In [2]:
```
#Antes de nada, debemos de preparar el ambiente de computación con el que
#vamos a trabajar.
#Para ello, vamos a utilizar el siguiente comando, que nos va a permitir
#obtener el ambiente adecuado de forma automática:

system2(command = './preparar_ambiente.sh', wait = TRUE)
```

INTRODUCCIÓN

En esta práctica vamos a llevar a cabo un análisis que va a consistir en explorar la distribución taxonómica de las secuencias homólogas a una proteína de interés, la proteína CHRM1. Esta, junto con las proteínas CHRNA3 y CHRNA7, se trata de una proteína que participa en la neurotransmisión de los humanos. La CHRM1, concretamente, es un receptor muscarínico de acetilcolina (CHolinergic Receptor Muscarinic 1 (NCBI)). Se conoce que el origen de estas familias de proteínas se encuentra en el linaje del último antepasado común entre cnidarios y cordados, por lo que surgieron durante la época inicial de la evolución de los animales. Esto implica, por tanto, que no se han encontrado proteínas homólogas en linajes anteriores como ctneóforos, poríferos, placozoa, hongos o plantas.

MÉTODOS

Este análisis de búsqueda de secuencias homólogas lo vamos a realizar a partir de la base de datos de Swissprot, ya que es una base de datos de secuencias de proteínas que está incluida en el ambiente de trabajo que estamos usando. Para la búsqueda utilizaremos el programa blastp, el cual nos permite ir introduciendo diferentes valores E. Esto nos va a servir para ver el ritmo al que aumenta la distribución taxonómica de las secuencias encontradas a medida que vamos disminuyendo el grado de similitud que exigimos que haya entre las secuencias homólogas y nuestra secuencia de interés.

RESULTADOS

- En primer lugar, vamos a llevar a cabo una primera búsqueda individual con blastp en la que vamos a establecer un valor E muy exigente, de 1e-50. De esta forma nos aseguramos de que solo nos aparezcan aquellas secuencias que sean extremadamente similares a la secuencia de nuestra proteína de interés CHRM1.

In [7]:
```
#El comando para hacer esta primera búsqueda es el siguiente:

BlastpOut01 <- system2(command = 'blastp',
                       args = c('-db', 'swissprot',
                                '-query', 'CHRM1.fas',
                                '-evalue', '1.0e-50',
                                '-outfmt', '"7 saccver pident length qstart q
                       stdout = TRUE)
```

In [8]:

```
#Como resultado de este comando, obtenemos un archivo "BlastOut01" que
#contiene mucho texto plano.
#Para poder observar de una forma más visual su contenido vamos a
#utilizar las funciones textConnection() y read.table(), que permiten
#transformarlo en un "data frame", es decir, en una tabla donde cada
#columna es una variable y la información está más organizada.

TablaOut01 <- read.table(textConnection(BlastpOut01),
                         sep = '\t',
                         col.names = c('saccver', 'pident', 'length', 'qstar
                                       'qend', 'sstart', 'send', 'evalue', 's
                                       'sciname', 'sblastname'))

#Para conocer las dimensiones de la tabla generada (nº de filas y
#columnas) utilizamos el siguiente comando:

dim(TablaOut01)
```

37 · 11

In [9]:

```
#Pedimos que nos muestre la tabla:

TablaOut01
```

A data.frame: 37 × 11

saccver	pident	length	qstart	qend	sstart	send	evalue	staxid	sciname
<fct>	<dbl>	<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<fct>
P11229.2	100.000	460	1	460	1	460	0.00e+00	9606	Homo sapiens
P56489.1	99.565	460	1	460	1	460	0.00e+00	9544	Macaca mulatta
Q5R949.1	99.565	460	1	460	1	460	0.00e+00	9601	Pongo abelii
P04761.1	99.130	460	1	460	1	460	0.00e+00	9823	Sus scrofa
P12657.2	98.913	460	1	460	1	460	0.00e+00	10090	Mus musculus
P08482.1	98.696	460	1	460	1	460	0.00e+00	10116	Rattus norvegicus
Q9N2A4.1	51.731	520	10	438	51	564	1.14e-170	9598	Pan troglodytes
P20309.1	51.731	520	10	438	51	564	2.19e-170	9606	Homo sapiens
Q9N2A3.1	51.737	518	10	436	51	562	1.08e-169	9595	Gorilla gorilla gorilla
P11483.1	50.769	520	10	438	51	564	3.15e-169	9823	Sus scrofa
Q9N2A2.1	51.670	509	19	438	62	564	3.59e-169	9600	Pongo pygmaeus

P41984.1	50.577	520	10	438	51	564	3.75e-169	9913	Bos taurus
P08483.1	51.670	509	19	438	61	563	3.87e-169	10116	Rattus norvegicus
P08912.2	51.935	491	23	441	28	518	4.60e-168	9606	Homo sapiens
Q9ERZ3.1	50.586	512	19	438	61	563	3.14e-167	10090	Mus musculus
P56490.1	51.527	491	23	441	28	518	6.41e-167	9544	Macaca mulatta
Q5IS53.1	51.324	491	23	441	28	518	9.87e-164	9598	Pan troglodytes
P49578.1	48.837	516	10	438	99	613	1.15e-156	9031	Gallus gallus
P17200.1	48.214	448	24	437	41	483	1.12e-141	9031	Gallus gallus
P32211.1	48.198	444	24	437	30	472	1.56e-138	10090	Mus musculus
P10980.2	47.153	439	23	437	21	459	9.08e-138	10116	Rattus norvegicus
P08173.2	47.973	444	24	437	31	472	9.18e-138	9606	Homo sapiens
P08485.1	47.973	444	24	437	30	471	4.95e-137	10116	Rattus norvegicus
Q9ERZ4.2	46.925	439	23	437	21	459	8.25e-137	10090	Mus musculus
P08172.1	46.697	439	23	437	21	459	3.31e-136	9606	Homo sapiens
P41985.2	45.796	452	23	437	20	458	6.36e-136	9913	Bos taurus
P06199.1	46.241	439	23	437	21	459	6.72e-135	9823	Sus scrofa
Q9N2A7.1	46.759	432	30	437	2	433	3.76e-134	9598	Pan troglodytes
P30372.1	44.444	459	2	437	3	459	2.76e-132	9031	Gallus gallus
P30544.1	45.975	472	18	438	25	478	1.12e-129	8355	Xenopus laevis
P08911.1	73.488	215	20	234	24	238	9.43e-111	10116	Rattus norvegicus
Q920H4.2	74.882	211	20	230	25	235	9.45e-110	10090	Mus musculus
P16395.2	53.061	245	12	251	87	323	3.30e-77	7227	Drosophila melanogaster

