

Análisis de secuencias

CLC Free Workbench

Introducción

CLC free workbench es una herramienta bioinformática para el análisis de secuencias de gran utilidad, versátil y de código libre. Por lo menos así lo es para la versión gratuita, pues existen diferentes versiones de este programa para cumplir diferentes necesidades bioinformáticas.

Un vistazo a las versiones de CLC workbench

Visite el siguiente enlace:

<http://www.clcbio.com/index.php?id=89>

¿Cuántas versiones de CLC workbench existen, que diferencias fundamentales hay entre ellas? ¿Para qué tipo de análisis cree usted que es más indicada una u otra versión de este programa?

En esta oportunidad trabajaremos con la versión FREE del programa, la cual puede ser descargada de la siguiente dirección: <http://www.clcbio.com/index.php?id=479>

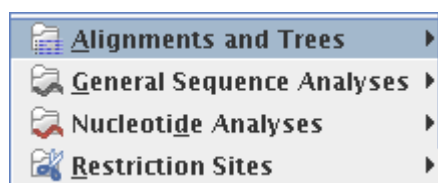
En esta oportunidad no descargaremos el software del sitio oficial sino del servidor de bioinformática del IBUN.

Descargue el programa de la siguiente dirección:

http://bioinf.ibun.unal.edu.co/cbib/estudiantes/tmp/CLCFreeWorkbench_4_0_1_JRE.exe

Haga doble click sobre el archivo que acaba de descargar y siga las instrucciones de instalación.

CLC permite realizar una serie de análisis sobre nuestras secuencias, dichos análisis posibles se pueden ver en el menú “toolbox”.

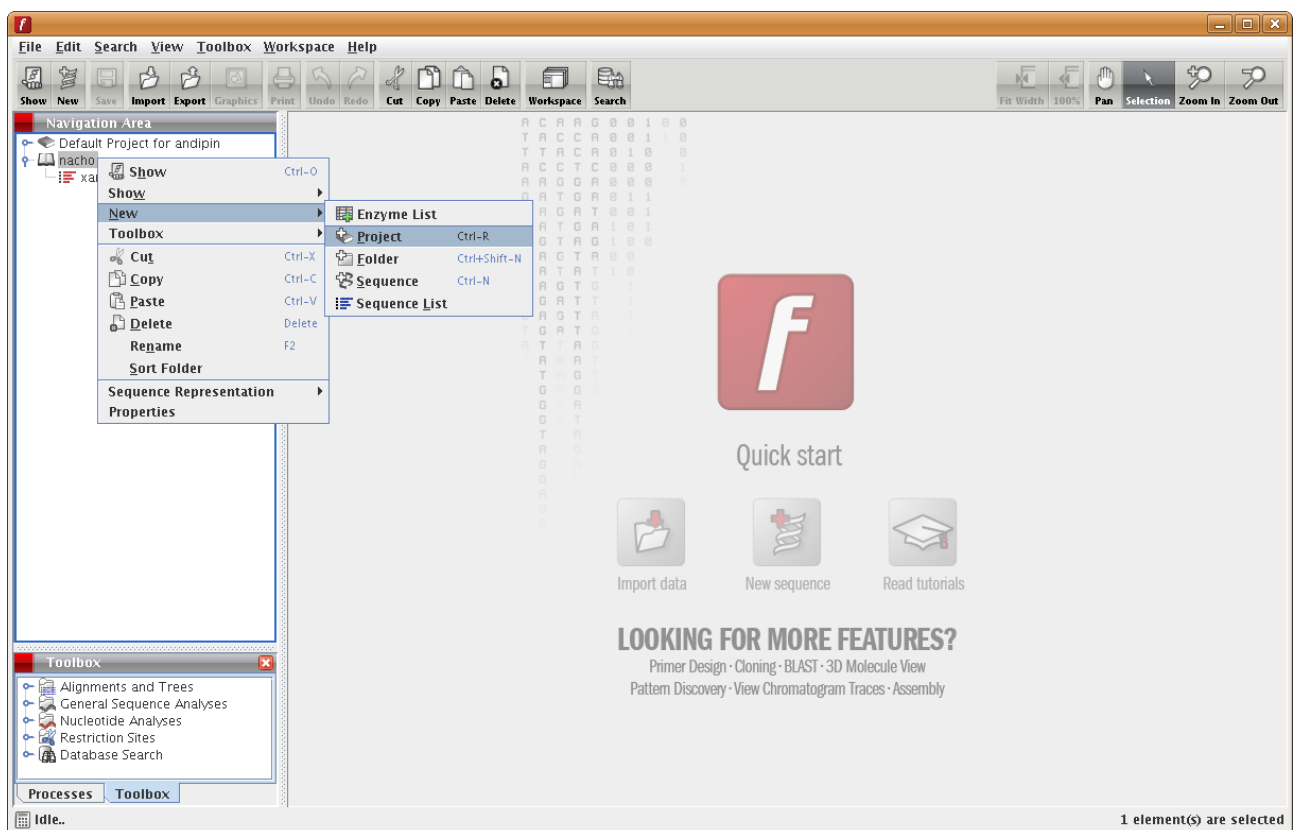


Asegúrese de familiarizarse con las diferentes opciones y menús en CLC.

