



I Curso de Bioinformática de la Estación Experimental de La Mayora

sol genomics network

Boyce Thompson Institute for Plant Research
Tower Road
Ithaca, New York 14853-1801
U.S.A.

by

Aureliano Bombarely Gomez



MATERIAL DEL CURSO:

***ftp://ftp.solgenomics.net/user_requests/
aubombarely/courses/BioinfoLaMayora/***



PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Ux y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Unix y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Casos prácticos del bloque de base de datos y anotación.



ubuntu





Cargar la máquina virtual Ubuntu en el programa VMware Player.

User = user

Password = password





Ubuntu

Suspend Take Snapshot Settings Unity Full Screen

ubuntu

Bioinformatics

Password:

Cancel Login

English (United States) USA Ubuntu Mon 4:43 PM

To direct input to this virtual machine, click inside the window.





The screenshot shows an Ubuntu desktop environment. At the top, there is a system menu with icons for Suspend, Take Snapshot, and Settings. The top right corner displays 'Unity Full Screen' and the user name 'user' with a power button icon. The time is 4:44 PM. A tooltip above the file manager window reads 'Access documents, folders and network places'. The file manager window displays a grid of files and folders:

- blast2go.jnlp
- EJERCICIO_01-BusquedaGenes
- EJERCICIO_01-Soluciones
- EJERCICIO_02-BusquedaMarcadores_genDR4
- EJERCICIO_02-Soluciones
- EJERCICIO_03-BusquedaGenes_sequia
- EJERCICIO_03-Material_y_Soluciones
- EJERCICIO_04-MapeadoSondas_SGN-Unigenes
- EJERCICIO_04-Material_y_Soluciones
- EJERCICIO_05-Material_y_Soluciones
- EJERCICIO_05-AnotacionSondasExpresadasDif
- Presentaciones

At the bottom of the window, there is a status bar with various system icons and a message: 'To direct input to this virtual machine, click inside the window.'





Caso práctico: Ejercicio_01:

Búsqueda de secuencias relacionadas con sequía en tres especies modelo: *Arabidosis thaliana*, *Oryza sativa* y *Solanum lycopersicum*

Caso práctico: Ejercicio_02:

Búsqueda de marcadores para el gen DRS1, un gen relacionado con sequia: DRS1.

Caso práctico: Ejercicio_05:

Anotacion de secuencias utilizando Blast2Go.



PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Ux y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Terminal Linux-Unix. Líneas de Comandos





Terminal Linux-Unix y Líneas de Comandos.

1. ¿ Qué es Linux, qué es Unix?
2. El Terminal.
3. Comandos más frecuentes.
4. Aplicación con archivos `.fasta`, `gff3` y `.tab`





Terminal Linux-Unix y Líneas de Comandos.

1. ¿ Qué es Linux, qué es Unix?
2. El Terminal.
3. Comandos más frecuentes.
4. Aplicación con archivos .fasta, gff3 y .tab





Linux es un [núcleo](#) de [sistema operativo libre](#) tipo [Unix](#). Es uno de los principales ejemplos de [software libre](#). Linux está licenciado bajo la [GPL v2](#) y está desarrollado por colaboradores de todo el mundo. El desarrollo del día a día tiene lugar en la *Linux Kernel Mailing List*.

El núcleo Linux fue concebido por el entonces estudiante de [ciencias de la computación finlandés](#), [Linus Torvalds](#), en 1991. Linux consiguió rápidamente desarrolladores y usuarios que adoptaron códigos de otros proyectos de software libre para su uso en el nuevo sistema operativo. El núcleo Linux ha recibido contribuciones de miles de programadores.

Normalmente Linux se utiliza junto a un empaquetado de software, llamado [distribución Linux](#). (Fuente: Wikipedia).

Ubuntu es una distribución de Linux.





Unix (registrado oficialmente como UNIX®) es un **sistema operativo** portable, multitarea y multiusuario; desarrollado, en principio, en 1969 por un grupo de empleados de los laboratorios Bell de AT&T, entre los que figuran Ken Thompson, Dennis Ritchie y Douglas McIlroy.