Q9U7D5.2	50.000	208	24	230	64	271	1.08e-66	6239	Caenorhabditis elegans
Q9JI35.2	33.333	408	39	433	49	427	4.11e-55	10141	Cavia porcellus
Q93126.1	29.451	455	30	444	38	458	1.06e-54	1232801	Amphibalanus amphitrite
Q64264.2	28.899	436	6	437	18	419	3.47e-51	10090	Mus musculus

In [10]:
```
#Como podemos observar, en la tabla hay 37 filas, es decir, se han
#encontrado 37 secuencias. De estas, 36 corresponden con las secuencias de
#proteínas con suficiente similitud a nuestra proteína de interés que ha
#encontrado el programa y, además, también está incluida la secuencia de la
#proteína de interés CHRM1, por eso hay 37.
```

- Ahora, vamos a llevar a cabo una búsqueda en serie con el blastp. Esto nos va a permtir ir utilizando diferentes valores E para poder observar cómo va aumentando el número de secuencias homólogas que encuentra el programa a medida que utilizamos un valor E menos exigente, es decir, un valor E mayor.

Esto podemos hacerlo ejecutando el comando anterior varias veces y cambiando cada vez el valor que le damos al parámetro '-evalue' para ver cómo va variando el resultado o podemos hacerlo de forma automática. Para hacerlo de esta segunda forma podemos usar la función 'lapply()'.

In [11]:
```
#Primero, vamos a generar un vector donde introduzcamos los diferentes
#valores E con los que queremos hacer el análisis anterior en serie:

Valores_E_maximos <- c('1.0e-50', '1.0e-40', '1.0e-30', '1.0e-20', '1.0e-1(
                       '1.0e-08', '1.0e-06', '1.0e-04', '1.0e-02', '1')
```

In [12]:

```
#Ahora, vamos a usar la función 'lapply()' para ejecutar la búsqueda en
#blastp para cada valor del vector `Valores_E_maximos`.
#Como resultado vamos a obtener una tabla como la anterior para cada valor
#E, por lo que tendremos una lista de tablas:

Lista_de_Tablas <- lapply(Valores_E_maximos,
                          function(x) {
                              BlastpOut <- system2(
                                  command = 'blastp',
                                  args = c('-db', 'swissprot',
                                           '-query', 'CHRM1.fas',
                                           '-evalue', x,
                                           '-outfmt',
                                           '"7 saccver pident length qstart
                                  stdout = TRUE)
                              read.table(textConnection(BlastpOut),
                                         sep = '\t',
                                         col.names = c('saccver', 'pident',
                                            'qend', 'sstart', 'send', 'eval
                                            'ssciname', 'sblastname'))
                          })
```

In [14]:

```
#Si queremos visualizar una de las 10 tablas generadas con el comando
#anterior, podemos indicar entre dobles corchetes el número de la
#tabla individual que queremos ver:

tail(Lista_de_Tablas[[3]])
```

A data.frame: 6 × 11

	saccver	pident	length	qstart	qend	sstart	send	evalue	staxid	ssciname
	<fct>	<dbl>	<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<fct>
89	Q09388.3	36.323	223	34	241	14	236	2.66e-37	6239	Caenorhabditis elegans
90	Q60476.1	26.699	412	34	433	56	445	1.07e-34	10141	Cavia porcellus
91	Q588Y6.1	33.761	234	3	235	26	247	1.10e-32	9685	Felis catus
92	Q25322.1	30.085	236	23	258	49	271	1.07e-31	7004	Locusta migratoria
93	Q25321.1	30.085	236	23	258	49	271	1.12e-31	7004	Locusta migratoria
94	P35404.1	33.645	214	22	235	46	246	6.95e-31	9267	Didelphis virginiana

In [15]:

```
#Como vemos, nos muestra la tabla con las secuencias homólogas encontradas
#partir de la búsqueda en blastp con el tercer valor E que habíamos dado,
#el de 1.0e-30.

#Podemos probar a buscar otras tablas con otros valores E:

tail(Lista_de_Tablas[[6]])
```

A data.frame: 6 × 11

	saccver	pident	length	qstart	qend	sstart	send	evalue	staxid	ssciname	sbla
	<fct>	<dbl>	<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<fct>	
572	Q86917.1	25.758	198	42	233	108	289	1.32e-09	10269	Sheeppox virus KS-1	
573	Q9P1P4.1	23.834	193	42	225	48	234	1.47e-09	9606	Homo sapiens	I
574	Q9Y5X5.2	27.338	139	22	160	144	280	1.58e-09	9606	Homo sapiens	I
575	Q6W3F4.1	30.709	127	39	165	21	145	1.87e-09	9615	Canis lupus familiaris	ca
576	Q64077.1	28.221	163	2	163	9	169	1.88e-09	10141	Cavia porcellus	
577	P32302.1	27.835	194	17	204	43	234	1.99e-09	9606	Homo sapiens	I

