

# ANOTACIÓN DE GENOMAS

### Anotación de Genomas

Este es un tutorial de cómo realizar la implementación de diferentes herramientas necesarias para lograr la anotación de genomas, se va utilizar el sistema operativo ubuntu ya que es un sistema amigable y fácil de usar aparte que hay herramientas que ya vienen preinstaladas y eso facilita el trabajo.

Brevemente, la anotación de genomas y genes es una labor que sirve para generar hipótesis y que serán aseguradas con el trabajo de laboratorio. En otras palabras es la identificación de repeticiones, además la predicción de genomas y genes que codifican para proteínas.

### Herramientas

Las herramientas ha utilizar son las siguientes:

- 1. Ubuntu / biolinux
- 2. Perl
- 3. CEGMA(pre-requisitos para instalar Cegma)
  - -Geneid ó Genewise
  - -Wise2
  - -Hmmer
  - -NCBI Blast+
- 4. Augustus(pre-requisitos para instalar Augustus)
   -Bowtie2
- 5. Maker (pre-requisitos para instalar Maker)
  - -SNAP
  - -RepeatMasker
  - -trf
  - -rmblast



Es un lenguaje de programación que puede hacer desde scripts para administrar servidores basados en unix o linux hasta proyectos completos realizados en este lenguaje como es el caso de varias herramientas que se verán en este contenido. Se realizará la actualización de Perl, en el caso de ubuntu ó biolinux ya trae pre-instalado el lenguaje Perl por lo que solo se necesitará actualizar los archivos, como se muestra en la imagen.

```
😰 🗖 🔲 vane@vane-VirtualBox: /etc/perl
To run a command as administrator (user "root"), use "sudo <command>".
<u>See "man</u> sudo root" for details.
vane@vane-VirtualBox:~$ cd /etc/perl
vane@vane-VirtualBox:/etc/perl$ ls
CPAN Net sitecustomize.pl
vane@vane-VirtualBox:/etc/perl$ sudo wget http://korflab.ucdavis.edu/Unix and Pe
rl/FAlite.pm
[sudo] password for vane:
--2016-12-10 17:33:08-- http://korflab.ucdavis.edu/Unix and Perl/FAlite.pm
Resolviendo korflab.ucdavis.edu (korflab.ucdavis.edu)... 128.120.143.145
Conectando con korflab.ucdavis.edu (korflab.ucdavis.edu)[128.120.143.145]:80...
conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 2770 (2,7K) [text/plain]
Guardando como: "FAlite.pm"
FAlite.pm
                   100%[================] 2,71K 9,74KB/s in 0,3s
2016-12-10 17:33:08 (9,74 KB/s) - "FAlite.pm" quardado [2770/2770]
vane@vane-VirtualBox:/etc/perl$
```



GeneWise es una herramienta que predice estructuras usando secuencias de proteinas similares. Es un algoritmo basado en los principios que utiliza los modelos ocultos de Markov, es una herramienta muy precisa y completa.



El comando **apt-get** es una herramienta para gestionar diferentes paquetes instalables, generalmente es un comando que se ejecuta como superusuario anteponiendo la palabra sudo para lograr ejecutar el programa como se muestra en la imagen. Con esta herramienta se inicia la instalación de Wise. De la misma manera se instalará la documentación del paquete Wise de la siguiente manera:

```
alle(wvalle-viicualbox.
vane@vane-VirtualBox:~$ sudo apt-get install wise-doc
Levendo lista de paquetes... Hecho
Creando árbol de dependencias
Levendo la información de estado... Hecho
Se instalarán los siguientes paquetes NUEVOS:
 wise-doc
0 actualizados, 1 nuevos se instalarán, 0 para eliminar y 142 no actualizados.
Se necesita descargar 817 kB de archivos.
Se utilizarán 1 317 kB de espacio de disco adicional después de esta operación.
Des:1 http://cr.archive.ubuntu.com/ubuntu yakkety/universe amd64 wise-doc all 2.
4.1-19 [817 kB]
Descargados 817 kB en 1s (425 kB/s)
Seleccionando el paquete wise-doc previamente no seleccionado.
(Levendo la base de datos ... 169547 ficheros o directorios instalados actualmen
te.)
Preparando para desempaquetar .../wise-doc 2.4.1-19 all.deb ...
Desempaquetando wise-doc (2.4.1-19) ...
Configurando wise-doc (2.4.1-19) ...
Procesando disparadores para doc-base (0.10.7) ...
Procesando 3 archivos doc-base añadidos...
vane@vane-VirtualBox:~S
```

