Introducción a la Bioinformática, práctica 3: Creación de árboles filogenéticos.

Introducción

En esta práctica conoceremos los fundamentos de los análisis filogenéticos por medios informáticos. Principalmente cubriremos el uso de algunas de las herramientas bioinformáticas más conocidas y la manera de crear filogramas y cladogramas con ellas.

Para seguir adelante con esta práctica, es fundamental que conozca la manera de recuperar secuencias y cómo crear alineamientos múltiples con estas. Si encuentra alguna dificultad en esto, por favor remitase a las dos prácticas anteriores: BLAST y recuperación de secuencias biológicas, Alineamiento múltiple e identificación y búsuqeda de motivos.

Nota: estas prácticas se basan en la premisa de que se conoce el marco teórico general sobre el cual éstas se basan, de esta manera no se pretende que estas prácticas sean un tutorial en sí mismas, sino más bien un complemento fundamental para poner en práctica los conceptos recibidos en las clases teóricas. Sin embargo, eventualmente encontrará descripciones y explicaciones de conceptos que así lo requieran, estos se encuentran encerrados en marcos de color verde.

Preparando nuestras secuencias

No exageremos si afirmamos que el exito o fracaso de un análisis filogenético radica en el alineamiento múltiple (AM) que obtengamos de nuestras secuencias. Básicamente debemos seguir los lineamientos planteados en la anterior guía acerca de AM.

“Garbage in, garbage out” (basura a la entrada, basura a la salida), es una vieja premisa en análisis bioinformáticos, y básicamente nos recuerda que no importa lo efectivo que sea un programa o algoritmo, la calidad de nuestros resultados siempre dependerá de la calidad de la información que le suministremos a este.

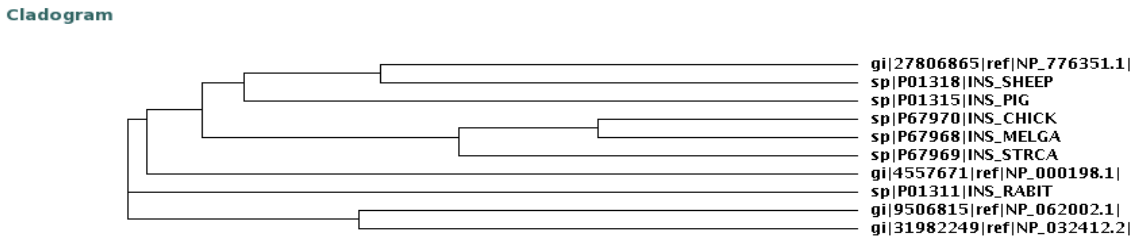
Recupere y realice un alineamiento múltiple (preferiblemente mediante clustalw, ya que necesitaremos una archivo .aln completamente válido más adelante) con las siguientes secuencias de insulina para diferentes especies:

1. **NP_000198**
2. **P30410**
3. **NP_062002**
4. **P01321**
5. **NP_032412**
6. **P01311**

7. **P01315**
8. **P01332**
9. **NP_776351**
10. **P01318**

Una vez realizado el alineamiento, preste atención al árbol guía calculado por clustalw, recuerde que este NO es un árbol filogenético. Entonces , ¿cuál es el significado de este árbol guía?

Guarde el archivo de alineamiento. Salve también el árbol guía, lo utilizaremos más adelante.



Edición de nuestro alineamiento

Generalmente, tendemos a pensar en un alineamiento múltiple como en un "archivo sagrado", intocable y portador de la verdad absoluta. Sin embargo, en la vida real es muy común realizar ediciones "manuales" de nuestros alineamientos, y esto puede responder a varias circunstancias. Por ejemplo, algunas veces resulta evidente para nuestros ojos que la remoción de una sección de gaps podría dar lugar a un mejor alineamiento (simplemente ningún algoritmo puede superar nuestra capacidad de raciocinio), sin llegar a alterar drásticamente nuestros resultados.

En el caso de la creación de árboles. Los gaps no son de mucha ayuda y por lo general trabajamos con las regiones muy conservadas, y es rutinario suprimir las regiones ricas en gaps y trabajar únicamente con dichos bloques de conservación.

En el caso de clustalw, es posible hacerle explícito que no queremos trabajar con regiones ricas en Gaps (más adelante detallaremos este proceso).

De esta manera es posible realizar una edición manual de nuestro alineamiento, esto es posible tanto desde **JalView**, como desde **BioEdit**, e incluso desde cualquier procesador de palabra como **Open Office Writer**, **Kword, Abiword** o

Ms Word.

