGene WHERE chrom NOT LIKE '%\_%';
+-----+
| chrom |
+-----+
```

Para el caballo:

```
mysql> TRUNCATE TABLE refGene;
Query OK, 0 rows affected (0.00 sec)

mysql> LOAD DATA LOCAL INFILE 'cab3.txt' INTO TABLE refGene;
Query OK, 76581 rows affected, 10 warnings (0.88 sec)
Records: 76581 Deleted: 0 Skipped: 0 Warnings: 10

mysql> SELECT name2, name, chrom, txStart, txEnd FROM refGene ORDER BY name2 LIMIT 10;
+-----+-----+-----+-----+-----+
| name2 | name          | chrom | txStart | txEnd |
+-----+-----+-----+-----+-----+
| A1BG  | XM_023651550.1 | chr10 | 28405572 | 28409265 |
| A1CF  | XM_014733382.2 | chr1  | 41813428 | 41889961 |
| A2ML1 | XM_023643099.1 | chr6  | 36959093 | 37019643 |
| A2ML1 | XM_014740589.2 | chr6  | 36982882 | 37019643 |
| A2ML1 | XM_014740588.2 | chr6  | 36982882 | 37019643 |
| A3F2  | XM_023631192.1 | chr28 | 37093256 | 37097738 |
| A3GALT2 | XM_023634015.1 | chr2  | 23764126 | 23778484 |
| A4GALT | XM_023631773.1 | chr28 | 40288953 | 40315193 |
| A4GNT | XM_023619735.1 | chr16 | 74432232 | 74442579 |
| A4GNT | NM_001163877.1 | chr16 | 74434598 | 74441156 |
+-----+-----+-----+-----+-----+
10 rows in set (0.00 sec)
```

1. Número de cromosomas distintos.

```
% SELECT DISTINCT chrom FROM refGene WHERE chrom NOT LIKE '%\_%';
```

Hay que restarle 1 de “chrom”.

Para el caballo: 33

```
student@ubuntu10131:~$  
File Edit View Search Termin  
chr20  
chr21  
chr22  
chr23  
chr24  
chr25  
chr26  
chr27  
chr28  
chr29  
chr3  
chr30  
chr31  
chr4  
chr5  
chr6  
chr7  
chr8  
chr9  
chrM  
chrom  
chrX  
+-----+  
34 rows in set (0.01 sec)
```

Para el pollo: 34

```
File Edit View Search Termin  
chr2  
chr20  
chr21  
chr22  
chr23  
chr24  
chr25  
chr26  
chr27  
chr28  
chr3  
chr30  
chr31  
chr32  
chr33  
chr4  
chr5  
chr6  
chr7  
chr8  
chr9  
chrom  
chrW  
chrZ  
+-----+  
35 rows in set (0.01 sec)
```

2. Número de genes distintos.

```
% SELECT COUNT(DISTINCT name2) FROM refGene;
```

Para la levadura:

```
mysql> SELECT COUNT(DISTINCT name2) FROM refGene;  
+-----+  
| COUNT(DISTINCT name2) |  
+-----+  
|          6126         |  
+-----+  
1 row in set (0.00 sec)
```

Caballo:

```
mysql> SELECT COUNT(DISTINCT name2) FROM refGene;
+-----+
| COUNT(DISTINCT name2) |
+-----+
|           29517 |
+-----+
1 row in set (0.02 sec)
```

3. Número de transcritos distintos.

% SELECT COUNT(DISTINCT name) FROM refGene;

Levadura:

```
mysql> SELECT COUNT(DISTINCT name) FROM refGene;
+-----+
| COUNT(DISTINCT name) |
+-----+
|           6126 |
+-----+
1 row in set (0.00 sec)
```

Caballo:

```
mysql> SELECT COUNT(DISTINCT name) FROM refGene;
+-----+
| COUNT(DISTINCT name) |
+-----+
|          76581 |
+-----+
1 row in set (0.03 sec)
```

4. Número de transcritos codificantes distintos. Para la levadura:

% SELECT COUNT(DISTINCT name) FROM refGene where name LIKE '%NM%';

```
mysql> SELECT COUNT(DISTINCT name) FROM refGene where name LIKE '%NM%';
+-----+
| COUNT(DISTINCT name) |
+-----+
|           5983 |
+-----+
1 row in set (0.01 sec)
```

Caballo:

```
mysql> SELECT COUNT(DISTINCT name) FROM refGene where name LIKE '%NM%';
+-----+
| COUNT(DISTINCT name) |
+-----+
|           1114 |
+-----+
1 row in set (0.02 sec)
```

5. Número de transcritos no codificantes distintos. Para el caballo:

% SELECT COUNT(DISTINCT name) FROM refGene where name LIKE '%NR%';

```
mysql> SELECT COUNT(DISTINCT name) FROM refGene where name LIKE '%NR%';
+-----+
| COUNT(DISTINCT name) |
+-----+
|           689 |
+-----+
1 row in set (0.03 sec)
```

```
mysql> SELECT COUNT(DISTINCT name) FROM refGene where name LIKE '%NR%';
+-----+
| COUNT(DISTINCT name) |
+-----+
|          18297      |
+-----+
1 row in set (0.06 sec)
```

6. Número de transcritos por cada gen (en promedio)

% SELECT name2, count(name) AS totales FROM refGene GROUP BY name2 LIMIT 10;

Para el caballo:

Primero creamos “totales”:

```
mysql> SELECT name2, COUNT(name) AS totales FROM refGene GROUP BY name2 LIMIT 10
+-----+
| name2 | totales |
+-----+
| A1BG   | 1       |
| A1CF   | 1       |
| A2ML1  | 3       |
| A3F2   | 1       |
| A3GALT2| 1       |
| A4GALT | 1       |
| A4GNT  | 2       |
| AAAS   | 3       |
| AACS   | 2       |
| AADACL2| 1       |
+-----+
```

Y calculamos la media:

% SELECT AVG(tmp.totales) from (SELECT name2, COUNT(name) AS totales FROM refGene GROUP BY name2) tmp;