Se descargará por medio del comando wget la herramienta wise. **Wget** lo que permite descargar ficheros desde internet de modo consola. Con el comando "**tar xvfz**" se desempaqueta las herramientas.

😣 🖨 🗉 🛛 vane@vane-VirtualBox: ~	😣 🖨 🗉 vane@vane-VirtualBox: ~
<pre>vane@vane-VirtualBox:~\$ wget http://korflab.ucdavis.edu/Datasets/cegma/wise2.2.3 -rc7.tar.gz</pre>	vane@vane-VirtualBox:~\$ tar xvfz wise2.2.3-rc7.tar.gz



Con "**cd**" se podrá navegar entre carpetas y "**pwd**" devuelve la ruta en la que esta situado.

Con "**nano ~/.bashrc**" se ingresa a editor de texto '**bash**' que es el intérprete de la mayoría de sistemas linux.



Con el comando "**nano** "**/.bashrc**" se mostrará la siguiente pantalla de texto donde se buscará el final del texto y se editará lo siguiente:

"export WISECONFIGDIR=/ubicacion/wise2.2.3-rc7/wisecfg/" como se ve en la imagen.

Hay que recordar que para saber la ubicación se utiliza el comando "pwd".

Una vez se guarde y se salga del editor de texto se editará el comando

". "/.bashrc" que servirá para actualizar sin necesidad de reiniciar el servidor.





Es un programa de predicción de genes y fue uno de los primeros programas que predicen estructuras exónicas completas de genes vertebrados. Geneid mantiene una estructura jerárquica(señal de exon a gen). Para descargar geneid se utilizará la siguiente dirección **wget ftp://** genome.crg.es/pub/software/geneid/geneid\_v1.4.4.Jan\_13\_2011.tar.gz y se colocará tal como se ve en la imagen.

😣 🗖 🗊 vane@vane-VirtualBox: ~	
<pre>vane@vane-VirtualBox:~\$ wget 4.Jan_13_2011.tar.gz</pre>	<pre>ftp://genome.crg.es/pub/software/geneid/geneid_v1.4</pre>

Seguidamente se debe descomprimir el paquete.



Una vez descomprimido el paquete de geneid se ingresará a la carpeta llamada geneid, y se utilizará el comando "**make**" que se encarga de leer todos los makefiles que son los que indican a make cúales y en que orden compilar los archivos fuentes de la herramienta.



Se moverá la terminal a la carpeta bin, y ahí se utilizará el comando "./ geneid –h" que prueba si la herramienta se compiló correctamente.



El comando "**sudo cp**" lo que va a realizar es la copia de todo lo que esta dentro de la carpeta bin en /usr/bin.



El **.profile** es un fichero de texto que el sistema operativo ejecuta de forma automática cuando se da una cierta condición. En el fondo lo que hace es mandar a bash a ejecutar archivos , hay que tener presente que solo hay una copia de .profile por lo que es muy importante tener cuidado con el contenido de este archivo. Este archivo se ejecutará cada vez que se inicie sesión de usuario.





Este es el editor .profile. La ruta que se deberá digitar será la siguiente: **PATH=\$PATH:~/src/geneid/bin/./** y export PATH, como se muestra en la imagen.

Para salir y guardar el archivo se podrá teclear control + C ó control + X dependiendo del computador.

Una vez se guarde y se salga del editor de texto se digitará el comando ". ~/.profile" que servirá para actualizar sin necesidad de reiniciar el servidor.