Las implementaciones más conocidas son:

- [Solaris](#) de [Sun Microsystems](#).
- [AIX](#) de [IBM](#).
- [Mac OS X](#) de Apple.
- [Red Hat Enterprise Linux](#). Apoya el [proyecto Fedora](#) del cual se beneficia y de ella se derivan distribuciones compatibles como [Oracle Enterprise Linux](#) y [CentOS](#), también distribuciones como [Mandriva Linux](#), se basó en una de sus primeras versiones.
- [SUSE Linux](#) de [Novell](#).
- [Debian GNU/Linux](#). Con una de las comunidades más grandes y antiguas del movimiento de [software libre](#), es base para distribuciones como [Xandros](#), [Mepis](#), [Linspire](#) y [Ubuntu](#).



Terminal Linux-Unix y Líneas de Comandos.

1. ¿ Qué es Linux, qué es Unix?

2. El Terminal.

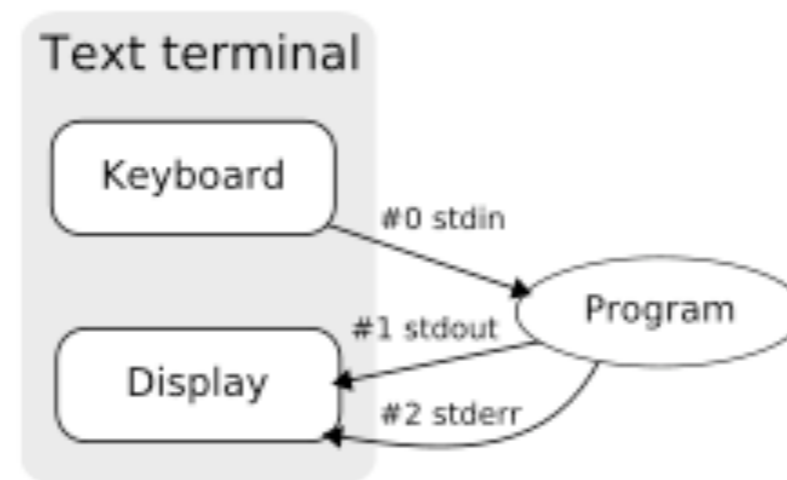
3. Comandos más frecuentes.

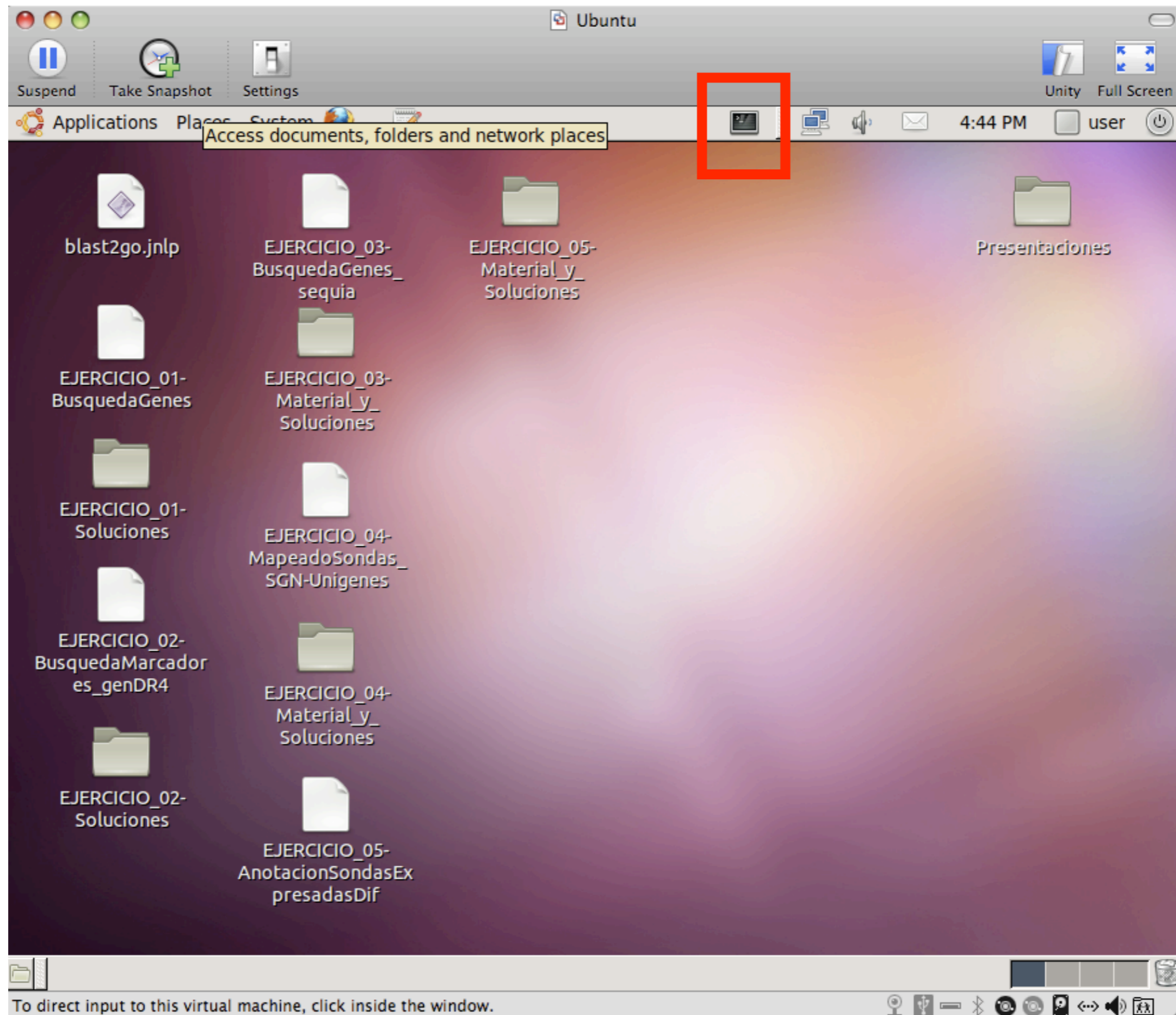
4. Aplicación con archivos .fasta, gff3 y .tab

