

Programa analizador de retrovirus

Roberto Reinoso

roberto117343@gmail.com

En el presente artículo se va a mostrar un programa, que sirve para realizar una serie de análisis a partir de secuencias provirales de retrovirus.

Este software esta basado en otro de mis programas (<https://github.com/roberto117343/rstrdrtrvs/blob/master/main.java>) que realiza un rastreo del ADN del provirus, en busca de secuencias aminoacídicas y las de nucleótidos que dan a las anteriores [1].

El código fuente del programa, que se presenta en este artículo, se encuentra escrito en Java bajo la licencia de software libre GNU GPL [2], en la siguiente dirección:

<https://github.com/roberto117343/Anresis/blob/master/main.java>

Podemos dividir el funcionamiento del programa en dos fases, donde en la primera se rastrea el ADN proviral buscando posibles proteínas que este contiene. El programa también arroja las cadenas de nucleótidos que dan lugar a dichas cadenas proteicas. La búsqueda no es perfecta, por lo que al acabar esta fase debemos hacer una selección para descartar secuencias incompletas, las que no son de nuestra utilidad y otros posibles artefactos. En la figura 1. se presenta una pequeña muestra de la tasa de acierto de esta fase.

La segunda fase corresponde a el análisis de las secuencias deseadas. Dicho análisis arroja la longitud de la secuencia de nucleótidos, la longitud de la secuencia de aminoácidos, el inicio y el final de la secuencia de nucleótidos (dentro de la cadena del provirus), el marco de lectura, la proporción de nucleótidos y aminoácidos (que contiene tanto una secuencia como la otra) y la masa molecular de la secuencia proteica (en este calculo no se

consideran los aminoácidos anómalos).

clone:pUG031-A2 (VIH-1)	gag	pol	rev
	Completa	Parcial	No Encontrada
	tat	vif	vpu
	Parcial	Completa	Completa
	nef	vpr	env
	Completa	Completa	Completa
	gag	pol	env
SDAU14A1 (ALV)	Completa	Parcial	Parcial
C3H (MMTV)	gag	env	sag
	Completa	Mayor	Completa
	env	orfa	orfb
NY 2003 (WDSV)	Completa	Completa	Completa
	orfC		
	Completa		
CV79 (HTLV-1)	gag	pol	tax
	Completa	Completa	Parcial
	env		
	Completa		
	gag	pol	bet
	Completa	Completa	Parcial
AG16 (SFV)	env	tas	
	Completa	Completa	
	env		
clone33 (FeLV)	Completa		

Figura 1. Muestra de la efectividad del software a la hora de encontrar secuencias proteicas y las nucleotídicas correspondientes.

Pasemos ahora a exponer como usar el programa.

Lo primero que se nos indica es que introduzcamos la cadena de ADN del provirus. Esta debe ir sin espacios, ni saltos de linea, etc.

Introduce La Secuencia De ADN Del Provirus:

Figura 2. Petición de la entrada del ADN del provirus desde el software.

Una vez introducida, el programa, la rastrea en busca de proteínas que esta posee y las secuencias de nucleótidos que dan a estas, expresadas en forma de ARN (con la tasa de acierto antes mencionada). A cada par de secuencias se le asigna un identificador (en el programa se encuentra como #ID) que corresponde al lugar del nucleótido que inicia la secuencia. En este punto es donde tenemos que hacer la selección de las secuencias validas.

```
#ID = 5808
Secuencia = MDPVDPNLEPWHPGSQPTTPCNKCYCKVC
-----

#ID = 6384
Secuencia = MLKHMIQKYIMSGRHMPVYPQTPHKKYTI
-----

#ID = 8787
Secuencia = MGGKWSKSSIVGWPQVRERMRQTPPAEGL
LKEKGGDLRLIYSRKRQEILDWVYNTQGYFPDQWNYTPGPG
LTHRAQELHPEFYKNC
-----

Introduce Identificador Para Analizar :
774
```

Figura 3. Petición de la entrada del identificador.

A continuación, el programa, nos pide que introduzcamos el identificador de la secuencia a analizar. Una vez hecho esto se nos muestra en pantalla el análisis realizado.

```
Longitudes:

Longitud De La Secuencia Problema (Nucleótidos): 1497
Longitud De La Secuencia Problema (Aminoácidos): 498
Inicio Y Final De La Secuencia Problema (Nucleótidos):
Marco De La Secuencia Problema: 3

Proporción De Nucleótidos Para La Secuencia Problema:

Proporción De Adenina : 36.47294589178357 %
Proporción De Uracilo : 19.03807615230461 %
Proporción De Guanina : 25.050100200400802 %
Proporción De Citosina : 19.43887775511022 %

Proporción De Nucleótidos Anómalos : 0.0 %

Proporción De Aminoácidos Para La Secuencia Problema :

S = 5.622489959839357 % R = 5.622489959839357 % L = 7.1
P = 6.626506024096385 % T = 5.220883534136546 % V = 4.1
D = 3.6144578313253013 % C = 2.0080321285140563 % Q = 1.1
K = 7.831325301204819 % F = 2.610441767068273 % Y = 1.1

Proporción De Aminoácidos Anómalos: 0.0 %

Masa Molecular ~ 55.286156640000016 kDa

¿Desea Analizar Otra Secuencia? S/N
```

Figura 4. Ejemplo de la salida del análisis del programa.

En el siguiente paso podemos analizar otra secuencia de la cadena proviral introducida o no. Si seleccionamos el no, podemos elegir analizar otro retrovirus.

```
¿Desea Analizar Otra Secuencia? S/N
n
¿Desea Analizar Otra Secuencia Proviral? S/N
```

Figura 5. Opciones para realizar nuevos análisis.

Los genomas provirales, usados para hacer las pruebas, han sido obtenidos de las bases de datos del NCBI (<https://www.ncbi.nlm.nih.gov/>).

Se recomienda usar este programa desde el software Eclipse [3]. Esta recomendación es para evitar ciertas limitaciones que puedan existir en relación a la entrada de las cadenas de texto, entre otros posibles contratiempos. Por otra parte esta sugerencia viene dada ya que este programa ha sido testeado desde dicho software.

Hay que indicar que este programa está en versión beta, por lo que de encontrar algún problema reportadlo y será solucionado lo antes posible. También no se dude en contactar si se tiene algún tipo de pregunta o problema a la hora de usarlo.

A modo de conclusión decir que este software es una herramienta que se puede utilizar para analizar retrovirus, con la finalidad de realizar trabajos de investigación, entre otras posibilidades. También remarcar que aunque el programa ha sido solamente probado en retrovirus, podría se utilizado para analizar secuencias biológicas de diferente procedencia.

Referencias

1. Roberto Reinoso. (2017, June 15). roberto117343/rstrdrtrvs 0.1. Zenodo. DOI: 10.5281/zenodo.808919
2. GNU General Public License. (2017, 17 de

mayo). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 02:08, junio 15, 2017 desde https://es.wikipedia.org/w/index.php?title=GNU_General_Public_License&oldid=99208527 .

3. Eclipse (software). (2017, 31 de mayo). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 02:09, junio 15, 2017 desde [https://es.wikipedia.org/w/index.php?title=Eclipse_\(software\)&oldid=99507966](https://es.wikipedia.org/w/index.php?title=Eclipse_(software)&oldid=99507966) .