```
File Edit View Search Terminal Help
mysql> SELECT AVG(tmp.totales) from (SELECT name2, COUNT(name) AS totales FROM refGene GROUP BY name2) tmp;
+-----+
| AVG(tmp.totales) |
+-----+
|          2.5945   |
+-----+
1 row in set (0.12 sec)
```

Para el pollo:

```
mysql> SELECT name2, COUNT(name) AS totales FROM refGene GROUP BY name2 LIMIT 10
+-----+
| name2 | totales |
+-----+
| 2KTGRL | 9       |
| 44G24.1 | 1      |
| A1CF   | 3       |
| A2M    | 2       |
| A2ML1  | 1       |
| A2ML2  | 6       |
| A2ML3  | 1       |
| A2ML4  | 1       |
| A4GALT | 2       |
| A4GNT  | 1       |
+-----+
10 rows in set (0.01 sec)
```

```
efGene GROUP BY name2) tmp;
+-----+
| AVG(tmp.totales) |
+-----+
|          2.6203   |
+-----+
1 row in set (0.09 sec)
```

Humano:

```
mysql> SELECT name2, COUNT(name) AS totales FROM refGene GROUP BY name2 LIMIT 10
+-----+-----+
| name2 | totales |
+-----+-----+
| A1BG  | 1       |
| A1BG-AS1 | 1     |
| A1CF  | 12      |
| A2M   | 5       |
| A2M-AS1 | 3     |
| A2ML1 | 9       |
| A2MP1 | 1       |
| A3GALT2 | 1     |
| A4GALT | 11      |
| A4GNT | 3       |
+-----+-----+
10 rows in set (0.00 sec)
```



```
mysql> SELECT AVG(tmp.totales) from (SELECT name2, COUNT(name) AS totales FROM r
efGene GROUP BY name2) tmp;
+-----+
| AVG(tmp.totales) |
+-----+
| 4.4814           |
+-----+
1 row in set (0.20 sec)
```

7. Número de exones por transcrito (en promedio)

% SELECT AVG(exonCount) FROM refGene;

Para la levadura:

```
mysql> SELECT AVG(exonCount) FROM refGene;
+-----+
| AVG(exonCount) |
+-----+
| 1.0583         |
+-----+
1 row in set (0.00 sec)
```

Para el caballo:

```
mysql> SELECT AVG(exonCount) FROM refGene;
+-----+
| AVG(exonCount) |
+-----+
| 12.2290        |
+-----+
1 row in set (0.02 sec)
```

Para el pollo:

```
mysql> SELECT AVG(exonCount) FROM refGene;
+-----+
| AVG(exonCount) |
+-----+
| 12.8119        |
+-----+
1 row in set (0.02 sec)
```

8. Número de nucleótidos por transcrito (en promedio)

% SELECT AVG(txEnd – txStart +1) FROM refGene;

Para la levadura:

```
mysql> SELECT AVG(txEnd-txStart+1) FROM refGene;
+-----+
| AVG(txEnd-txStart+1) |
+-----+
| 1467.6996           |
+-----+
1 row in set (0.00 sec)
```

Para el caballo:

```
mysql> SELECT AVG(txEnd-txStart+1) FROM refGene;
+-----+
| AVG(txEnd-txStart+1) |
+-----+
|          66577.4442 |
+-----+
1 row in set (0.03 sec)
```

Para el pollo:

```
mysql> SELECT AVG(txEnd-txStart+1) FROM refGene;
+-----+
| AVG(txEnd-txStart+1) |
+-----+
|          47617.6525 |
+-----+
1 row in set (0.02 sec)

mysql> TRUNCATE TABLE refGene;
Query OK, 0 rows affected (0.00 sec)
```

Genoma	1	2	3	4	5	6	7	8
<i>H. sapiens</i> (hg38)	24	38552	163989	60847	18297	4.48	11.6	76709.7
<i>E. caballus</i> (EcuCab3.0)	33	29517	76581	1114	689	2.6	12.23	66576.4
<i>G. gallus</i> (GRCg6a)	34	23727	62161	6401	776	2.62	12.8	47616.7
<i>S. cerevisiae</i> (sacCer3)	17	6126	6126	5983	123	1	1.06	1466.7

2.Datos de expresión.

Visualizamos los datos de " **GTEX_Analysis_v8_Annotations_SubjectPhenotypesDS.txt**", que se han descargado con el nombre de Phenotypes.txt. Se elimina la primera línea con grep -v SUBJID, y lo llamamos Pheno.txt.

```
student@ubuntuM0151:~$ head Phenotypes.txt
SUBJID SEX AGE DTHRDY
GTEX-1117F 2 60-69 4
GTEX-111CU 1 50-59 0
GTEX-111FC 1 60-69 1
GTEX-111VG 1 60-69 3
GTEX-111YS 1 60-69 0
GTEX-11220 2 60-69 0
GTEX-1128S 2 60-69 2
GTEX-113IC 1 60-69
GTEX-113JC 2 50-59 2
student@ubuntuM0151:~$ grep -v SUBJID Phenotypes.txt > Pheno.txt
student@ubuntuM0151:~$ head Pheno.txt
GTEX-1117F 2 60-69 4
GTEX-111CU 1 50-59 0
GTEX-111FC 1 60-69 1
GTEX-111VG 1 60-69 3
GTEX-111YS 1 60-69 0
GTEX-11220 2 60-69 0
GTEX-1128S 2 60-69 2
GTEX-113IC 1 60-69
GTEX-113JC 2 50-59 2
GTEX-117XS 1 60-69 2
student@ubuntuM0151:~$
```

Del archivo de las muestras, " **GTEx_Analysis_v8_Annotations_SampleAttributesDS.txt** ", seleccionamos las columnas necesarias.