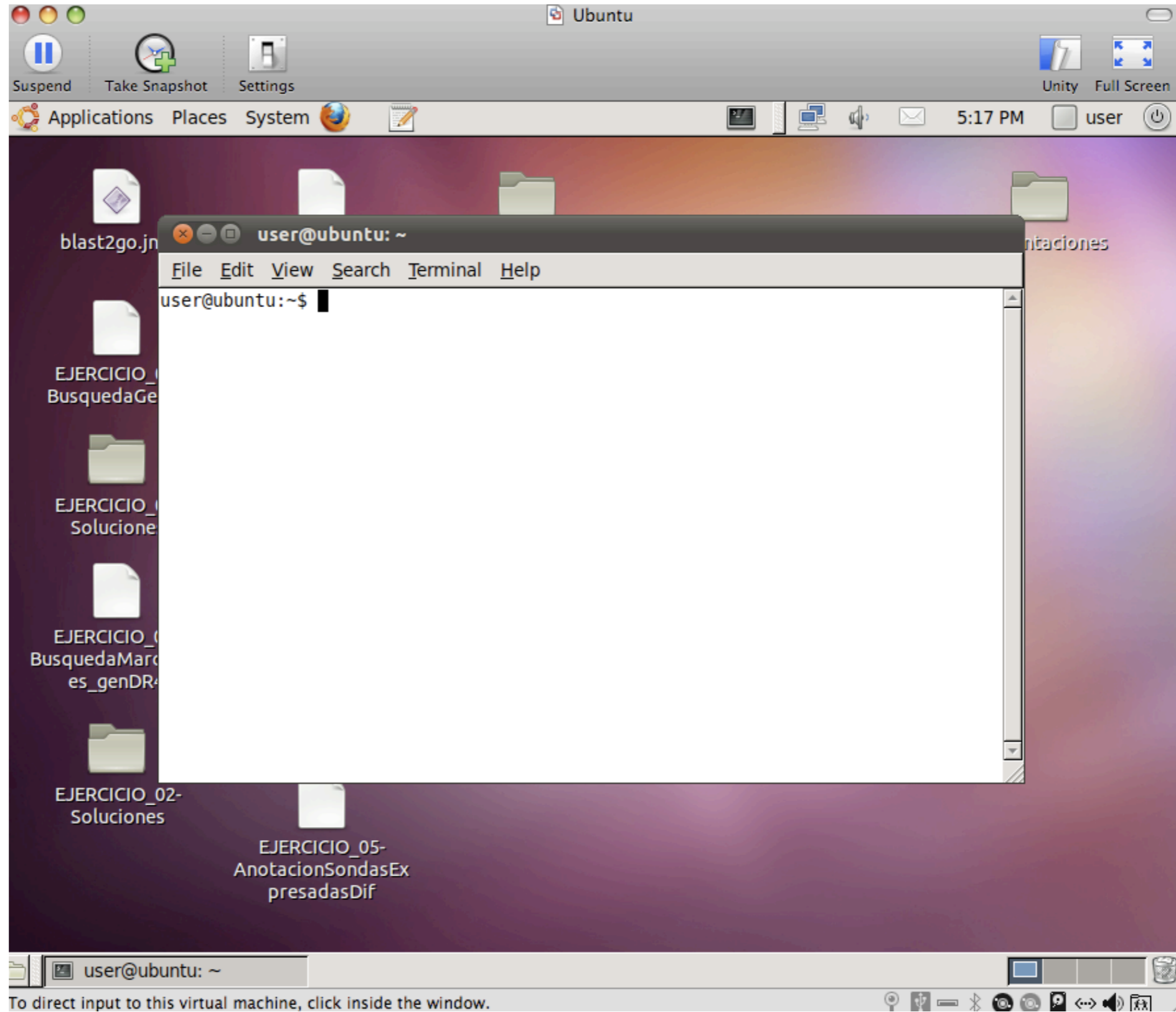




La terminal, consola de comandos o simplemente **consola** es la interface para comunicarse con el ordenador, con un modo de entrada de datos u órdenes y un modo de salida o visualización de texto (wikipedia).