* * * *

Creación de árboles filogenéticos con clustalw

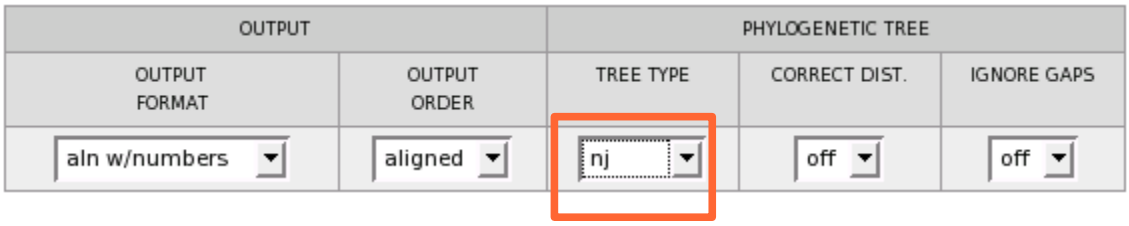
Existen herramientas mucho más “sofisticadas” que clustalw para la creación de árboles (más adelante veremos algunas de ellas), que nos permiten controlar cada uno de los aspectos de la creación de nuestro árbol, para lo cual debemos intimar con muchos de los detalles detrás de los algoritmos existentes para este fin. Particularmente con clustalw resulta muy sencillo crear nuestros árboles.

...visite nuevamente la página de clustalw en el sitio web del EBI: http://www.ebi.ac.uk/clustalw

e ingrese el alineamiento múltiple creado en el paso anterior.

Esta vez clustalw, entenderá que hemos ingresado un AM y que no debe crear uno nuevamente! Sin embargo debemos darle algunas indicaciones antes de continuar. Lo primero que debemos decirle es el tipo de método que utilizaremos para la creación de nuestro árbol:

Seleccione **NJ** en el menú desplegable: “**TREE TYPE”**, de la sección “**PHYLOGENETIC TREE”**.



Con esta opción hemos decidido crear nuestro árbol mediante el método **Neighbor Joining**, que es tal vez el método de mayor aceptación para este tipo de análisis.

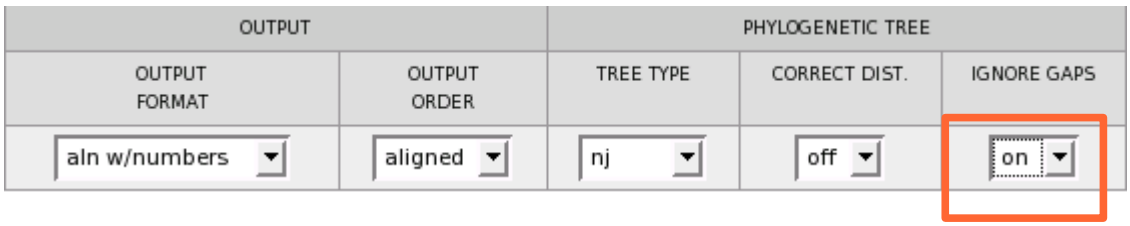
Métodos para la creación de árboles

Básicamente existen 2 categorías de métodos para la creación de árboles en estudios filogenéticos: Métodos basados en distancia (dentro de los que se encuentran UPGMA y NJ), y los métodos basados en la composición de las secuencias (acá se encuentran el método de máxima parsimonia ML, y el de máxima probabilidad ML).

Cada uno de estos métodos tiene sus fortalezas y sus debilidades, pero es más ampliamente usado NJ, ya que ha probado ser bastante eficaz.

Como fue mencionado anteriormente, generalmente no es conveniente crear nuestro árbol haciendo uso de las regiones ricas en gaps, si no hemos editado nuestro alineamiento manualmente debemos decirle a clustalw que no use dichas regiones.

Seleccione "**On**" en el menú desplegable: "**Ignore Gaps**" de la sección "**PHYLOGENETIC TREE**".