Blast2 es una herramienta interactiva que utiliza el motor BLAST para la comparación de secuencias con pares de ADN-AND o Proteinas-Proteinas y se basa en el mismo algoritmo de Smith-Waterman y estadisticas de alineación que Blast. Nuevamente se utilizará el comando "**wget**" para poder descargar los ficheros de blast, con la dirección que se ve a continuación:

😣 🗖 🗊 vane@vane-VirtualBox: ~	
<pre>vane@vane-VirtualBox:~\$ wget https://ftp.ncbi.nlm.nih.gov/blast/executables/blas t+/LATEST/ncbi-blast-2.5.0+-x64-linux.tar.gz</pre>	

Con el comando "tar xvfz" se desempaqueta el blast descargado.

🕒 🗉 🛛 vane@vane-VirtualBox: ~

vane@vane-VirtualBox:~\$ tar xvfz ncbi-blast-2.5.0+-x64-linux.tar.gz

Hay que navegar a la carpeta "**bin**" como se muestra en la imagen y se ingresará al editor de texto de ./profile.



Como se muestra es el mismo texto donde se colocó la dirección de geneid y seguidamente se deberá poner la dirección de blast y guardar el archivo.



Por último el comando ". ~/.profile" que como se mencionó anteriormente servirá para actualizar los archivos sin necesidad de reiniciar.

😕 🗖 🔲 vane@vane-VirtualBox: ~/ncbi-blast-2.5.0+/bin		
vane@vane-VirtualBox:~/ncbi-blast-2.5.0+/bin\$ . ~/.profile vane@vane-VirtualBox:~/ncbi-blast-2.5.0+/bin\$		



HMMER se utiliza para hacer alineaciones de secuencias y buscar bases de datos de secuencias homologas, basadas en modelos probabilisticos ocultos de Markov. Esta diseñado para detectar los homologos con la mayor sensibilidad posible. Se utilizará el comando "wget"mpara descargar el paquete de HMMER.



Se descomprime la herramienta con el comando "**tar xvfz**" y se pone el nombre del paquete HMMER.



Se navega a la carpeta hmmer y se utilizará el comando "./configure" que sirve para configurar las caracteristicas del sistema, además de crear archivos makefile con dependencias y opciones definidas.

	vane@vane-VirtualBox: ~/hmmer-3.1b2-linux-intel-x86_64	
vane@va vane@va	ne-VirtualBox:~\$ cd hmmer-3.1b2-linux-intel-x86_64/ ne-VirtualBox:~/hmmer-3.1b2-linux-intel-x86_64\$ ./configure	

Se utilizará el comando "make" para compilar, "make check" para la verificación y "make install" para la instalación del programa.



wane@vane-VirtualBox: ~/hmmer-3.1b2-linux-intel-x86\_64
vane@vane-VirtualBox:~/hmmer-3.1b2-linux-intel-x86\_64\$ sudo make install
[sudo] password for vane:

El primer comando "**pwd**" devuelve la ruta en la que esta situado. El segundo comando nos permite ingresar al archivo .profile para poder direccionar hmmer y el tercer comando actualizará el archivo sin necesidad de reiniciar el sistema, este ultimo se utilizará una vez y el archivo .profile a sido modificado

y guardado.



Se muestra como va colocado en el archivo .profile la direccion de hmmer.





CEGMA (Core Eukaryotic Genes Mapping Approach) es una herramienta que permite construir un conjunto inicial de anotaciones de genes fiables de genomas Eucariotas, incluye el uso de los modelos ocultos de Markov para garantizar la fiabilidad de la estructura de los genes. Se utilizará el comando "wget" para descargar el paquete de cegma y se descomprimirá con el comando "tar xvfz".