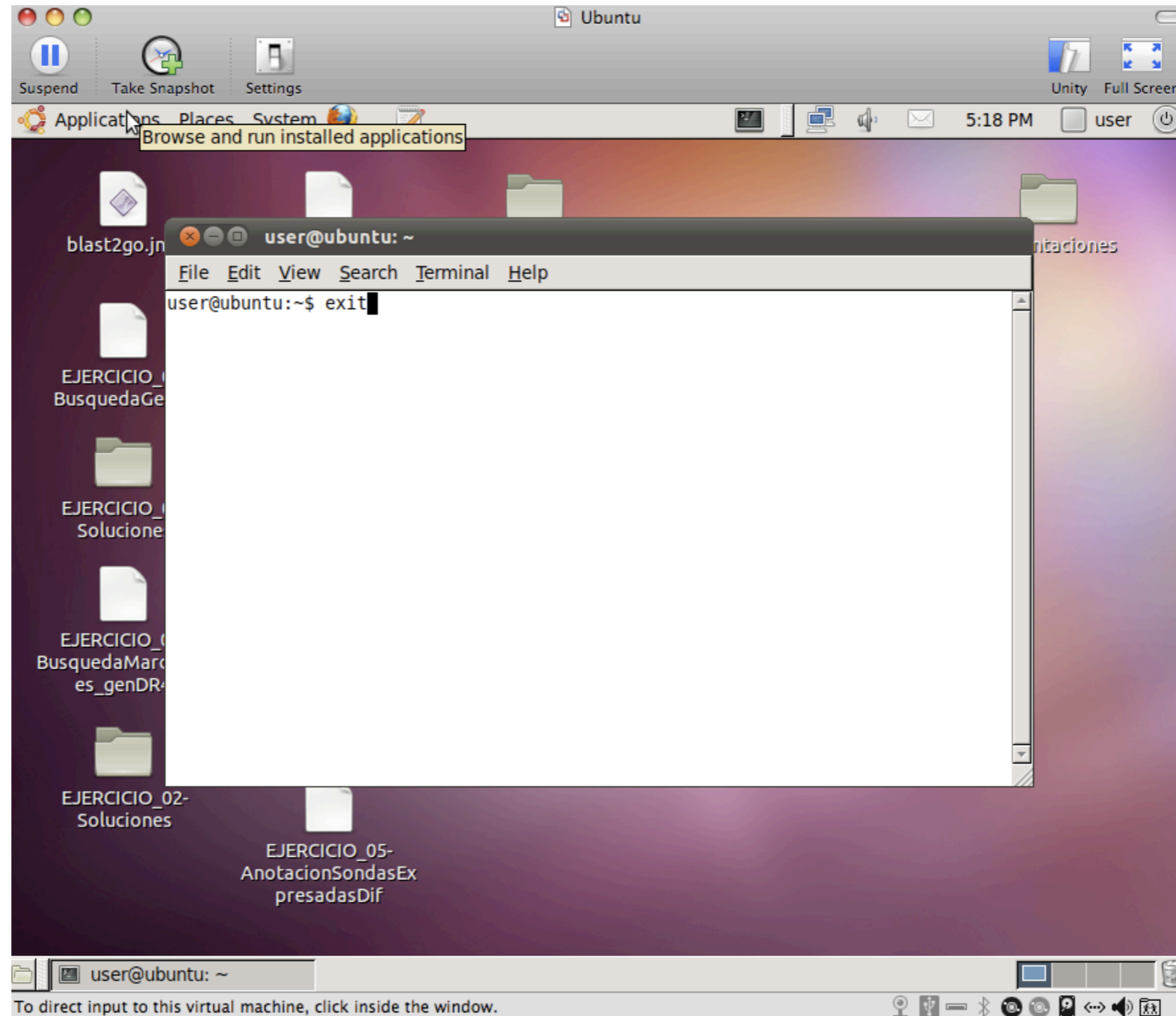






Para ejecutar cualquier orden o comando, siempre hay que pulsar **RETURN** (**ENTER**) tras el comando.

Para salir del terminal: **exit**





Terminal Linux-Unix y Líneas de Comandos.

1. ¿ Qué es Linux, qué es Unix?
2. El Terminal.
3. Comandos más frecuentes.
4. Aplicación con archivos .fasta, gff3 y .tab





Comandos para conocer, manipular o moverse a través de los directorios de archivos.

Commando	Función	Ejemplo:
pwd	imprimir el directorio de trabajo	pwd
cd	cambiar de directorio de trabajo	cd /home/user/Desktop
ls	imprimir un listado de archivos y directorios	ls -lh /home/user/Desktop

Comandos para obtener información de otro comando

Commando	Función	Ejemplo:
man comando	para salir pulsar "q"	man pwd
comando --help	imprime ayuda	ls --help



Comandos para manipular archivos

Commando	Función	Ejemplo:
touch	crea un archivo vacío	touch test.txt
rm	elimina un archivo	rm test.txt
cat	imprimir el contenido de un archivo	cat EJERCICIO_01-BusquedaGenes
head	imprimir las primeras líneas de un archivo	head -n 10 EJERCICIO_01-BusquedaGenes
tail	imprimir las últimas líneas de un archivo	tail -n 10 EJERCICIO_01-BusquedaGenes
less	abre un archivo para visualizado (“q” para salir)	less EJERCICIO_01-BusquedaGenes
cp	copia un archivo	cp test.txt test2.txt
mv	mueve un archivo	cp test2.txt ../

Comandos para manipular directorios

Commando	Función	Ejemplo:
mkdir	crea un directorio	mkdir testdir
rmdir	elimina un directorio vacío	rmdir testdir
rm -Rf	elimina un directorio y todo lo que contiene	rm -Rf testdir
cp -R	copia un directorio y todo lo que contiene	cp -R testdir



Comandos para imprimir selectivamente el contenido de un archivo

Commando	Función	Ejemplo:
cut	corta e imprime columnas de un archivo	cut -f1,5 test.tab
sed	reemplaza e imprime contenido de un archivo	sed 's/Oryza/Patata/' file
diff	compara archivos e imprime la diferencia	diff test1.txt test2.txt
sort	ordena por fila e imprime	row test1.tab
grep	busca e imprime las lineas coincidentes	grep 'thaliana' file.tab

Comandos para visualizar procesos

Commando	Función	Ejemplo:
top	para salir pulsar "q"	top



Comandos para volcar (bajar) archivo con una URL

Commando	Ejemplo:	Función
wget	para bajar datos de una dirección URL	wget ftp://ftp.solgenomics.net/unigene_builds/README.unigene

Comandos para trabajar con bloques

Commando	Función	Ejemplo:
;	Ejecuta varios comandos secuencialmente	cd; ls;
&&	Ejecuta varios comandos solo si el anterior fue exitoso	ls dir && mv file dir/
	Ejecuta varios comandos solo si el anterior falló	ls dir mkdir dir
	Usa la salida del comando anterior como entrada en el segundo	ls sort
>	Crea un archivo con la salida del archivo anterior	ls > dir_list.txt
>>	Añade la salida anterior al archivo especificado	ls >> dir_list.txt
<	Usa el archivo como entrada de datos	echo < dir_list.txt



Comandos para trabajar con permisos

Commando	Función	Ejemplo:
ls -l	imprime el listado de archivos con permisos	ls -l
chmod	Cambia los permisos del archivo/directorio	ls dir && mv file dir/
chgrp	Cambia el usuario y grupo del archivo/directorio	ls dir mkdir dir

r => 4 => readable, puede leerse
w => 2 => writable, puede escribirse
x => 1 => executable, puede ejecutarse

10 columnas donde la primera indica si es o no un directorio (d/-), los siguientes tres si es r,w,d para el propietario, las siguientes tres para el grupo y las últimas tres para todos los usuarios.