```
student@ubuntuM0151:~  
File Edit View Search Terminal Help  
student@ubuntuM0151:~$ grep -v SAMPID SampleAtt.txt | gawk 'BEGIN {FS=OFS="\t"}{  
print $1, $7,$55,$6,$48}'> Sample.txt  
student@ubuntuM0151:~$ head -11 Sample.txt  
GTEX-1117F-0003-SM-58Q7G Whole Blood Blood  
GTEX-1117F-0003-SM-5DWSB Whole Blood Blood  
GTEX-1117F-0003-SM-6WBT7 Whole Blood Blood  
GTEX-1117F-0011-R10a-SM-AHZ7F Brain - Frontal Cortex (BA9) Brain  
GTEX-1117F-0011-R10b-SM-CYKQ8 Brain - Frontal Cortex (BA9) Brain  
GTEX-1117F-0226-SM-5GZZ7 Adipose - Subcutaneous 1.19993e+07 Adipose  
Tissue 6.68332e+07  
GTEX-1117F-0426-SM-5EGHI Muscle - Skeletal 1.15502e+07 Muscle 5  
.99636e+07  
GTEX-1117F-0526-SM-5EGHJ Artery - Tibial 1.10003e+07 Blood Vessel 5  
.98394e+07  
GTEX-1117F-0626-SM-5N9CS Artery - Coronary 1.16179e+07 Blood Ve  
ssel 6.98592e+07  
GTEX-1117F-0726-SM-5GIEN Heart - Atrial Appendage 9.26607e+06 H  
eart 6.28857e+07  
GTEX-1117F-1326-SM-5EGHH Adipose - Visceral (Omentum) 9.63106e+06 A  
dipose Tissue 5.97958e+07  
student@ubuntuM0151:~$
```

Hay dos columnas separadas por un tabulador. Llamamos al nuevo archivo Sample.txt.

Creamos un database "gtex" y creamos la tabla "pheno", en la que cargaremos los datos de Pheno.txt. Como ha valores ausentes, para evitar que aparezca como un 0 escribimos:

**% LOAD DATA LOCAL INFILE 'Pheno.txt' INTO TABLE pheno(subid, sex, age, @vdthrdy) SET
dthrdy = NULLIF(@vdthrdy, '');**

```
File Edit View Search Terminal Help  
Your MySQL connection id is 60  
Server version: 5.7.33-0ubuntu0.16.04.1 (Ubuntu)  
Copyright (c) 2000, 2021, Oracle and/or its affiliates.  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
mysql> USE gtex;  
Reading table information for completion of table and column names  
You can turn off this feature to get a quicker startup with -A  
Database changed  
mysql> CREATE TABLE pheno  
-> (subid VARCHAR (20) NOT NULL,  
-> sex INT,  
-> age TEXT,  
-> dthrdy INT,  
-> PRIMARY KEY (subid));  
Query OK, 0 rows affected (0.05 sec)
```

```
File Edit View Search Terminal Help  
Database changed  
mysql> SHOW TABLES;  
+-----+  
| Tables_in_gtex |  
+-----+  
| fenotipos |  
| pheno |  
| samples |  
+-----+  
3 rows in set (0.02 sec)  
mysql> DESCRIBE pheno;  
+-----+-----+-----+-----+-----+  
| Field | Type | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+  
| subid | varchar(20) | NO | PRI | NULL | |  
| sex | int(11) | YES | | NULL | |  
| age | text | YES | | NULL | |  
| dthrdy | int(11) | YES | | NULL | |  
+-----+-----+-----+-----+-----+  
4 rows in set (0.01 sec)  
mysql> LOAD DATA LOCAL INFILE 'Pheno.txt' INTO TABLE pheno (subid, sex, age, @vd  
thrdy) SET dthrdy = NULLIF(@vdthrdy, '');
```


Comprobamos:

```
mysql> LOAD DATA LOCAL INFILE 'Pheno.txt' INTO TABLE pheno (subid, sex, age, @vdthhrdy) SET dthhrdy = NULLIF(@vdthhrdy, '');
Query OK, 980 rows affected (0.04 sec)
Records: 980 Deleted: 0 Skipped: 0 Warnings: 0

mysql> SELECT subid, sex, age, dthhrdy FROM pheno LIMIT 10;
```

subid	sex	age	dthhrdy
GTEX-1117F	2	60-69	4
GTEX-111CU	1	50-59	0
GTEX-111FC	1	60-69	1
GTEX-111VG	1	60-69	3
GTEX-111YS	1	60-69	0
GTEX-11220	2	60-69	0
GTEX-11285	2	60-69	2
GTEX-113IC	1	60-69	NULL
GTEX-113JC	2	50-59	2
GTEX-117XS	1	60-69	2

1. Rango de edad más frecuente del que se tienen muestras.

Contamos y ordenamos la columna age de forma descendiente.

```
% SELECT age, COUNT(age) FROM pheno GROUP BY age ORDER BY COUNT(age) DESC;
```

La franja de 60-69 es la que tiene más muestras.

```
student@ubuntuM0151: ~
File Edit View Search Terminal Help
mysql> SELECT age, COUNT(age) FROM pheno GROUP BY age ORDER BY COUNT(age) DESC;
```

age	COUNT(age)
60-69	317
50-59	315
40-49	153
20-29	84
30-39	78
70-79	33

```
6 rows in set (0.02 sec)

mysql> SELECT dthhrdy, COUNT(dthhrdy) FROM pheno WHERE sex='2' GROUP BY dthhrdy
ORDER BY COUNT(dthhrdy) DESC;
```

dthhrdy	COUNT(dthhrdy)
0	194
2	50
4	46
3	20
1	12
NULL	0

2. Tipo de muerte más frecuente por género.

El tipo de muerte es la columna dthhrdy. Se utiliza sex=1 o 2 para distinguir entre los dos géneros:

```
% SELECT dthhrdy , COUNT(dthhrdy) FROM pheno WHERE sex='1' GROUP BY dthhrdy ORDER
BY COUNT(dthhrdy) DESC;
```

En ambos géneros la causa más frecuente es la '0', Ventilator Case.