In [16]:

```
tail(Lista_de_Tablas[[10]])
```

A data.frame: 6 × 11

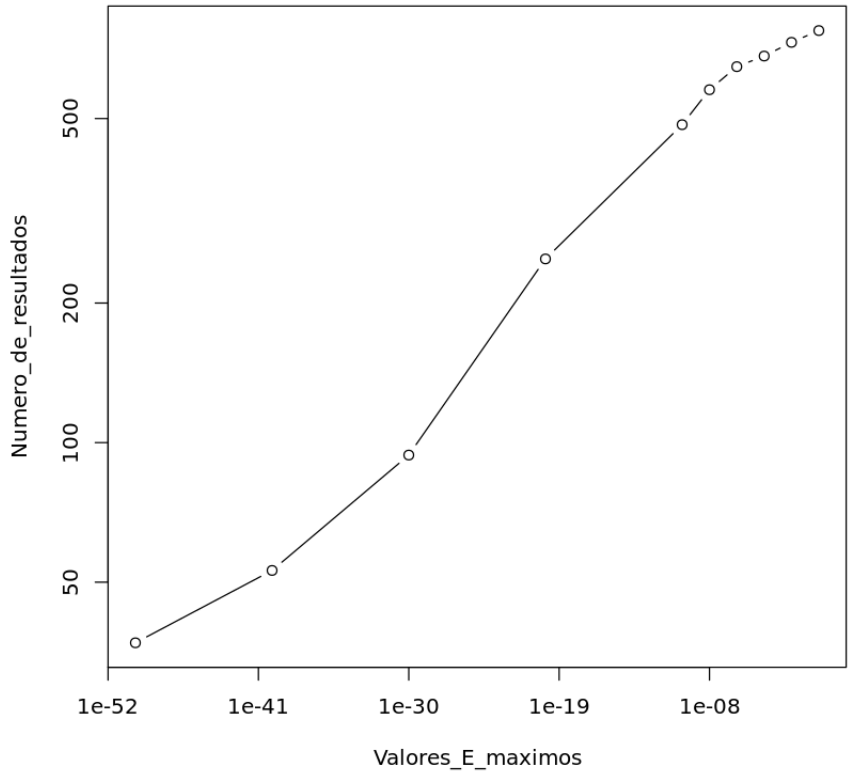
	saccver	pident	length	qstart	qend	sstart	send	evalue	staxid	ssciname	s
	<fct>	<dbl>	<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<fct>	
769	O62795.1	28.477	151	22	170	34	181	1.32e-09	39089	Phoca groenlandica	
770	Q86917.1	25.758	198	42	233	108	289	1.32e-09	10269	Sheeppox virus KS-1	
771	Q9P1P4.1	23.834	193	42	225	48	234	1.47e-09	9606	Homo sapiens	
772	Q9P1P4.1	32.143	84	350	433	241	324	4.38e-06	9606	Homo sapiens	
773	Q9Y5X5.2	27.338	139	22	160	144	280	1.58e-09	9606	Homo sapiens	
774	Q9Y5X5.2	30.380	79	359	430	370	447	2.83e-04	9606	Homo sapiens	

In [17]:
```
#Como usamos el comando tail, solamente nos muestra una tabla con las
#últimas secuencias encontradas por el programa blastp con los diferentes
#valores E que nosotros especificamos. Con esto podemos observar y verifica
#que cuanto mayor es el valor E utilizado, mayor número de secuencias
#homólogas se encuentran, porque el programa es menos restrictivo a la hora
#de buscar secuencias homólogas a nuestra proteína original CHRM1.
```

In [18]:
```
#Si queremos saber el número de secuencias homólogas que ha encontrado el
#blastp en la base de datos de Swissprot con cada uno de los valores E que
#le hemos indicado al elaborar la Lista de tablas, podemos utilizar la
#función 'dim()' para que nos dé las dimensiones de las diferentes tablas
#que ha elaborado, de manera que, sabiendo el número de filas y columnas
#podremos conocer el número de secuencias, ya que cada fila corresponde cor
#una secuencia.
#Por tanto, al utilizar la función 'dim()', especificamos que nos dé solo e
#primer valor, correspondiente con el número de filas, indicando un 1 entre
#corchetes:

Numero_de_resultados <- sapply(Lista_de_Tablas, function(x) dim(x)[1])

#Pedimos que nos dé el resultado del número de secuencias homólogas
#encontradas para cada valor E:

Numero_de_resultados
```

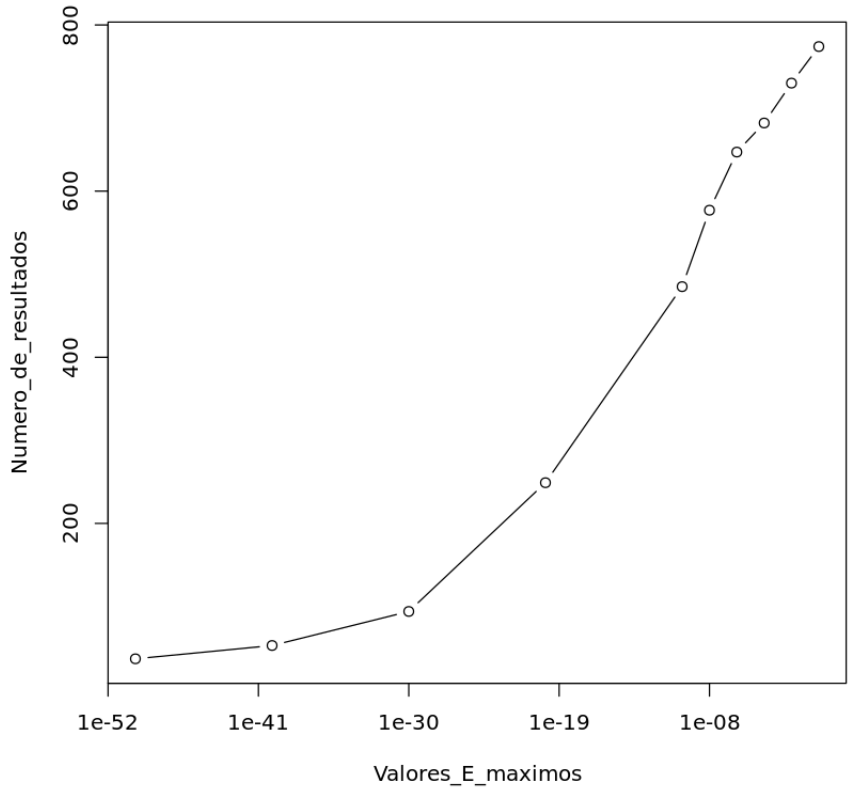
$37 \cdot 53 \cdot 94 \cdot 249 \cdot 485 \cdot 577 \cdot 647 \cdot 682 \cdot 730 \cdot 774$