🔵 🔲 🛛 vane@vane-VirtualBox: ~ vane@vane-VirtualBox:~\$ wget http://korflab.ucdavis.edu/Datasets/cegma/cegma\_v2. 4.010312.tar.gz --2016-12-10 19:07:15-- http://korflab.ucdavis.edu/Datasets/cegma/cegma v2.4.01 0312.tar.qz Resolviendo korflab.ucdavis.edu (korflab.ucdavis.edu)... 128.120.143.145 Conectando con korflab.ucdavis.edu (korflab.ucdavis.edu)[128.120.143.145]:80... conectado. Petición HTTP enviada, esperando respuesta... 200 OK Longitud: 13352428 (13M) [application/x-tar] Guardando como: "cegma v2.4.010312.tar.gz" in 69s 2016-12-10 19:08:24 (190 KB/s) - "cegma\_v2.4.010312.tar.gz" guardado [13352428/1 3352428] vane@vane-VirtualBox:~\$

👅 🗉 🕛 🖉 vane@vane-virtualBox: ~

vane@vane-VirtualBox:~\$ tar xvfz cegma\_v2.4.010312.tar.gz

Se navegará a la carpeta cegma con el comando "cd" y se ejecutará el comando "make" para compilar los archivos.

```
😕 🗖 🔲 vane@vane-VirtualBox: ~/cegma_v2.4.010312
vane@vane-VirtualBox:~$ cd cegma v2.4.010312/
vane@vane-VirtualBox:~/cegma_v2.4.010312$ make
###
### RUNNING MAKEFILE
###
### sáb 10 dic 19:10:02 CST 2016 -- vane
###
mkdir ./bin;
### Finishing PERL script from "src/cegma.pl" -> "bin/cegma"
### Finishing PERL script from "src/genome map.pl" -> "bin/genome map"
### Finishing PERL script from "src/local map.pl" -> "bin/local map"
### Finishing PERL script from "src/parsewise.pl" -> "bin/parsewise"
### Finishing PERL script from "src/hmm_select.pl" -> "bin/hmm select"
### Finishing PERL script from "src/completeness.pl" -> "bin/completeness"
### Finishing PERL script from "src/geneid-train.pl" -> "bin/geneid-train"
### Finishing PERL script from "src/make_paramfile.pl" -> "bin/make paramfile"
###
### MAKEFILE DONE...
###
vane@vane-VirtualBox:~/cegma_v2.4.010312$
```

Se navega a la carpeta bin y se probará que cegma se logrará compilará bien con



Se va ingresar nuevamente al .profile para agregar la ruta de cegma como se muestra en la segunda imagen y exportarla, una vez guardada se actualiza el sistema con .~/.profile.





Primero se verificará con "pwd" la ruta en la que esta situado.

Seguido "**nano ~**/**.bashrc**" se ingresa al editor de texto y se verificará que el bash sea similar al que se muestra en la segunda imagen, sino fuera así deberá agregar los export que se muestran verificando en donde se encuentra ubicado cada una de las herramientas en su sistema.

Por último, se guardará el archivo y se actualizará con ".~/.bashrc".

Una vez reiniciado el sistema se ingresará a la carpeta "sample" para realizar una pequeña prueba, se utilizará el siguiente comando:



Esta prueba dura un poco asi que se debe tener paciencia, se verá al finalizar la corrida algo muy similar a la imagen.