```
student@ubuntuM0151: ~  
File Edit View Search Terminal Help  
+-----+  
| dthhrdy | COUNT(dthhrdy) |  
+-----+  
| 0 | 194 |  
| 2 | 50 |  
| 4 | 46 |  
| 3 | 20 |  
| 1 | 12 |  
| NULL | 0 |  
+-----+  
6 rows in set (0.01 sec)  
  
mysql> SELECT dthhrdy, COUNT(dthhrdy) FROM pheno WHERE sex='1' GROUP BY dthhrdy  
ORDER BY COUNT(dthhrdy) DESC;  
+-----+  
| dthhrdy | COUNT(dthhrdy) |  
+-----+  
| 0 | 317 |  
| 2 | 189 |  
| 4 | 73 |  
| 3 | 37 |  
| 1 | 23 |  
| NULL | 0 |  
+-----+
```

3. ¿Cuántos tipos de regiones del cerebro (brain) podemos encontrar?

Los tipos de regiones se encuentran en la columna regions, y contienen la palabra 'Brain'.

Creamos la tabla 'sample' y cargamos 'Sample.txt', teniendo en cuenta que hay datos ausentes:

```
mysql> CREATE TABLE sample(sampleid VARCHAR (50) NOT NULL,  
-> regions VARCHAR(200) NOT NULL,  
-> splitreads DECIMAL (10,0),  
-> tipomuestra VARCHAR(100) NOT NULL,  
-> mapeunique DECIMAL(10,0),  
-> PRIMARY KEY (sampleid));  
Query OK, 0 rows affected (0.03 sec)  
  
mysql> LOAD DATA LOCAL INFILE 'Sample.txt' INTO TABLE sample(sampleid, regions,  
@vsplitreads, tipomuestra, @vmappedunique) SET splitreads=NULLIF(@vsplitreads, '  
' ),mapeunique = NULLIF(@vmappedunique, '');  
Query OK, 22951 rows affected (0.28 sec)
```

% SELECT DISTINCT(regions) FROM sample WHERE regions LIKE '%Brain%' ORDER BY regions;

```
mysql> Sysql> SELECT DISTINCT(regions) FROM sample WHERE regions LIKE '%Brain%' ORDER BY regions;  
ERROR 1305 (42000): FUNCTION gtex.DISTINCTDISTINCT does not exist  
mysql> SELECT DISTINCT(regions) FROM sample WHERE regions LIKE '%Brain%' ORDER BY regions;  
+-----+  
| regions |  
+-----+  
| Brain - Amygdala |  
| Brain - Anterior cingulate cortex (BA24) |  
| Brain - Caudate (basal ganglia) |  
| Brain - Cerebellar Hemisphere |  
| Brain - Cerebellum |  
| Brain - Cortex |  
| Brain - Frontal Cortex (BA9) |  
| Brain - Hippocampus |  
| Brain - Hypothalamus |  
| Brain - Nucleus accumbens (basal ganglia) |  
| Brain - Putamen (basal ganglia) |  
| Brain - Spinal cord (cervical c-1) |  
| Brain - Substantia nigra |  
+-----+  
13 rows in set (0.01 sec)
```

Existen 13 tipos.

4. Paciente que más muestras tiene:

Primero se crea una nueva columna 'subid' en la tabla sample:

% ALTER TABLE sample ADD subid VARCHAR(20) NOT NULL;

Esta columna nos permite relacionar sample con pheno.

% ALTER TABLE sample ADD FOREIGN KEY (subid) REFERENCES pheno(subid);

Se rellenan los datos teniendo en cuenta que hay valores ausentes:

```
mysql> UPDATE sample
  -> SET subid= IF(sampleid IS NULL, 0, substring_index(sampleid, '-',2));
Query OK, 22951 rows affected (0.41 sec)
Rows matched: 22951  Changed: 22951  Warnings: 0
```

Comprobamos:

```
File Edit View Search Terminal Help
-----+-----
10 rows in set (0.01 sec)

mysql> SELECT * FROM sample LIMIT 10;
-----+-----+-----+-----+-----+-----+
| sampleid | regions | splitreads | tipomuestra | mappeuniqu | subid |
-----+-----+-----+-----+-----+-----+
| GTEX-1117F-0003-SM-58Q7G | Whole Blood | NULL | Blood | NULL | GTEX-1117F |
| GTEX-1117F-0003-SM-5DMSB | Whole Blood | NULL | Blood | NULL | GTEX-1117F |
| GTEX-1117F-0003-SM-6WBT7 | Whole Blood | NULL | Blood | NULL | GTEX-1117F |
| GTEX-1117F-0011-R10a-SM-AHZ7F | Brain - Frontal Cortex (BA9) | NULL | Brain | NULL | GTEX-1117F |
| GTEX-1117F-0011-R10b-SM-CYKQ8 | Brain - Frontal Cortex (BA9) | NULL | Brain | NULL | GTEX-1117F |
| GTEX-1117F-0226-SM-5GZ27 | Adipose - Subcutaneous | 11999300 | Adipose Tissue | 66833200 | GTEX-1117F |
| GTEX-1117F-0426-SM-5EGHI | Muscle - Skeletal | 11550200 | Muscle | 59963600 | GTEX-1117F |
| GTEX-1117F-0526-SM-5EGHJ | Artery - Tibial | 11000300 | Blood Vessel | 59839400 | GTEX-1117F |
| GTEX-1117F-0626-SM-5N9CS | Artery - Coronary | 11617900 | Blood Vessel | 69859200 | GTEX-1117F |
| GTEX-1117F-0726-SM-5GIEN | Heart - Atrial Appendage | 9266070 | Heart | 62885700 | GTEX-1117F |
-----+-----+-----+-----+-----+-----+
10 rows in set (0.01 sec)
```

Pacientes con más muestras:

% SELECT subid, COUNT(subid) FROM sample GROUP BY subid ORDER BY COUNT(subid) DESC LIMIT 10;

```
>
mysql> SELECT subid, COUNT(subid) FROM sample GROUP BY subid ORDER BY count(subid) DESC LIMIT 10;
-----+-----+
| subid | COUNT(subid) |
-----+-----+
| K-562 | 217 |
| GTEX-NPJ8 | 72 |
| GTEX-RU72 | 59 |
| GTEX-Q2AG | 58 |
| GTEX-N7MS | 56 |
| GTEX-YEC3 | 51 |
| GTEX-N7MT | 50 |
| GTEX-RNOR | 50 |
| GTEX-OHPN | 49 |
| GTEX-T5JC | 49 |
-----+-----+
0 rows in set (0.02 sec)
```

El K-562 es el que tiene más: 217.

5. Cuáles son las 7 muestras con más Split Reads.

% SELECT subid, sampleid, splitreads FROM sample ORDER BY splitreads DESC LIMIT 7;