In [19]:
```
#Podemos representar el número de resultados obtenidos para cada valor E en
#una gráfica. Esta gráfica puede estar en escala logarítmica solamente en
#un eje ("log='x'"), en los dos, o en ninguno.
#En este caso, la hacemos en escala logarítmica para los dos ejes:

plot(Valores_E_maximos, Numero_de_resultados, log = 'xy', type = 'b')
```



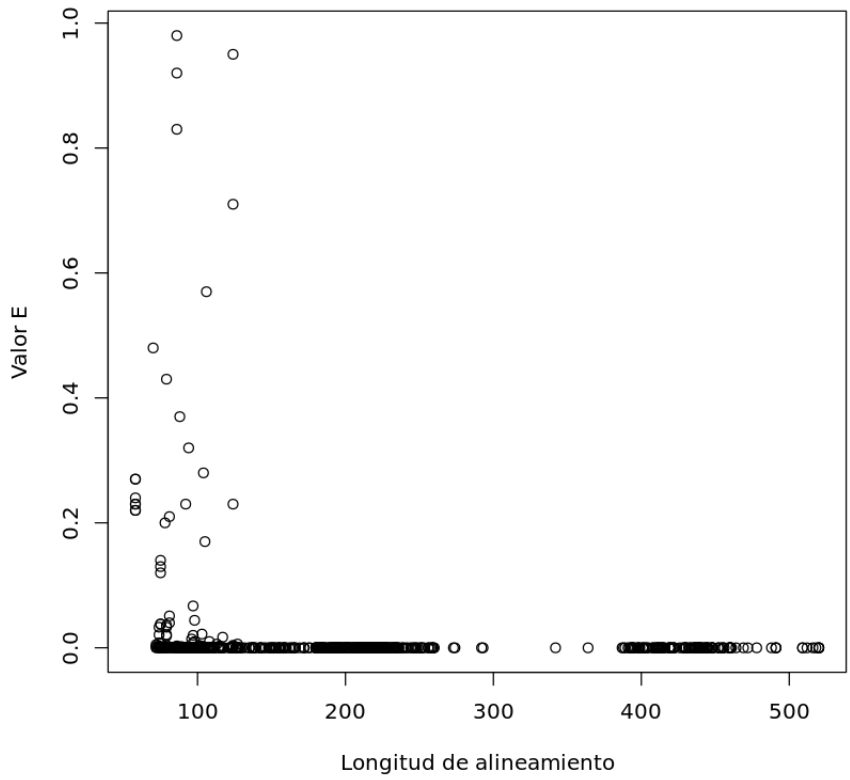
In [20]:

```
#Probamos, también, a ver la gráfica resultante si solamente ponemos en
#escala logarítmica los valores E:

plot(Valores_E_maximos, Numero_de_resultados, log = 'x', type = 'b')
```



In [21]:

```
#Ahora vamos a ver a partir de la última tabla elaborada con el blastp con
#el último valor E, cuál es la relación entre la longitud del alineamiento
#y el valor E, ya que en esta tabla número 10 se ha usado un valor E de 1 y
#por tanto, es el menos restrictivo, por lo que tiene un mayor número de
#resultados.

#Para llevar a cabo esta acción, ejecutamos el siguiente comando, donde el
#símbolo $ nos va a permitir averigurar en la tabla que especificamos, la
#información que se encuentra dentro de la columna que se indica después de
#este símbolo:

plot(Lista_de_Tablas[[10]]$length, Lista_de_Tablas[[10]]$evalue,
     xlab = 'Longitud de alineamiento', ylab = 'Valor E')
```



In [22]:

```
#La proteína CHRM1 sabemos que presenta 460 aa, sin embargo, vemos que hay
#varios alineamientos que se sitúan por debajo de este valor, llegando
#incluso a los 100 aa o menos, lo que nos indica que hay homologías
#parciales entre las secuencias que nos está encontrando el blastp con
#nuestra secuencia de interés.

#Vamos a pedir que nos enseñe las diferentes longitudes que presentan las
#secuencias que se encuentran cuando se utiliza el valor E menos restrictiv
#poniendo el siguiente comando:

Lista_de_Tablas[[10]]$length
```