MAPPING PROTEINS TO GENOME (TELASTN)
RUNNING: genome_map -n genome -p 6 -o 5000 -c 2000 -t 16 /path/to/CEGMA/sample.prot /path/to/CEGMA/sample.sample.dna 2>output.cegma.errors
Building a new DB, current time: 04/30/2012 11:42:26 New DB name: /tmp/genome86278.blaatdb Sequence type: Nucleotide Sequence type: Nucleotide Keep MBits T Maximum file size: 1073741824B Adding sequences from FASTA; added 1 sequences in 0.0666069 seconds. Found 86 candidate regions in /path/to/CEGMA/sample/sample.dna
•• NAKING INITIAL GENE PREDICTIONS FOR CORE GENES (GENEWISE + GENEID) ••
RUNNING: local_map -n local -f -h /path/to/CEGMA/data/hmm_profiles -i KOG genome.chunks.fa 2>output.cegma.errors NOTE: created 23 geneid predictions
FILTERING INITIAL PROTEINS PRODUCED BY GENEID (HMMER)
RUNNING: hmm_select -i KOG -o local -t 16 /path/to/CEGNA/data/hmm_profiles local.geneid.fa /path/to/CEGNA/data/profiles_cutoff.tbl 2>output.cegma.errors NOTE: Found 15 geneid predictions with scores above threshold value
CALCULATING GENEID PARAMETERS FROM SELECTED GENEID PREDICTIONS
RUNNING: geneid-train local.geneid.selected.gff local.geneid.selected.dna geneid_params 2>output.cegma.errors DAT& COLLECTED: 15 Coding sequences containing 48 introns RUNNING: make_paramfile /path/of/SCRM/data/self.param.template \ geneid_params/coding.initial.s.logs geneid_params/coding.transition.5.logs \ geneid_params/startlogs geneid_params/acclogs geneid_params/don.logs \ geneid_params/intron.max > geneid_params/self.param
•• ACCURATE LOCAL MAPPING ••
RUNNING: local_map -n local_self -g local.genewise.gff -d geneid_params/self.param -h /path/to/CEGMA/data/hmm_profiles -i KOG genome.chunks.fa 2>output.cegma.errors NOTE: Will use specifed local.genewise.gff file instead of running genewise NOTE: created 21 geneid predictions
FINAL FILTERING
RUNNING: hmm_select -i KOG -o local_self -t 16 /path/to/CEGMA/data/hmm_profiles local_self.geneid.fa /path/to/CEGMA/data/profiles_cutoff.tbl 2>output.cegma.errors NOTE: Foud 15 geneid predictions with scores above threshold value
CONVERTING LOCAL COORDINATES INTO GENORE-WIDE COORDINATES
•• EVALUATING RESULTS AND COMPARING TO SET OF 248 HIGHLY CONSERVED CEGS ••
RUNNING: completeness local self.hmm select.aln /path/to/CEGMA/data/completeness cutoff.tbl > output.completeness report.

### Nota

Hay que tomar en cuenta que dependiendo de las versiones podría generar errores en el momento de querer usar cegma.

Un error común es "**geneid-train did not work properly**" cuando se intenta correr la prueba y se podría solucionar de una manera muy sencilla:

Ingresamos en la carpeta cegma – folder lib - en los archivos geneid.pm en la línea 108 y HMMStar.pm en la línea 41 y se sustituirá el siguiente código:

Antes: \$code .= "\$tab foreach my \\$c\$i qw(A C G T) { \n";

Después: \$code .= "\$tab foreach my \\$c\$i (qw(A C G T)) { \n";

Guardan los archivos reinician sistema y vuelven a intentar la prueba y debería funcionar sin problema.



## Augustus [gene prediction]

AUGUSTUS es un programa que predice genes en secuencias eucarióticas genómicas. Augustus esta creado para predecir genes en genomas de especies novedosas. También le permite predecir genes en una secuencia del genoma con parámetros ya entrenados.

Para instalar Augustus es importante tener numpy instalado, es un paquete fundamental para el trabajo de computación científica con python.

En el primer comando se verifica si esta instalado, en mi caso no lo tenía instalado, por lo que se agregará con el comando "sudo apt-get install



La subversion es una herramienta mas conocida como svn diseñado para mantener bajo control la vesión, documentación y código fuente de una aplicación. Y se lográ instalar con el comando " sudo apt-get install

subversion".

🛚 🗖 🔲 vane@vane-VirtualBox: ~

vane@vane-VirtualBox:~\$ svn
El programa «svn» no está instalado. Puede instalarlo escribiendo:
sudo apt install subversion
vane@vane-VirtualBox:~\$ sudo apt-get install subversion

Se podrá descargar Augustus desde el link mostrado en la imagen con el comando"wget" una vez descargado se descomprimirá con el comando "tar –xzf".