```
mysql> SELECT subid, sampleid, splitreads FROM sample ORDER BY splitreads DESC LIMIT 10;
-----+-----+-----+
| subid | sampleid | splitreads |
-----+-----+-----+
| GTEX-14BMU | GTEX-14BMU-1526-SM-5TDE6 | 99228600 |
| GTEX-14JFF | GTEX-14JFF-0526-SM-62LFL | 81794300 |
| GTEX-QMRM | GTEX-QMRM-0426-SM-4R1K2 | 80654100 |
| GTEX-1JMCI | GTEX-1JMCI-2026-SM-CMKGP | 68568100 |
| GTEX-1JJE9 | GTEX-1JJE9-0006-SM-CQOE8 | 64031800 |
| GTEX-13G51 | GTEX-13G51-0011-R8b-SM-5LZZ4 | 60748600 |
| GTEX-18D9B | GTEX-18D9B-0008-SM-EAZBT | 57394000 |
| GTEX-11DJU | GTEX-11DJU-0326-SM-CGQEE | 57307700 |
| GTEX-UJMC | GTEX-UJMC-1026-SM-4IHKH | 56677200 |
| GTEX-1LKK1 | GTEX-1LKK1-0005-SM-CXZIT | 56454200 |
-----+-----+
10 rows in set (0.01 sec)
```

6. Mediante el comando “join” responder a, ¿cuántas mujeres han muerto de manera violenta y tienen muestras de sangre? ¿y cuál es la media de “ mapped unique” de esta selección?

Subid es el elemento común entre pheno y sample: pheno.subid=sample.subid para hacer el JOIN.

Las otras variables a tener en cuenta son: sex=2 (mujer), dthhrdy=1(muerte violenta), tipomuestra que contenga “Blood”.

```
% SELECT DISTINCT(pheno.subid), pheno.sex, pheno.dthhrdy, sample.subid,  
sample.tipomuestra FROM pheno JOIN sample on pheno.subid=sample.subid WHERE sex=2  
AND dthhrdy=1 AND tpomuestra LIKE '%Blood%' AND tipomuestra NOT LIKE '%Vessel%';
```

```
mysql> SELECT DISTINCT(pheno.subid), pheno.sex, pheno.dthhrdy, sample.subid, sample.tipomuestra FROM ph  
eno JOIN sample ON pheno.subid=sample.subid WHERE sex=2 AND dthhrdy=1 AND tipomuestra LIKE '%Blood%'AND  
tipomuestra NOT LIKE '%Vessel%';  
+-----+-----+-----+-----+-----+  
| subid | sex | dthhrdy | subid | tipomuestra |  
+-----+-----+-----+-----+-----+  
| GTEX-12ZZX | 2 | 1 | GTEX-12ZZX | Blood |  
| GTEX-15ER7 | 2 | 1 | GTEX-15ER7 | Blood |  
| GTEX-1F7RK | 2 | 1 | GTEX-1F7RK | Blood |  
| GTEX-1GZHY | 2 | 1 | GTEX-1GZHY | Blood |  
| GTEX-1PBJI | 2 | 1 | GTEX-1PBJI | Blood |  
| GTEX-1POEN | 2 | 1 | GTEX-1POEN | Blood |  
| GTEX-RNOR | 2 | 1 | GTEX-RNOR | Blood |  
| GTEX-T2IS | 2 | 1 | GTEX-T2IS | Blood |  
| GTEX-WRHU | 2 | 1 | GTEX-WRHU | Blood |  
| GTEX-YFC4 | 2 | 1 | GTEX-YFC4 | Blood |  
| GTEX-ZVZQ | 2 | 1 | GTEX-ZVZQ | Blood |  
+-----+-----+-----+-----+-----+  
11 rows in set (0.01 sec)
```

11 mujeres.

Y hacemos la media de mapped unique:

```
% SELECT AVG(mappeunique) FROM pheno JOIN sample ON pheno.subid=sample.subid  
WHERE sex = 2 AND dthhrdy = 1 AND tipomuestra LIKE '%Blood%' AND tipomuestra NOT LIKE  
'%Vessel%';
```

```
mysql> SELECT AVG(mappeunique) FROM pheno JOIN sample ON pheno.subid=sample.subid WHERE sex=2 AND dthhr  
dy=1 AND tipomuestra LIKE '%Blood%'AND tipomuestra NOT LIKE '%Vessel%';  
+-----+  
| AVG(mappeunique) |  
+-----+  
| 92580733.3333 |  
+-----+  
1 row in set (0.02 sec)
```

3. Servicio web de consulta de genes.

En el database “genoma” se crean 4 tablas, una para cada una de las especies.

EINES INFORMÀTIQUES PER A LA BIOINFORMÀTICA. PAC2

Amelia Martínez Sequera.

```
student@ubuntuM0151:~$  
File Edit View Search Terminal Help  
+-----+  
| Tables_in_genoma |  
+-----+  
| refGene          |  
| refGeneCer       |  
+-----+  
2 rows in set (0.00 sec)  
  
mysql> CREATE TABLE refGeneCab LIKE refGene;  
Query OK, 0 rows affected (0.01 sec)  
  
mysql> CREATE TABLE refGeneGal LIKE refGene;  
Query OK, 0 rows affected (0.01 sec)  
  
mysql> LOAD DATA LOCAL INFILE 'cer3.txt' INTO TABLE refGeneCer;  
Query OK, 6126 rows affected, 10 warnings (0.05 sec)  
Records: 6126 Deleted: 0 Skipped: 0 Warnings: 10  
  
mysql> LOAD DATA LOCAL INFILE 'cab3.txt' INTO TABLE refGeneCab;  
Query OK, 76581 rows affected, 10 warnings (0.93 sec)  
Records: 76581 Deleted: 0 Skipped: 0 Warnings: 10  
  
mysql> LOAD DATA LOCAL INFILE 'g6a.txt' INTO TABLE refGeneGal;  
Query OK, 62171 rows affected, 10 warnings (0.72 sec)  
Records: 62171 Deleted: 0 Skipped: 0 Warnings: 10  
  
mysql>
```

Código HTML: (html2.html)