Creación de un proyecto

En CLC se usa el concepto de proyecto, que es simplemente un espacio de trabajo en el que tenemos un conjunto de secuencias y análisis restringidos a dicho espacio de trabajo. Esto es útil en casos en los que estamos trabajando en múltiples proyectos y queremos mantener los datos y análisis separados y organizados.

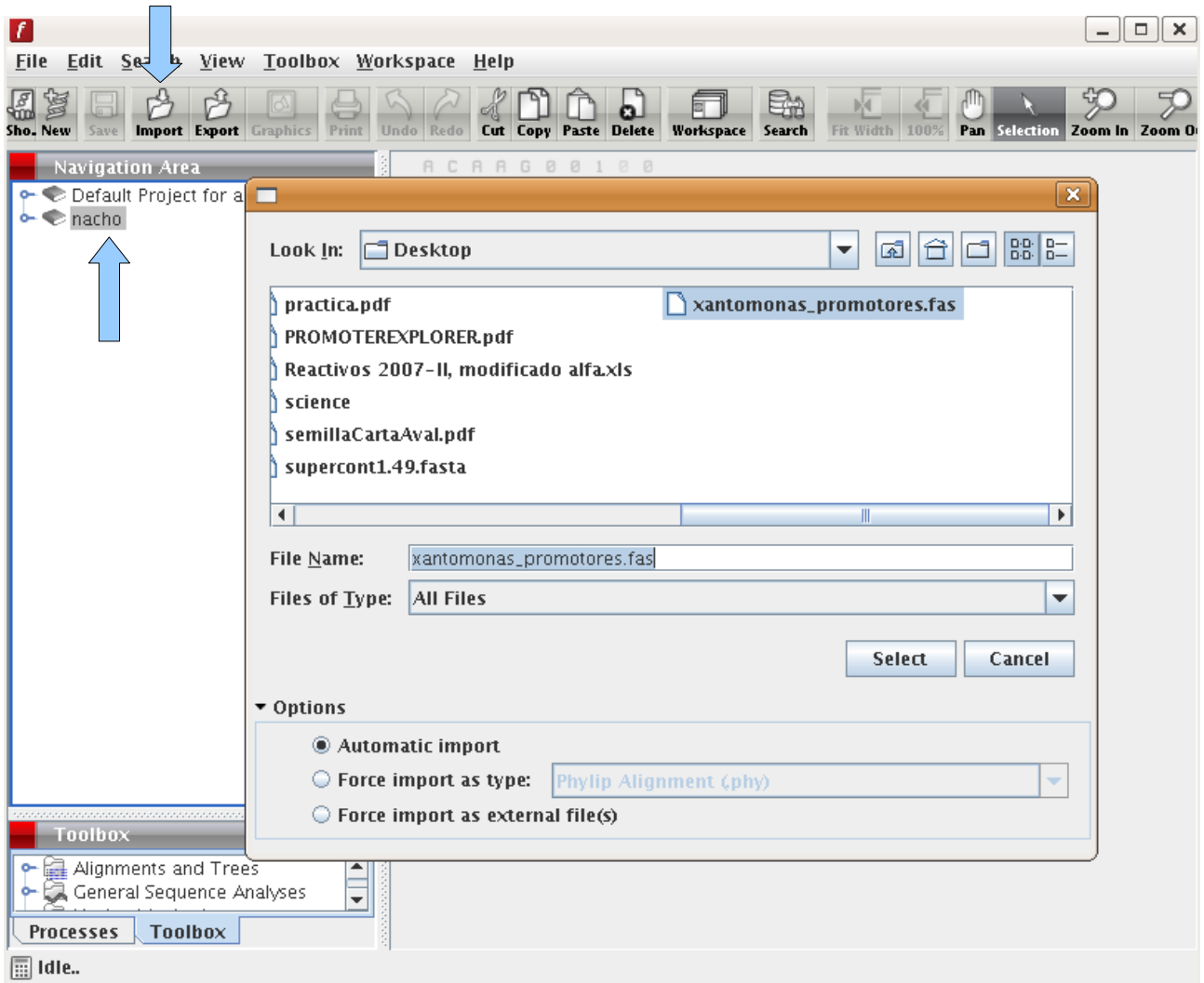
De click derecho sobre la ventana lateral izquierda “Navigation area” y a continuación en “new”, “project”. Digite el nombre que quiera dar a su proyecto. De ahora en adelante allí se guardaran los datos y análisis realizados.



Descargue el siguiente set de secuencias:

http://bioinf.ibun.unal.edu.co/cbib/estudiantes/tmp/xantomonas_promotores.fas

Este set de secuencias debe ser importado al proyecto que justo acabamos de crear. En este caso hemos llamado al proyecto “nacho”, así que debemos hacer click sobre el icono de nuestro proyecto y hacer click nuevamente sobre el icono con el título “import”, en la barra superior y seleccionar el archivo descargado.