#### 😣 🗖 🔲 🛛 vane@vane-VirtualBox: ~

vane@vane-VirtualBox:~\$ sudo wget https://storage.googleapis.com/google-code-arc hive-downloads/v2/code.google.com/augustus/augustus-0.5.3.0.tar.gz --2016-12-11 17:19:49-- https://storage.googleapis.com/google-code-archive-down loads/v2/code.google.com/augustus/augustus-0.5.3.0.tar.gz Resolviendo storage.googleapis.com (storage.googleapis.com)... 74.125.134.128 Conectando con storage.googleapis.com (storage.googleapis.com)[74.125.134.128]:4 43... conectado. Petición HTTP enviada, esperando respuesta... 200 OK Longitud: 5192530 (5,0M) [application/octet-stream] Guardando como: "augustus-0.5.3.0.tar.gz" augustus-0.5.3.0.ta 100%[===========] 4,95M 69,6KB/s in 40s 2016-12-11 17:20:31 (128 KB/s) - "augustus-0.5.3.0.tar.gz" guardado [5192530/519 2530]

vane@vane-VirtualBox:~\$ tar -xzf augustus.2.5.5.tar.gz

Ahora, hay que navegar a la carpeta augustus.2.5.5/src (tomando en cuenta que esa es la versión que yo utilice).Se inicia la compilación de la herramienta con el comando "**make**".

😣 🗖 🗊 vane@vane-V	irtualBox: ~/augustus.2.5	.5/src
vane@vane-VirtualBo vane@vane-VirtualBo	x:~/augustus.2.5.5\$ x:~/augustus.2.5.5/s	cd src/ rc\$ ls
augustus.cc	extrinsicinfo.cc	merkmal.cc
baumwelch.cc	fastBlockSearch.cc	motif.cc
cflags	genbank.cc	namgene.cc
commontrain.cc	gene.cc	pp_fastBlockSearcher.cc
consensus.cc	geneticcode.cc	pp_hitseq.cc
consensusFinder.cc	hints.cc	pp_prepare_align.cc
contentmodel.cc	igenicmodel.cc	pp_profile.cc
curve2hints.cc	igenictrain.cc	pp_scoring.cc
dummy.cc	intronmodel.cc	projectio.cc
etraining.cc	introntrain.cc	properties.cc
evaluate.cc	licence.txt	statemodel.cc
evaluation.cc	LICENCE.TXT	types.cc
exonmodel.cc	lldouble.cc	utrmodel.cc
exon_seg.cc	logdoubletest.cc	utrtrain.cc
exon_seg.orig.cc	makedepend.pl	vitmatrix.cc
exontrain.cc	Makefile	
vane@vane-VirtualBo	x:~/augustus.2.5.5/s	rc\$ make

Se ingresará al bash recordando el comando "**nano ~/.bashrc**" y se cambiará la variable de entorno con la dirección en donde se encuentra augustus. Por último se actualizará el sistema con ". **~/.bashrc**".





SNAP es un nuevo alineador de lecturas cortas y largas más precisa, es decir con menos errores y es 10-100x mas rápido que herramientas de última generación.

Primero se descargará el paquete de la web en la siguiente dirección :

#### <u>https://drive.google.com/open?id=0B-</u> DuA1JwraQuVEQyV19UZHNpRXc

Navegar hasta la carpeta donde esta SNAP. Hay que recordar que si necesitará descomprimir el paquete hay que usar el siguiente comando.

```
vane@vane-VirtualBox: ~
                                 WUSCZIZIJTU
vane@vane-VirtualBox:~$ tar xzvf snap-2013-11-29.tar.gz
snap/
snap/00README
snap/. cds-trainer.pl
snap/cds-trainer.pl
snap/depend
snap/DNA/
snap/example.zff
snap/._exonpairs.c
snap/exonpairs.c
snap/. fathom.c
snap/fathom.c
snap/. forge.c
snap/forge.c
snap/HMM/
snap/hmm-assembler.pl
snap/. hmm-info.c
snap/hmm-info.c
snap/. LICENSE
snap/LICENSE
snap/. Makefile
snap/Makefile
snap/._noncoding-trainer.pl
```