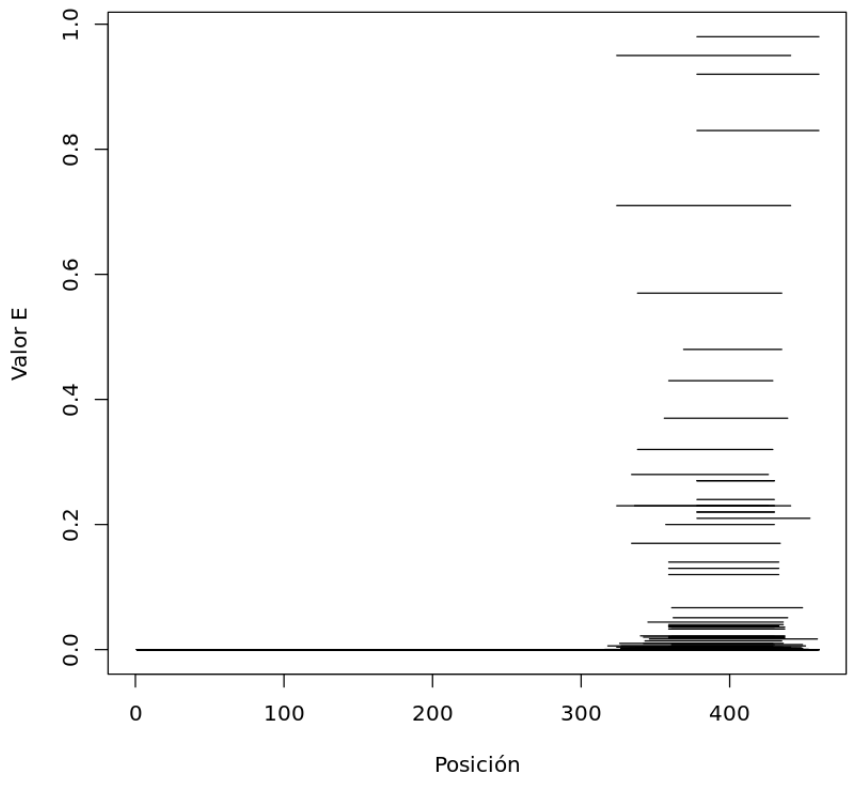
460 · 460 · 460 · 460 · 460 · 460 · 520 · 520 · 518 · 520 · 509 · 520 · 509 · 491 · 512 · 491 · 491 · 516 · 448 · 444 · 439 · 444 · 444 · 439 · 439 · 452 · 439 · 432 · 459 · 472 · 215 · 109 · 211 · 109 · 245 · 85 · 208 · 164 · 456 · 488 · 464 · 408 · 461 · 455 · 453 · 430 · 469 · 460 · 439 · 436 · 448 · 455 · 444 · 448 · 445 · 448 · 448 · 445 · 414 · 415 · 410 · 414 · 414 · 436 · 394 · 394 · 412 · 414 · 430 · 404 · 437 · 394 · 394 · 439 · 437 · 415 · 401 · 448 · 433 · 412 · 394 · 428 · 419 · 430 · 430 · 437 · 430 · 443 · 434 · 419 · 408 · 445 · 447 · 442 · 429 · 429 · 441 · 455 · 478 · 440 · 440 · 442 · 183 · 183 · 412 · 393 · 393 · 183 · 426 · 422 · 422 · 439 · 439 · 223 · 187 · 387 · 433 · 402 · 412 · 422 · 422 · 410 · 421 · 410 · 439 · 438 · 234 · 107 · 420 · 236 · 236 · 214 · 110 · 205 · 93 · 405 · 256 · 127 · 419 · 224 · 96 · 411 · 198 · 107 · 193 · 97 · 242 · 127 · 228 · 183 · 419 · 221 · 96 · 364 · 238 · 124 · 185 · 108 · 194 · 107 · 391 · 401 · 227 · 93 · 230 · 96 · 213 · 79 · 403 · 259 · 127 · 197 · 95 · 342 · 223 · 259 · 127 · 413 · 251 · 127 · 274 · 126 · 225 · 90 · 247 · 96 · 244 · 96 · 207 · 90 · 231 · 96 · 235 · 98 · 234 · 130 · 228 · 87 · 411 · 396 · … · 223 · 86 · 152 · 104 · 142 · 226 · 75 · 180 · 81 · 219 · 122 · 138 · 228 · 156 · 224 · 88 · 137 · 209 · 156 · 72 · 225 · 109 · 183 · 220 · 90 · 226 · 58 · 226 · 58 · 220 · 158 · 197 · 88 · 228 · 58 · 203 · 58 · 217 · 74 · 235 · 74 · 226 · 58 · 137 · 155 · 124 · 155 · 124 · 224 · 88 · 222 · 94 · 137 · 220 · 93 · 212 · 74 · 166 · 137 · 221 · 107 · 197 · 104 · 196 · 105 · 197 · 88 · 202 · 58 · 104 · 208 · 196 · 96 · 220 · 90 · 217 · 88 · 224 · 93 · 143 · 176 · 437 · 160 · 166 · 78 · 135 · 103 · 135 · 94 · 191 · 74 · 212 · 159 · 226 · 196 · 81 · 192 · 124 · 181 · 175 · 150 · 75 · 202 · 88 · 208 · 123 · 96 · 213 · 185 · 224 · 158 · 221 · 181 · 83 · 201 · 273 · 104 · 210 · 70 · 172 · 81 · 172 · 193 · 97 · 202 · 79 · 98 · 202 · 79 · 172 · 198 · 108 · 143 · 292 · 389 · 205 · 75 · 202 · 79 · 202 · 79 · 212 · 436 · 208 · 200 · 260 · 196 · 79 · 184 · 194 · 87 · 205 · 75 · 212 · 97 · 197 · 98 · 158 · 196 · 100 · 127 · 117 · 163 · 233 · 115 · 193 · 86 · 127 · 191 · 97 · 185 · 209 · 139 · 116 · 200 · 145 · 78 · 195 · 101 · 126 · 115 · 127 · 103 · 210 · 86 · 126 · 115 · 149 · 78 · 211 · 101 · 210 · 210 · 86 · 151 · 198 · 193 · 84 · 139 · 79

In [23]:

```
#Se nos muestra el vector con los datos de longitudes de secuencias con los
#que se ha elaborado la tabla anterior.
#De esta forma podemos observar que hay muchos números menores de 460, que
#es el número de aminoácidos de la proteína CHRM1.
#Esto implica que hay una gran cantidad de proteínas que se parecen a
#nuestra proteína original solamente en un fragmento.
```

In [25]:

```
#Procedemos a averiguar qué posiciones de la secuencia de la proteína CHRM
#son las que alinean parcialmente con el resto de secuencias encontradas
#con el valor E menos estricto.
#Para ello, primero damos nombre a los diferentes parámetros que queremos
#conocer de las secuencias:

inicios <- Lista_de_Tablas[[10]]$qstart
finales <- Lista_de_Tablas[[10]]$qend
valoresE <- Lista_de_Tablas[[10]]$evalue

#Hacemos, en primer lugar, lo que llamamos el Alineamiento Máximo, porque
#vamos a elaborar una gráfica y tenemos que determinar el rango del eje
#horizontal, estableciendo como valor máximo el "quend".
#Luego elaboramos un gráfico con los rangos adecuados, pero vacío (type='n
#Por último, utilizamos la función segments() para añadir segmentos al
#gráfico llevado a cabo previamente:

AlineamientoMaximo <- max(Lista_de_Tablas[[10]]$qend)

plot(c(0, AlineamientoMaximo), range(valoresE), type='n', xlab='Posición',

segments(inicios, valoresE, finales, valoresE)
```



In [26]:
```
#El gráfico resultante nos muesta que cuando se buscan secuencias homólogas
#con valores E menos estrictos la homología se concentra sobre todo en
#posiciones alrededor de 400. Esto nos quiere decir que las secuencias
#encontradas se parecen sobre todo a la parte final de la secuencia de la
#CHRM1, lo que puede significar que esta parte final corresponde con un
#fragmento más conservado de la proteína y que es compartido por proteínas
#de diferentes especies.
```