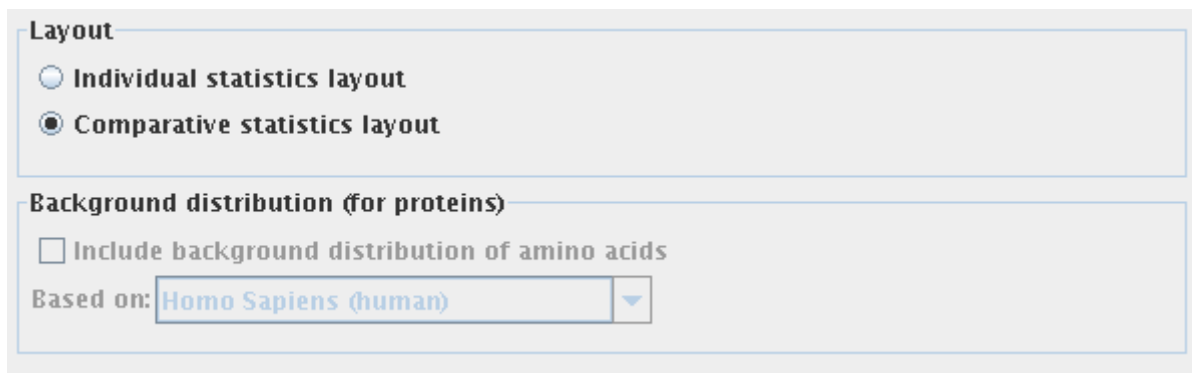
Inmediatamente el archivo es agregado a la carpeta de nuestro proyecto.

Analizando nuestras secuencias

Haga click derecho sobre el archivo que acaba de descargar y a continuación:

toolbox->general sequence analysis->create sequence statistics

Aparece una ventana que le permite escoger la secuencia(s) sobre la cual realizar el análisis, por defecto ya viene seleccionada nuestra secuencia, haga click en “next”. aparece una nueva ventana ofreciendo diversas opciones Asegúrese de dejar las opciones por defecto, como lo muestra la gráfica a continuación:



Nuevamente haga click en el botón “next”. Y deje la opción por defecto “Open”. La diferencia entre seleccionar “open” o “Save”, radica en que con la primera opción los resultados nos van a ser mostrados de una vez en pantalla, la segunda opción no los muestra inmediatamente sino que primero nos pide guardarlos.
Haga click en el botón “finish”.

Aparecerá una pestaña de resultados como la siguiente:

1 Nucleotide sequence statistics

1.1 Sequence information

Information	AvrBs2_xan_cam_aves	*AvrRx01_xan_cam_aves	AvrXa10_xan_ory_ory	AvrXa7-2M_xan_ory_ory	XopA_xan_cam_aves	XopB_xan_cam_aves	XopC_xan_cam_aves	XopF1_xan_cam_aves	XopF2_xan_cam_aves	XopJ_xan_cam_aves	XopM_xan_cam_aves
Sequence type	DNA	DNA	DNA	DNA	DNA	DNA	DNA	DNA	DNA	DNA	DNA
Length	180 nuc	156 nuc	255 nuc	222 nuc	201 nuc	201 nuc	201 nuc	255 nuc	193 nuc	200 nuc	200 nuc
Organism	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available
Name	AvrBs2_xan_cam_aves	*AvrRx01_xan_cam_aves	AvrXa10_xan_ory_ory	AvrXa7-2M_xan_ory_ory	XopA_xan_cam_aves	XopB_xan_cam_aves	XopC_xan_cam_aves	XopF1_xan_cam_aves	XopF2_xan_cam_aves	XopJ_xan_cam_aves	XopM_xan_cam_aves
Description	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available
Modification Date	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available

Examine y describa los resultados obtenidos.

En términos generales, todas las opciones de análisis de CLC funcionan de igual manera a la descrita anteriormente, es decir que se siguen los pasos casi de igual manera para todo tipo de análisis.

Con base en las indicaciones dadas previamente, realice los siguientes análisis sobre el set de secuencias de trabajo:

1. Alineamiento múltiple de las secuencias de DNA. Fíjese en la barra lateral derecha en la pestaña de resultados, allí encontrará una serie de opciones de visualización e información de su alineamiento. Comente brevemente que utilidad tiene cada una de las opciones bajo el menú: **Alignment info**. ¿Qué regiones de la secuencia cuentan con mayor conservación (ej, la región 180-200)? ¿Cuál(es) la menos conservada?
2. Concatene las diferentes secuencias en una sola secuencia.
3. Convierta la secuencia concatenada a RNA.
4. ¿Cuántos marcos de lectura abiertos (ORF) tiene dicha secuencia?

Guía elaborada por Andrés M. Pinzón V., del Centro de Bioinformática del Instituto de Biotecnología en la Universidad Nacional de Colombia y del Laboratorio de Micología y Fitopatología de la Universidad de los Andes, y está distribuida bajo licencia:



Creative Commons

Bogotá Colombia – Febrero de 2007.

Cualquier sugerencia o inquietud dirigirla a:

ampinzonv@una.edu.co ó andrespinzon@gmail.com