# Una vez descargado y descomprimido se navegará a la carpeta snap, y compilaremos la herramienta con "make".

😣 🗖 🗊 vane@va	ne-VirtualBox: ~/snap		
vane@vane-Virtua vane@vane-Virtua 00README cds-trainer.pl	alBox:~\$ cd snap/ alBox:~/snap\$ ls exonpairs.c fathom.c	hmm-info.c LICENSE	snap.c zff2gff3.pl
depend DNA example.zff vane@vane-Virtu	forge.c HMM hmm-assembler.pl alBox:~/snapS make	Makefile noncoding-trainer.pl patch-hmm.pl	Zoe

Ahora se ingresará al bash, recordando el comando "**nano ~/.bashrc**" y se cambiará la variable de entorno "**PATH**" como se ve en la imagen y se guarda los cambios.

Por último se actualizará el sistema con ". ~/.bashrc".

80	vane@var	ne-VirtualBox: ~	/snap			
nano	2.6.3	Ar	chivo: /home/	/vane/.bashrc		Modificado
export export export export export export export	WISECONFI CEGMA=/ho PERL5LIB= CEGMATMP= WISE2_PAT PATH="/ho AUGUSTUS_ ZOE=/home	GDIR=/home/v me/vane/cegm /home/vane/c /home/vane/c H=/usr/bin/ me/vane/augu CONFIG_PATH= /vane/snap/Z	ane/wise2.2.3 a_v2.4.010312 egma_v2.4.010 egma_v2.4.010 stus.2.5.5/bi /home/vane/au oe	3-rc7/wisecfg )312/lib )312/ .n/:/home/vane Igustus.2.5.5/	e/snap: <mark>\$</mark> PATH" 'config/	
<mark>^G</mark> Ver <mark>^X</mark> Sali	ayuda <mark>^0</mark> ir <mark>^R</mark>	Guardar ^W Leer fich.^\	Buscar <mark>^</mark> Reemplazar <mark>^</mark> l	Cortar Tex <mark>^3</mark> Pegar txt <mark>^1</mark>	Justificar <mark>^C</mark> Ortografía <mark>^</mark>	Posición Ir a línea







### RMBlast

Es una versión compatible con RepeatMasker y es de la familia de NCBI BLAST. Incluye soporte para matrices personalizadas, el algoritmo de búsqueda de Smith-Waterman.

### **Tandem Repeats Finder TRF**

Es un programa para localizar y mostrar repeticiones tandem (una secuencia de dos o más contiguos) en secuencias de ADN. Para utilizar el programa, el usuario envía una secuencia en formato FASTA. La salida consta de dos archivos: un archivo de tabla de repetición y un archivo de alineación. El programa es muy rápido, analizando secuencias del orden de 5Mb en unos segundos.

### Maker

Es un pipeline de anotacion de genoma portable y fácil de configurar. Su propósito es permitir que los proyectos de genoma eucariotas y procariotas más pequeños anoten sus genomas de forma independiente y crear bases de datos de genomas. MAKER debe resultar especialmente útil para proyectos de organismos modelo con una experiencia mínima en bioinformática y recursos informáticos.

### RepeatMasker

Es un programa que detecta secuencias de ADN para repeticiones intercaladas y secuencias de ADN de baja complejidad. La salida del programa es una anotación detallada de las repeticiones que están presentes en la secuencia de consulta, así como una versión modificada de la secuencia de consulta en la que todas las repeticiones anotadas han sido enmascaradas.

Primero se descargará los paquete de la web en el siguiente vínculos ó las paginas oficiales de las herramientas.