```
Applications Places System  
emac@ubuntuM0151  
File Edit Options Buffers Tools HTML SGML Text Help  
Save Undo  
  
<html>  
  <head>  
    <title>Human gene catalog</title>  
    <link rel="stylesheet" href="css.css">  
  </head>  
  <body>  
    <h1>Reference sequence and Gene Ontology catalog</h1>  
    <p>  
      <b>1.Search for a especie:</b>  
      <form action=php3.php method=post>  
        <select name=organism>  
          <option value=1>Human</option>  
          <option value=2>Horse</option>  
          <option value=3>Chicken</option>  
          <option value=4>Sacharomyces</option>  
        </select>  
        <br><br>  
        <b>2.Select a gene:</b>  
        <br>  
        <input type=text name=gene>  
        <br><br>  
        <input type=submit name=submit value=Submit>  
      </form>  
    </p>  
  </body>  
</html>  
  
---- html2.html All L1 (HTML)  
For information about GNU Emacs and the GNU system, type C-h C-a.  
student@ubuntuM0151 [Human gene catalog ... EINES.PEC2.docx - Mi... emac@ubuntuM0151
```

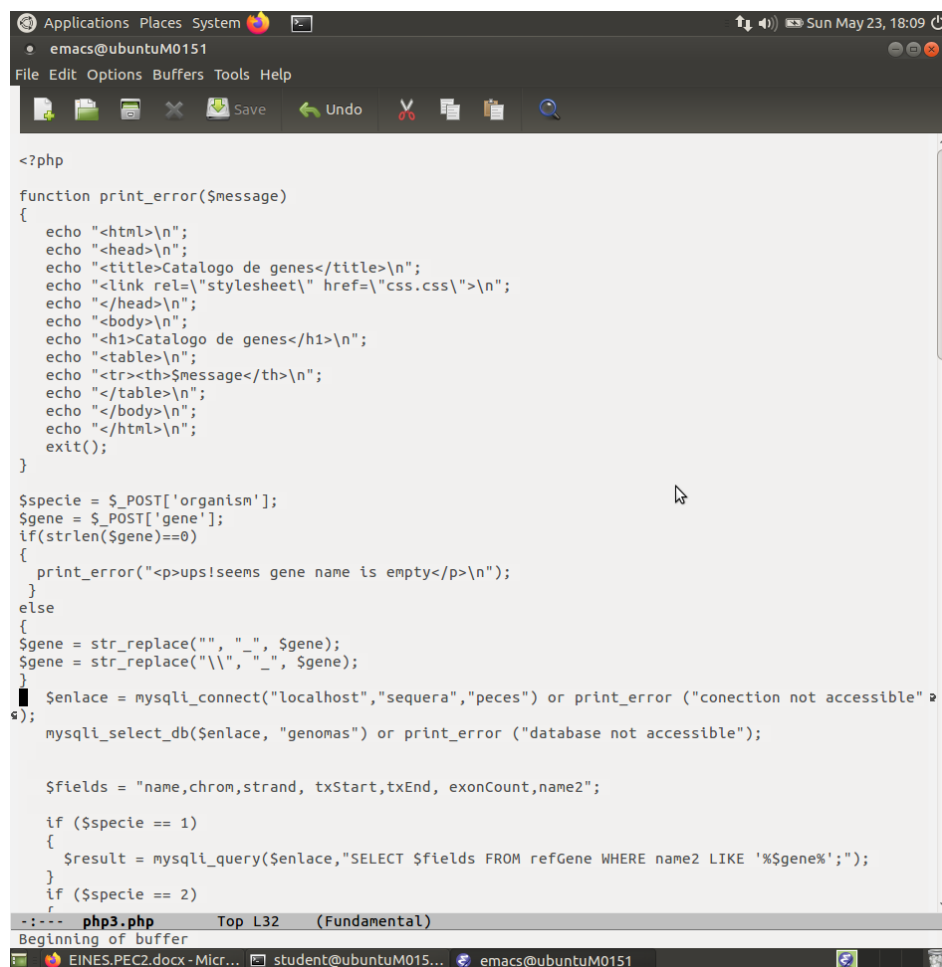
PHP: (php3.php)

Primero establecemos la función `print_error`. Luego las 2 variables a introducir (form `method=post`), una tipo `<select>` y la otra tipo `<input type=text>`.

Con `$enlace` nos conectamos a `mysql`. Luego seleccionamos database con `mysqli_select_db`.

Antes de hacer la query, establecemos las columnas a consultar con `$fields`, y establecemos un condicional en función de la especie seleccionada. Se muestra el resultado mediante `mysqli_affected_rows` en una tabla html, además de un link para UCSC, también en función de la especie seleccionada.

Se incluyen, además, mensajes de error para campos en blanco o genes que no existen.



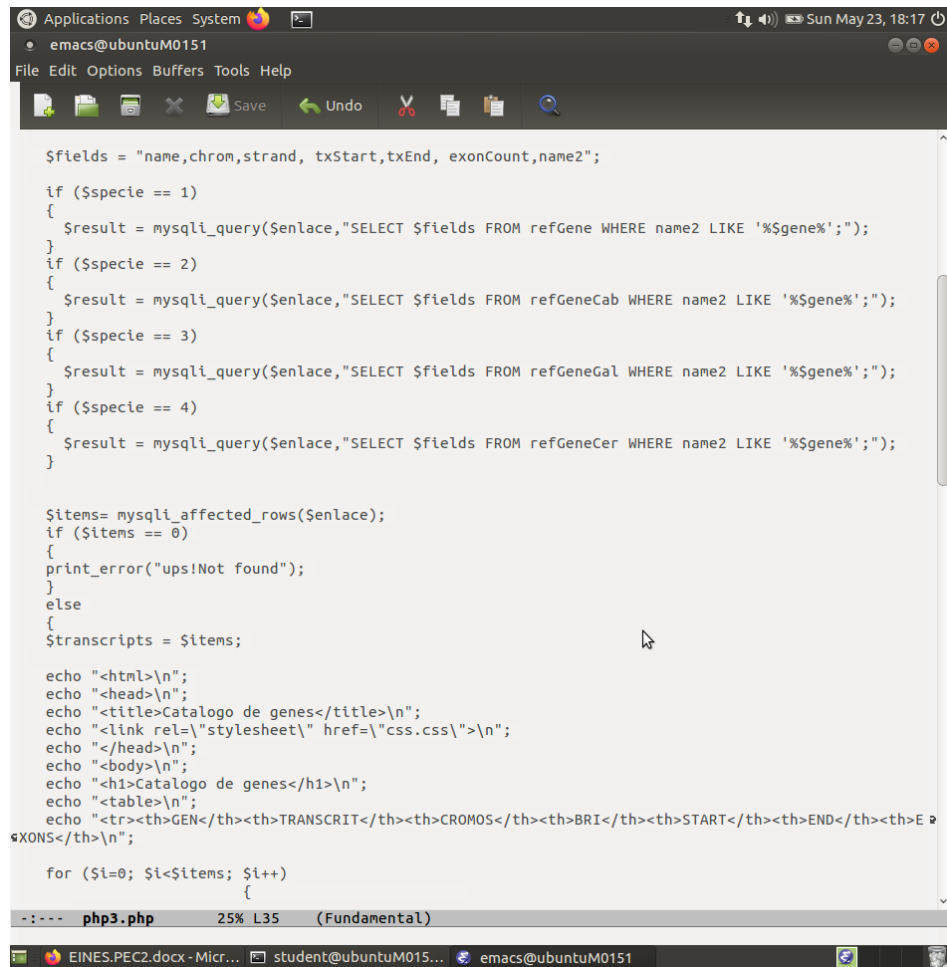
```
<?php