In [29]:
```
#En último lugar, vamos a explorar la distribución taxonómica de las
#secuencias encontradas con los diferentes valores E.
#Recordemos que, en las tablas que hemos elaborado, tenemos una columna
#'sblastname' que nos da el nombre del grupo taxonómico de rango superior
#al que pertenece cada secuencia con homología encontrada. De esta forma
#tenemos las secuencias clasificadas para facilitar nuestra interpretación.

#Para llevar a cabo el análisis de la distribución taxonómica, una opción
#es contar en cada tabla generada las secuencias que se han encontrado de
#cada grupo taxonómico de la columna 'sblastname'. Esto lo podemos hacer
#aplicando la función `table()` sobre esa columna de cada tabla en
#`Lista_de_Tablas`:
```

In [30]:
```
lapply(Lista_de_Tablas, function(x) table(x$sblastname))
```

```
[[1]]

               birds         crustaceans even-toed ungulates                f
lies
                   3                   1                   5
1
      frogs & toads           nematodes             primates              rod
ents
                   1                   1                  13
12

[[2]]

               birds         bony fishes         crustaceans even-toed ungul
ates
                   3                   4                   1
7
               flies       frogs & toads               moths           nemat
odes
                   1                   1                   1
1
           primates             rodents
                  15                  19

[[3]]

               birds         bony fishes          carnivores         crustac
eans
                   4                   6                   2
1
even-toed ungulates               flies       frogs & toads          gastro
pods
```

```
                   8                    2                    1
1
       grasshoppers          marsupials                moths            nemat
odes
                   2                    1                    3
2
 odd-toed ungulates             primates     rabbits & hares              rod
ents
                   2                   26                    1
32

[[4]]

               birds         bony fishes          carnivores         crustac
eans
                   6                   18                   16
2
even-toed ungulates               flies       frogs & toads          gastro
pods
                  23                   12                    8
2
       grasshoppers        insectivores          marsupials               m
oths
                   2                    2                    4
5
          nematodes  odd-toed ungulates          placentals            prim
ates
                   7                    3                    7
52
    rabbits & hares             rodents
                   5                   75

[[5]]

               birds            bivalves         bony fishes          carniv
ores
                  11                    1                   25
29
        cephalopods         crustaceans even-toed ungulates               f
lies
                   5                    3                   44
23
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  13                    6                    2
2
          lancelets             lizards          marsupials               m
oths
                   2                    1                    4
7
          nematodes  odd-toed ungulates          placentals            prim
ates
                  14                    4                    7
100
    rabbits & hares             rodents             viruses
                  11                  170                    1

[[6]]

               birds            bivalves         bony fishes          carniv
ores
```

```
                  11                    1                   27
36
        cephalopods         crustaceans even-toed ungulates                f
lies
                   6                    3                   49
28
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  15                    6                    2
2
          lancelets             lizards          marsupials                m
oths
                   2                    1                    6
7
          nematodes  odd-toed ungulates          placentals             prim
ates
                  14                    4                    7
117
    rabbits & hares             rodents             viruses
                  13                  218                    2

[[7]]

               birds             bivalves         bony fishes           carniv
ores
                  12                    1                   32
43
        cephalopods         crustaceans even-toed ungulates                f
lies
                   6                    3                   58
29
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  18                    7                    2
2
          lancelets             lizards          marsupials                m
oths
                   2                    1                    6
8
          nematodes  odd-toed ungulates          placentals             prim
ates
                  14                    5                    7
132
    rabbits & hares             rodents             viruses
                  13                  244                    2

[[8]]

               birds             bivalves         bony fishes           carniv
ores
                  12                    1                   32
44
        cephalopods         crustaceans even-toed ungulates                f
lies
                   6                    3                   61
29
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  20                    9                    2
2
          lancelets             lizards          marsupials                m
```

```
oths
                  2                   1                   7
8
          nematodes  odd-toed ungulates          placentals             prim
ates
                 14                   5                   7
140
    rabbits & hares             rodents             viruses
                 14                 261                   2

[[9]]

              birds            bivalves         bony fishes           carniv
ores
                 14                   1                  34
47
        cephalopods         crustaceans even-toed ungulates                f
lies
                  7                   3                  67
30
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                 21                   9                   2
2
          lancelets             lizards          marsupials                m
oths
                  2                   1                   7
8
          nematodes  odd-toed ungulates          placentals             prim
ates
                 15                   5                   7
150
    rabbits & hares             rodents             viruses
                 17                 279                   2

[[10]]

              birds            bivalves         bony fishes           carniv
ores
                 15                   2                  35
50
        cephalopods         crustaceans even-toed ungulates                f
lies
                 10                   3                  73
32
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                 22                   9                   2
2
          lancelets             lizards          marsupials                m
oths
                  3                   1                   7
9
          nematodes  odd-toed ungulates          placentals             prim
ates
                 15                   5                   7
162
    rabbits & hares             rodents             viruses
                 17                 291                   2
```

In [31]:

```
#Como resultado del último comando, obtenemos una lista donde aparece,
#para cada tabla generada a partir de cada valor E, un recuento de los
#diferentes grupos taxonómicos que se encuentran con secuencias homólogas a
#nuestra proteína de interés. Hemos mencionado anteriormente que esta
#proteína surge entre la evolución de cnidarios y cordados, por lo que
#entre los grupos taxonómicos que aparecen no deberían de encontrarse
#ctneóforos, poríferos, placozoa, ni hongos ni plantas.

#Entre las secuencias encontradas para los diferentes valores E, ninguna
#pertenece a los grupos anteriormente mencionados,por lo que sí que podemos
#afirmar que las homologías pertenecen a proteínas que han tenido una
#evolución posterior y sí podrían tener una base taxonómica.

#Cabe destacar que en la lista de grupos taxonómicos obtenida a partir del
#último blastp (umbral de valor E de 1), uno de los grupos taxonómicos que
#aparece es el de virus, cosa que me ha resultado interesante.
```