<u>https://drive.google.com/open?id=0B-</u> DuA1JwraQueHo4NEhVOWZneUk

Si necesitará descomprimir los paquetes habría que realizarlo de la siguiente manera:

😣 🗖 🗉 vane@vane-VirtualBox: ~
<pre>vane@vane-VirtualBox:~\$ tar xzvf rmblastn-2.2.28-x64.tar.gz</pre>
😰 🗖 🗉 vane@vane-VirtualBox: ~
vane@vane-VirtualBox:~\$ tar xzvf RepeatMasker.tar.gz
😣 🗖 🔲 vane@vane-VirtualBox: ~
vane@vane-VirtualBox:~\$ tar xvfz maker-2.31.8.tgz Archivos

El primer paquete que se va a instalar va ser Maker, para eso se navegará a la carpeta **maker/src** y se ejecutará un comando con perl generalmente instalado en ubuntu por defecto, que revisará que todas las versiones y requerimientos se encuentren correctamente.En este caso se generó una advertencia de "build

installdeps". 😣 亘 🔲 vane@vane-VirtualBox: ~/maker/src vane@vane-VirtualBox:~\$ cd maker/ vane@vane-VirtualBox:~/maker\$ ls bin data GMOD INSTALL lib LICENSE MWAS perl README RELEASE src vane@vane-VirtualBox:~/maker\$ cd src/ vane@vane-VirtualBox:~/maker/src\$ ls bin \_build inc locations MYMETA.json version blib Build.PL lib MANIFEST MYMETA.yml vane@vane-VirtualBox:~/maker/src\$ perl Build.PL Checking prerequisites... requires: ! IO::All is not installed ! Inline::C is not installed ! Perl::Unsafe::Signals is not installed ! Want is not installed ! forks is not installed ! forks::shared is not installed ERRORS/WARNINGS FOUND IN PREREQUISITES. You may wish to install the versions of the modules indicated above before proceeding with this installation Run 'Build installdeps' to install missing prerequisites.

Se configurará el paquete pendiente con el comando "sudo ./build installdeps" como se ve en la imagen.



Ahora se deberá instalar Maker con el comando "./Build install", sin problemas.



También Maker se ingresará al bash con el comando "**nano ~/.bashrc**" y se modificará la variable de entorno "**PATH**" como se ve en la imagen y se guarda los cambios. Por último se actualizará el sistema con "**. ~/.bashrc**".



Se navegará a la carpeta RepeatMasker, una vez se este en la carpeta se colocará el comando "perl ./configure" y debería ver la siguiente imagen. Generalmente el sistema por determinado pone la dirección donde se encuentra RepeatMasker, sino



Se va instalar un programa Ilamado Tandem Repeat Finder o mas conocido con sus siglas TRF. Se recomienda ubicar el archivo TRF dentro de la carpeta de RepeatMasker para mayor facilidad de ubicación. Una vez ingresada la ubicación de la herramienta seguimos con enter.



Ahora se va desplegar un menu y se eligira la opción 2.



De igual manera hay que decirle a la configuración de RepeatMasker la ubicación donde se encuentra el RMBlast, como se muestra en la imagen.



Lo siguiente en el menú es la opción 5 para verificar que la configuración se realizó correctamente. Se mostrará la siguiente leyenda mostrada en la imagen.

🔕 🖻 🗉 🛛 vane@vane-VirtualBox: ~/RepeatMasker
<ol> <li>CrossMatch: [ Un-configured ]</li> <li>RMBlast - NCBI Blast with RepeatMasker extensions: [ Configured, Default</li> <li>WUBlast/ABBlast (required by DupMasker): [ Un-configured ]</li> <li>HMMER3.1 &amp; DFAM: [ Un-configured ]</li> </ol>
5. Done
Enter Selection: 5 Setting perl interpreter
Congratulations! RepeatMasker is now ready to use. The program is installed with a minimal repeat library by default. This library only contains simple, low-complexity, and common artefact ( contaminate ) sequences. These are adequate for use with your own custom repeat library. If you plan to search using common species specific repeats you will need to obtain the complete RepeatMasker repeat library from GIRI ( www.giriinst.org ) and install it in /home/vane/RepeatMasker.
Further documentation on the program may be found here: /home/vane/RepeatMasker/repeatmasker.help
vane@vane-VirtualBox:~/RepeatMaskerS

### Todo salio bien!