function print_error($message)
{
    echo "<html>\n";
    echo "<head>\n";
    echo "<title>Catalogo de genes</title>\n";
    echo "<link rel='stylesheet' href='css.css'>\n";
    echo "</head>\n";
    echo "<body>\n";
    echo "<h1>Catalogo de genes</h1>\n";
    echo "<table>\n";
    echo "<tr><th>$message</th>\n";
    echo "</table>\n";
    echo "</body>\n";
    echo "</html>\n";
    exit();
}

$specie = $_POST['organism'];
$gene = $_POST['gene'];
if(strlen($gene)==0)
{
    print_error("<p>ups!seems gene name is empty</p>\n");
}
else
{
    $gene = str_replace(" ", "_", $gene);
    $gene = str_replace("\\", "\\", $gene);
}
$enlace = mysqli_connect("localhost","sequera","peces") or print_error ("conection not accessible");
mysqli_select_db($enlace, "genomas") or print_error ("database not accessible");

$fields = "name,chrom,strand, txStart,txEnd, exonCount,name2";

if ($specie == 1)
{
    $result = mysqli_query($enlace,"SELECT $fields FROM refGene WHERE name2 LIKE '%$gene%'");
}
if ($specie == 2)
{
```



```
$fields = "name,chrom,strand, txStart,txEnd, exonCount,name2";

if ($specie == 1)
{
    $result = mysqli_query($enlace,"SELECT $fields FROM refGene WHERE name2 LIKE '%$gene%'");
}
if ($specie == 2)
{
    $result = mysqli_query($enlace,"SELECT $fields FROM refGeneCab WHERE name2 LIKE '%$gene%'");
}
if ($specie == 3)
{
    $result = mysqli_query($enlace,"SELECT $fields FROM refGeneGal WHERE name2 LIKE '%$gene%'");
}
if ($specie == 4)
{
    $result = mysqli_query($enlace,"SELECT $fields FROM refGeneCer WHERE name2 LIKE '%$gene%'");
}

$items= mysqli_affected_rows($enlace);
if ($items == 0)
{
    print_error("ups!Not found");
}
else
{
    $transcripts = $items;

    echo "<html>\n";
    echo "<head>\n";
    echo "<title>Catalogo de genes</title>\n";
    echo "<link rel='stylesheet' href='css.css'>\n";
    echo "</head>\n";
    echo "<body>\n";
    echo "<h1>Catalogo de genes</h1>\n";
    echo "<table>\n";
    echo "<tr><th>GEN</th><th>TRANSCRIPT</th><th>CROMOS</th><th>BRI</th><th>START</th><th>END</th><th>E  
XONS</th>\n";

    for ($i=0; $i<$items; $i++)
    {
```

--- php3.php 25% L35 (Fundamental)

Applications Places System

emcs@ubuntuM0151

File Edit Options Buffers Tools Help

Save Undo

```

echo "<html>\n";
echo "<head>\n";
echo "<title>Catalogo de genes</title>\n";
echo "<link rel=\"stylesheet\" href=\"css.css\">\n";
echo "</head>\n";
echo "<body>\n";
echo "<h1>Catalogo de genes</h1>\n";
echo "<table>\n";
echo "<tr><th>GEN</th><th>TRANSCRIT</th><th>CROMOS</th><th>BRI</th><th>START</th><th>END</th><th>EXONS</th>\n";

for ($i=0; $i<$items; $i++)
{
$row= mysqli_fetch_array($result);

$name2 = $row["name2"];
$name = $row["name"];
$chrom = $row["chrom"];
$strand = $row["strand"];
$start = $row["txStart"];
$end = $row["txEnd"];
$exonCount = $row["exonCount"];

echo "<tr><td>$name2</td><td>$name</td><td>$chrom</td><td>$strand</td>";
echo "<td>$start</td><td>$end</td><td>$exonCount</td>\n";
}