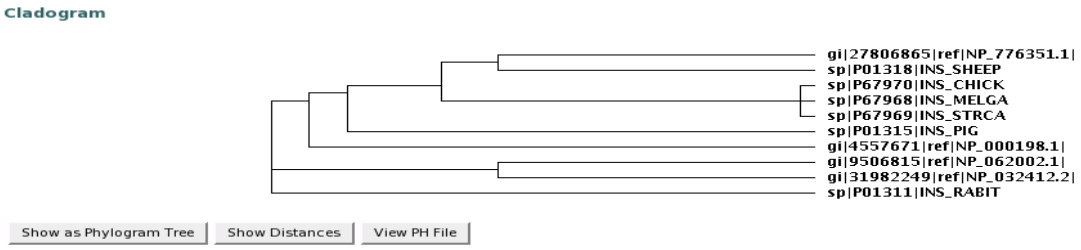


Después de unos segundos, aparecerán nuestros resultados. La página de resultados es similar a la presentada al hacer un alineamiento múltiple, sin embargo su contenido es un poco diferente.

Results of search	
Number of sequences	10
Sequence format	Clustal
Sequence type	aa
ClustalW version	1.83
Output file	clustalw-20060710-17163441.output
Neighbor-Joining tree file	clustalw-20060710-17163441.nj
Phylip tree file	clustalw-20060710-17163441.ph
Your input file	clustalw-20060710-17163441.input
SUBMIT ANOTHER JOB	

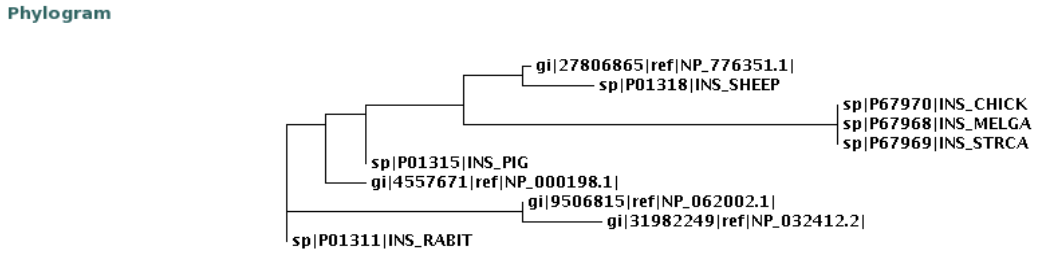
Cabe anotar que se crea por defecto un árbol Phylip, además del que hemos pedido (NJ).

Por defecto también, clustalw muestra primero un cladograma, que se ubica al final de la página de resultados:



Esta vez nos interesa conocer las distancias relativas de las ramas en nuestros árboles y no solamente su topología, así que hes más conveniente que veamos el árbol como un filograma:

Presione el botón "**Show as phylogram Tree**".



Compare este árbol con el árbol guía creado por clustalw al realizar el alineamiento. ¿Encuentra diferencias y/o similitudes entre ellos? ¿De qué manera se organizan lo OTUs en cada uno de estos?

Puede ser interesante guardar nuestro árbol para futuros análisis y tal vez la idea de tomar capturas de pantalla cada vez que queramos guardar nuestros árboles no resulte muy atractiva. Clustalw nos ofrece la opción de guardar nuestro árbol en un formato conocido como "**Newick**".

Presione el botón "**View PH File**" que se encuentra en la parte inferior del filograma.

Después de unos segundos será llevado a una página web en la cual nuestro árbol ha sido convertido a dicho formato, que se caracteriza por su gran cantidad de paréntesis.

Creación de árboles filogenéticos mediante Phylip

Como fue mencionado anteriormente, clustalw es básicamente una herramienta para alineamiento múltiple de secuencias, sin embargo hemos visto que es posible realizar muy buenos árboles mediante su utilización.

Para el mundo de la filogenia, sin embargo, existen otros programas mucho más usados, y que de alguna manera podemos caracterizar como “de gama alta”, es decir son programas (y suites) específicamente desarrollados para este propósito, generalmente reservados para los “gurúes” en este tipo de análisis.

Dentro de esta gama alta de programas contamos con PAUP y PHYLIP. En esta oportunidad trabajaremos con este último.

Visite la siguiente dirección:

http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html

Phylip puede ser considerado mejor como un “paquete de programas para análisis filogenético” más que un programa individual. Esta es la razón por la cual se encuentra frente a un listado de programas en este momento.

Phylogeny : Phylip programs

Documentation.
FAQ (Frequently Asked Questions).

- **Programs for molecular sequence data [sequence.doc]**
  - DNA
    dnadist [advanced form] [dnadist.doc]
    Distances from DNA sequences.
    dnapars [advanced form] [dnapars.doc]
    Parsimony method for DNA.
    dnaml
    (Maximum likelihood method) has been removed ; please use rather fastDNAml,
  - Proteins
    protdist [advanced form] [protdist.doc]
    Distances from protein sequences.
    protpars [advanced form] [protpars.doc]
    Parsimony method for protein sequences.