In [33]:

```
#Para finalizar, vamos a ver la relación que hay entre la longitud de los
#alineamientos que se han encontrado y la distribución taxonómica que
#acabamos de observar.
#Queremos saber cuáles son los grupos taxonómicos en los que hay secuencias
#que se parecen más a nuestra proteína de interés, por lo que vamos a pedir
#que nos muestre aquellos grupos en los que realmente haya encontrado
#secuencias con un grado de similitud elevado. Para ello, vamos a repetir
#el recuento de los grupos taxonómicos, pero teniendo en cuenta solo
#aquellas secuencias que presenten un alineamiento de al menos 300
#aminoácidos:

lapply(Lista_de_Tablas, function(x) {
       filtro <- x$length >= 300
       table(x[filtro, 'sblastname'])
    }
)
```

```
[[1]]

              birds         crustaceans even-toed ungulates                f
lies
                  3                   1                   5
0
      frogs & toads           nematodes             primates              rod
ents
                  1                   0                  13
10

[[2]]

              birds         bony fishes         crustaceans even-toed ungul
ates
                  3                   4                   1
7
              flies       frogs & toads               moths           nemat
odes
                  0                   1                   1
0
           primates             rodents
```

```
                  15                  15

[[3]]

               birds         bony fishes          carnivores          crustac
eans
                   4                   6                   1
1
even-toed ungulates               flies       frogs & toads          gastro
pods
                   8                   1                   1
1
       grasshoppers           marsupials               moths           nemat
odes
                   0                   0                   3
0
 odd-toed ungulates             primates     rabbits & hares             rod
ents
                   2                  25                   1
26

[[4]]

               birds         bony fishes          carnivores          crustac
eans
                   4                  10                   4
1
even-toed ungulates               flies       frogs & toads          gastro
pods
                  10                   2                   3
1
       grasshoppers        insectivores           marsupials               m
oths
                   0                   2                   1
3
          nematodes  odd-toed ungulates          placentals            prim
ates
                   3                   3                   7
32
    rabbits & hares             rodents
                   2                  33

[[5]]

               birds            bivalves         bony fishes          carniv
ores
                   4                   0                  10
4
        cephalopods         crustaceans even-toed ungulates               f
lies
                   0                   1                  13
2
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                   4                   1                   0
2
          lancelets             lizards          marsupials               m
oths
                   0                   0                   1
3
          nematodes  odd-toed ungulates          placentals            prim
```

```
ates
                  3                   3                   7
36
    rabbits & hares             rodents             viruses
                  2                  43                   0

[[6]]

               birds             bivalves         bony fishes           carniv
ores
                  4                   0                  10
4
        cephalopods         crustaceans even-toed ungulates                 f
lies
                  0                   1                  14
3
      frogs & toads           gastropods        grasshoppers         insectiv
ores
                  4                   1                   0
2
          lancelets             lizards          marsupials                 m
oths
                  0                   0                   1
3
          nematodes  odd-toed ungulates          placentals             prim
ates
                  3                   3                   7
37
    rabbits & hares             rodents             viruses
                  2                  46                   0

[[7]]

               birds             bivalves         bony fishes           carniv
ores
                  4                   0                  10
4
        cephalopods         crustaceans even-toed ungulates                 f
lies
                  0                   1                  15
3
      frogs & toads           gastropods        grasshoppers         insectiv
ores
                  4                   1                   0
2
          lancelets             lizards          marsupials                 m
oths
                  0                   0                   1
3
          nematodes  odd-toed ungulates          placentals             prim
ates
                  3                   3                   7
38
    rabbits & hares             rodents             viruses
                  2                  46                   0

[[8]]

               birds             bivalves         bony fishes           carniv
ores
                  4                   0                  10
```

```
4
        cephalopods         crustaceans even-toed ungulates               f
lies
                  0                   1                  15
3
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                  4                   1                   0
2
          lancelets             lizards          marsupials               m
oths
                  0                   0                   1
3
          nematodes  odd-toed ungulates          placentals             prim
ates
                  3                   3                   7
38
    rabbits & hares             rodents             viruses
                  2                  46                   0

[[9]]

              birds            bivalves         bony fishes           carniv
ores
                  4                   0                  10
4
        cephalopods         crustaceans even-toed ungulates               f
lies
                  0                   1                  15
3
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                  4                   1                   0
2
          lancelets             lizards          marsupials               m
oths
                  0                   0                   1
3
          nematodes  odd-toed ungulates          placentals             prim
ates
                  3                   3                   7
38
    rabbits & hares             rodents             viruses
                  2                  46                   0

[[10]]

              birds            bivalves         bony fishes           carniv
ores
                  4                   0                  10
4
        cephalopods         crustaceans even-toed ungulates               f
lies
                  0                   1                  15
3
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                  4                   1                   0
2
          lancelets             lizards          marsupials               m
oths
```

```
                  0                   0                   1
3
          nematodes  odd-toed ungulates          placentals             prim
ates
                  3                   3                   7
38
    rabbits & hares             rodents             viruses
                  2                  46                   0
```