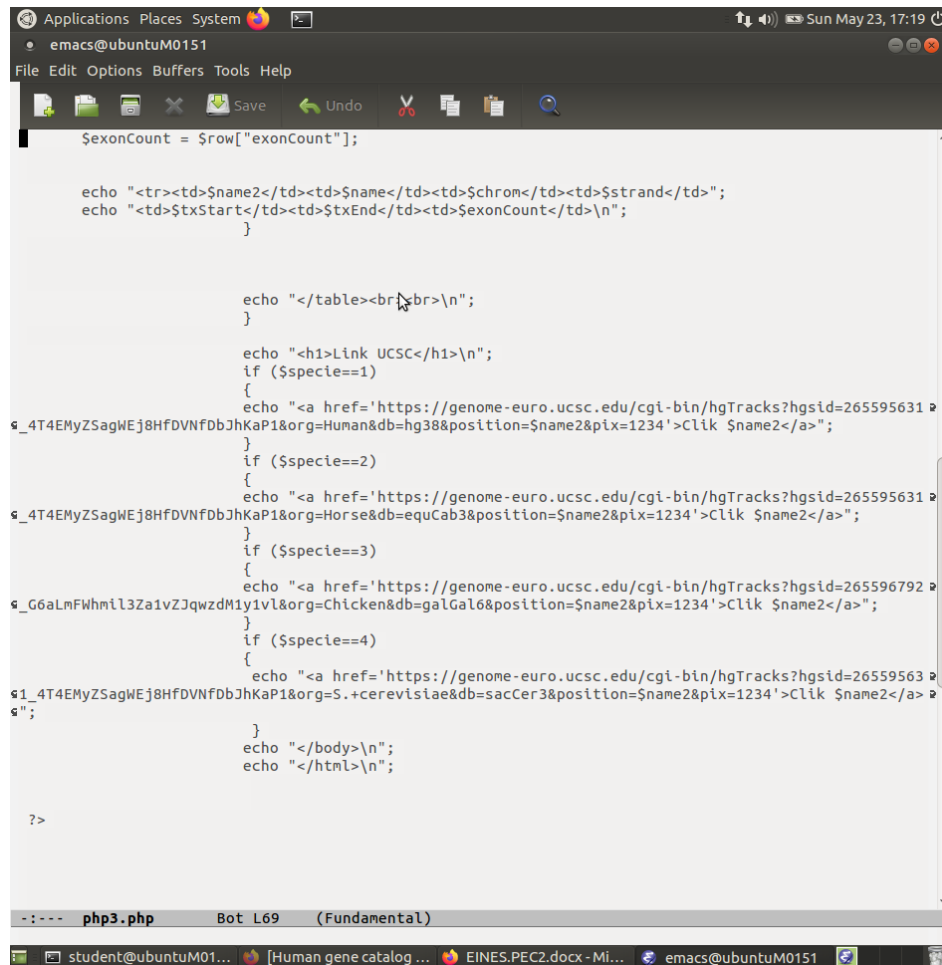
echo "</table><br><br>\n";
}

echo "<h1>Link UCSC</h1>\n";
if ($specie==1)
{
echo "<a href='https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgId=265595631&org=Human&db=hg38&position=$name2&pix=1234'>Clik $name2</a>";
}
if ($specie==2)
{
echo "<a href='https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgId=265595631&org=Human&db=hg38&position=$name2&pix=1234'>Clik $name2</a>";
}

```

47% L79 (Fundamental)

EINES INFORMÀTIQUES PER A LA BIOINFORMÀTICA. PAC2
Amelia Martínez Sequera.



```
$exonCount = $row["exonCount"];

echo "<tr><td>$name2</td><td>$name</td><td>$chrom</td><td>$strand</td>";
echo "<td>$txStart</td><td>$txEnd</td><td>$exonCount</td>\n";
}

echo "</table><br><br>\n";
}

echo "<h1>Link UCSC</h1>\n";
if ($specie==1)
{
echo "<a href='https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgside=265595631&_4T4EMyZSagWEj8HfDVNFdbJhKaP1&org=Human&db=hg38&position=$name2&pix=1234'>Clik $name2</a>";
}
if ($specie==2)
{
echo "<a href='https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgside=265595631&_4T4EMyZSagWEj8HfDVNFdbJhKaP1&org=Horse&db=equCab3&position=$name2&pix=1234'>Clik $name2</a>";
}
if ($specie==3)
{
echo "<a href='https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgside=265596792&_G6aLmFwhmil3Za1vZJqwdM1y1vL&org=Chicken&db=galGal6&position=$name2&pix=1234'>Clik $name2</a>";
}
if ($specie==4)
{
echo "<a href='https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgside=265595631&_4T4EMyZSagWEj8HfDVNFdbJhKaP1&org=S.+cerevisiae&db=sacCer3&position=$name2&pix=1234'>Clik $name2</a>";
}
}
echo "</body>\n";
echo "</html>\n";

?>
```

CSS: (css.css)

body

```
{
    background-color:#DDDDDD;
    padding-top: 20px;
    padding-bottom: 20px;
    padding-left: 50px;
    padding-right: 50px;
}
```

h1

```
{
    background: white;
    width: 100%;
```

```
border: 0px solid black;  
border-radius: 25px;  
color: darkblue;  
height: 35px;  
font-family: verdana, sans-serif;  
font-size: 18pt;  
font-weight: bold;  
text-align: center;  
}
```

```
p  
{  
  font-family: verdana, sans-serif;  
}
```

```
table  
{  
  border: 5px solid #FFA500;  
  border-radius: 25px;  
  padding: 20;  
  width: 100%;  
  font-family: verdana, sans-serif;  
}
```

```
td  
{  
  padding: 30;  
  font-weight: bold;  
  font-size: 12pt;  
  border: 1px solid #D0D0D0;  
  border-radius: 25px;  
}
```

```
th
{
    padding: 10;
    font-weight: bold;
    font-size: 18pt;
    text-align: left;
    text-transform: uppercase;
}
```

```
a:link, a:visited
{
    color: blue;
    text-decoration: none;
    font-weight: bold;
}
```

```
a:hover
{
    color: white;
    background: darkblue;
}
```

```
input
{
    width:300px;
}
```