Terminal Linux-Unix y Líneas de Comandos.

1. ¿ Qué es Linux, qué es Unix?
2. El Terminal.
3. Comandos más frecuentes.
4. Aplicación con archivos `.fasta`, `gff3` y `.tab`





Conjunto de comandos para manipular archivos

Commando	Función
<code>grep '>' file cut -d ' ' -f1 sed 's/>/'</code>	selecciona IDs en un archivo fasta
<code>awk '{ print \$2"\t"\$1}' file</code>	reordena columnas (2,1)
<code>awk '{ if (\$7 > 80) print \$0}' file</code>	imprime lineas donde la septima columna es mayor de 80
<code>cat file1 file2 > file3</code>	crea un nuevo archivo con los archivos 1 y 2
<code>grep -v '#' file.gff3 cut -f1,3</code>	ignora lineas con '#' y corta las primera y tercera columna
<code>cut -f1 file sort -u wc -l</code>	corta la primera columna, ordena unicos y cuenta numero de lineas



Caso práctico: Ejercicio_03:

Búsqueda de genes expresados diferencialmente en un experimento de sequía.



PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Uinx y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Bases de datos.

1. ¿Qué es una base de datos?
2. Sistemas gestores de bases de datos.
3. Comandos SQL más frecuentes.





Bases de datos.

1. ¿ Qué es una base de datos?
2. Sistemas gestores de bases de datos.
3. Comandos SQL más frecuentes.





Una **base de datos** o banco de datos (en ocasiones abreviada con la sigla BD o con la abreviatura b. d.) es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

Existen programas denominados **sistemas gestores de bases de datos**, abreviado SGBD, que permiten almacenar y posteriormente acceder a los datos de forma rápida y estructurada.

Una **base de datos relacional** es una base de datos que cumple con el **modelo relacional** (En este modelo todos los datos son almacenados en relaciones, y como cada relación es un conjunto de datos, el orden en el que estos se almacenen no tiene relevancia). Cada relación se utiliza como si fuese una tabla que está compuesta por registros (cada fila de la tabla sería un registro o tupla), y columnas (también llamadas campos).

El lenguaje más habitual para construir las consultas a bases de datos relacionales es **SQL, Structured Query Language o Lenguaje Estructurado de Consultas**, un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales.





Bases de datos.

1. ¿Qué es una base de datos?
- 2. Sistemas gestores de bases de datos.**
3. Comandos SQL más frecuentes.





Los **sistemas gestores de bases de datos**, más usados son:

❖ SGBD libres:

- Postgres. <http://www.postgresql.org/>
- MySQL. <http://dev.mysql.com/>
- SQLite. <http://www.sqlite.org/>

❖ SGBD no libres.

- Oracle. <http://www.oracle.com/index.html>
- Microsoft Access. <http://office.microsoft.com/en-us/access/>



Bases de datos.

1. ¿Qué es una base de datos?
2. Sistemas gestores de bases de datos.
3. Comandos SQL más frecuentes.





Los comandos SQL más usados son:

❖ Búsqueda:

- SELECT cols FROM table WHERE condition.
- SELECT columns INTO TABLE table FROM table WHERE condition.
- SELECT columns FROM table1 JOIN table2 ON table1.col1=table2.col2

❖ Creación de tablas:

- CREATE TABLE table1 (colname datatype).

❖ Insertar, modificar, eliminar o copiar datos.

- INSERT INTO table VALUES (col1, col2...)
- UPDATE table SET column = newvalue WHERE column = oldvalue (or ID)
- DELETE FROM table WHERE column = value
- COPY table FROM file.
- COPY table TO file

❖ Eliminar tabla

- DROP TABLE table



Más información:

Postgres manual: <http://www.postgresql.org/docs/8.4/static/index.html>

W3school Tutorial: <http://www.w3schools.com/sql/default.asp>



Caso práctico: Ejercicio_04: Mapeado de datos. Sondas y Unigenes.





PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Unix y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.