In [34]:

```
#Podemos comprobar que el número de secuencias encontradas con una
#homología de 300 aminoácidos es menor.

#Vamos a comprobar, también, cuántas secuencias alinean exactamente con
#nuestra proteína de interés y a qué grupos taxonómicos pertenecen:

lapply(Lista_de_Tablas, function(x) {
       filtro <- x$length >= 460
       table(x[filtro, 'sblastname'])
    }
)
```

```
[[1]]

              birds         crustaceans even-toed ungulates                f
lies
                  1                   0                   3
0
      frogs & toads           nematodes            primates             rod
ents
                  1                   0                  10
4

[[2]]

              birds         bony fishes         crustaceans even-toed ungul
ates
                  1                   0                   0
3
              flies       frogs & toads              moths             nemat
odes
                  0                   1                   0
0
           primates             rodents
                 10                   4

[[3]]

              birds         bony fishes          carnivores         crustac
eans
                  1                   0                   0
0
even-toed ungulates               flies       frogs & toads          gastro
pods
                  4                   0                   1
0
       grasshoppers          marsupials              moths             nemat
odes
                  0                   0                   0
0
```

```
 odd-toed ungulates             primates     rabbits & hares              rod
ents
                   0                   13                   0
5

[[4]]

               birds          bony fishes          carnivores          crustac
eans
                   1                    0                    0
0
even-toed ungulates                flies       frogs & toads           gastro
pods
                   4                    0                    1
0
       grasshoppers         insectivores          marsupials                m
oths
                   0                    0                    0
0
          nematodes  odd-toed ungulates          placentals             prim
ates
                   1                    0                    0
13
    rabbits & hares              rodents
                   0                    5

[[5]]

               birds             bivalves         bony fishes          carniv
ores
                   1                    0                    0
0
        cephalopods         crustaceans even-toed ungulates                 f
lies
                   0                    0                    4
0
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                   1                    0                    0
0
          lancelets             lizards          marsupials                m
oths
                   0                    0                    0
0
          nematodes  odd-toed ungulates          placentals             prim
ates
                   1                    0                    0
13
    rabbits & hares              rodents             viruses
                   0                    5                    0

[[6]]

               birds             bivalves         bony fishes          carniv
ores
                   1                    0                    0
0
        cephalopods         crustaceans even-toed ungulates                 f
lies
                   0                    0                    4
0
```

```
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  1                   0                   0
0
          lancelets             lizards          marsupials               m
oths
                  0                   0                   0
0
          nematodes  odd-toed ungulates          placentals            prim
ates
                  1                   0                   0
13
    rabbits & hares             rodents             viruses
                  0                   5                   0

[[7]]

               birds            bivalves         bony fishes          carniv
ores
                  1                   0                   0
0
        cephalopods         crustaceans even-toed ungulates               f
lies
                  0                   0                   4
0
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  1                   0                   0
0
          lancelets             lizards          marsupials               m
oths
                  0                   0                   0
0
          nematodes  odd-toed ungulates          placentals            prim
ates
                  1                   0                   0
13
    rabbits & hares             rodents             viruses
                  0                   5                   0

[[8]]

               birds            bivalves         bony fishes          carniv
ores
                  1                   0                   0
0
        cephalopods         crustaceans even-toed ungulates               f
lies
                  0                   0                   4
0
      frogs & toads          gastropods        grasshoppers        insectiv
ores
                  1                   0                   0
0
          lancelets             lizards          marsupials               m
oths
                  0                   0                   0
0
          nematodes  odd-toed ungulates          placentals            prim
ates
                  1                   0                   0
```

```
13
    rabbits & hares              rodents              viruses
                  0                    5                    0

[[9]]

               birds             bivalves         bony fishes           carniv
ores
                   1                    0                    0
0
        cephalopods         crustaceans even-toed ungulates                f
lies
                   0                    0                    4
0
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                   1                    0                    0
0
          lancelets              lizards          marsupials                m
oths
                   0                    0                    0
0
          nematodes  odd-toed ungulates          placentals             prim
ates
                   1                    0                    0
13
    rabbits & hares              rodents              viruses
                   0                    5                    0

[[10]]

               birds             bivalves         bony fishes           carniv
ores
                   1                    0                    0
0
        cephalopods         crustaceans even-toed ungulates                f
lies
                   0                    0                    4
0
      frogs & toads          gastropods        grasshoppers         insectiv
ores
                   1                    0                    0
0
          lancelets              lizards          marsupials                m
oths
                   0                    0                    0
0
          nematodes  odd-toed ungulates          placentals             prim
ates
                   1                    0                    0
13
    rabbits & hares              rodents              viruses
                   0                    5                    0
```

In [35]:
```
#El grupo taxonómico donde mayor cantidad de secuencias homólogas se
#encuentra es el de los primates, lo que tiene mucho sentido, ya que
#estamos buscando homologías para una proteína humana.

#Sin embargo, debemos de tener en cuenta que, realmente, que hayan alineado
#460 aminoácidos no quieren decir que se trate de una homología completa
#con nuestra proteína de interés, ya que puede haber alineado en porciones
#distintas y, por tanto, que solo por una parte de la secuencia de la
#proteína CHRM1 sean homólogas.
```

DISCUSIÓN

Como conclusión podemos sacar que la proteína CHRM1 es una proteína humana con un origen muy temprano, presente en gran cantidad de grupos taxonómicos desde su aparición entre cnidarios y cordados. Esta presenta 460 aminoácidos, de los cuales la mayoría de especies presentan proteínas con homologías en las últimas posiciones, alrededor de la posición 400, lo que indica la presencia de un posible fragmento muy conservado a lo largo de la evolución.

Para comprobar esto, podríamos llevar a cabo una búsqueda de la secuencia en la base de datos Pfam (Mistry et al, 2021) para comprobar si esta parte de la proteína corresponde con algún dominio conservado típico de alguna familia de proteínas.

BIBLIOGRAFÍA

- Lucas Henriques Viscardi, Danilo Oliveira Imparato, Maria Cátira Bortolini, Rodrigo Juliani Siqueira Dalmolin, Ionotropic Receptors as a Driving Force behind Human Synapse Establishment, Molecular Biology and Evolution, Volume 38, Issue 3, March 2021, Pages 735–744, doi:10.1093/molbev/msaa252 (https://doi:10.1093/molbev/msaa252).

- Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers and David J. Lipman. Basic local alignment search tool, Journal of Molecular Biology, Volume 215, Issue 3, 1990, Pages 403-410, doi:10.1016/S0022-2836(05)80360-2 (https://doi.org/10.1016/S0022-2836(05)80360-2).

- Eric W Sayers, Jeff Beck, J Rodney Brister, Evan E Bolton, Kathi Canese, Donald C Comeau, Kathryn Funk, Anne Ketter, Sunghwan Kim, Avi Kimchi, Paul A Kitts, Anatoliy Kuznetsov, Stacy Lathrop, Zhiyong Lu, Kelly McGarvey, Thomas L Madden, Terence D Murphy, Nuala O'Leary, Lon Phan, Valerie A Schneider, Françoise Thibaud-Nissen, Bart W Trawick, Kim D Pruitt, James Ostell, Database resources of the National Center for Biotechnology Information, Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D9–D16, https://doi.org/10.1093/nar/gkz899

- Jaina Mistry, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A Salazar, Erik L L Sonnhammer, Silvio C E Tosatto, Lisanna Paladin, Shriya Raj, Lorna J Richardson, Robert D Finn, Alex Bateman. Pfam: The protein families database in 2021, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D412–D419, doi:10.1093/nar/gkaa913 (https://doi:10.1093/nar/gkaa913)