- **Programs for distance matrix data [distance.doc]**

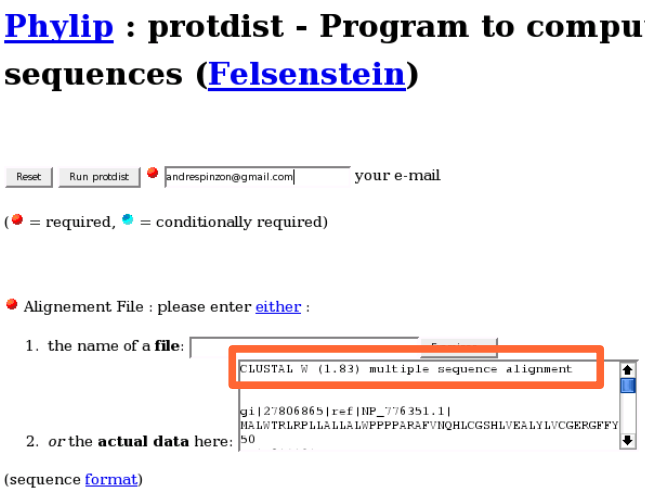
  neighbor [advanced form] [neighbor.doc]
  Neighbor-joining and UPGMA methods
  fitch [advanced form] [fitch.doc]
  Fitch-Margoliash and least-squares methods
  kitsch [advanced form] [kitsch.doc]
  Fitch-Margoliash and least-squares methods with molecular clock

- **Programs for trees**

El “truco” para trabajar con Phylip es utilizar los programas en orden y conocer de antemano lo que queremos lograr.

Nuestro objetivo es crear un árbol basado en distancias, así que lo primero que debemos hacer es crear la matriz de distancias de nuestro alineamiento múltiple.

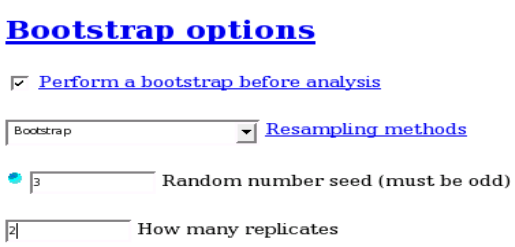
En la primera sección “**Programs for molecular sequence data**” en la sección “**Proteins**” presione el enlace “**advanced form**” del programa “**protdist**”. Será llevado-a la siguiente página web:



Ingrese allí su dirección de correo electrónico, también copie y pegue el alineamiento múltiple realizado anteriormente (Incluyendo la línea que dice CLUSTALW).

Dependiendo de la longitud del alineamiento, los resultados pueden aparecer en pantalla inmediatamente o ser enviados a la dirección de correo que se ha ingresado, esto depende de la carga del servidor y no existe manera de controlar este comportamiento.

Siga el enlace a Bootstrap en la parte inferior del formulario.



Este formulario nos permite definir los parámetros bajo los cuales realizaremos el bootstrap de nuestro árbol.

Seleccione la casilla “**Perform a bootstrap before analysis**”. En la casilla “**Random Number seed”,** introduzca “**3**” y en la opción “**How many replicates**” 2.

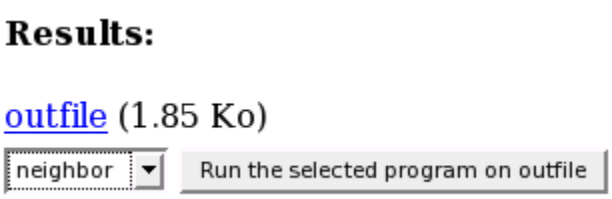
El número de réplicas por lo general debe ser como mínimo 100. Esto implica sin embargo una gran carga para el servidor, por ahora dejaremos este número en 2, así obtendremos nuestros resultados mucho más rápido y, en general, el concepto se podrá comprender igualmente.

Despues de unos minutos llegará una serie de correos a la direccion que ingresó anteriormente.

El más importante de estos correos es el primero de ellos: “”**acces all protdist results**”, los otros correos contienen detalles acerca de los análisis realizados. El contenido de este correo es una URL que le llevará a una página web con los resultados.

Siga dicho link a la página de resultados.

Seleccione “**neighbor**” del menú desplegable en la parte superior de la página de resultados.



Mediante esta opción estamos especificando que usaremos el programa **Neighbor,** que convierte nuestra matriz de distancia en un árbol NJ.

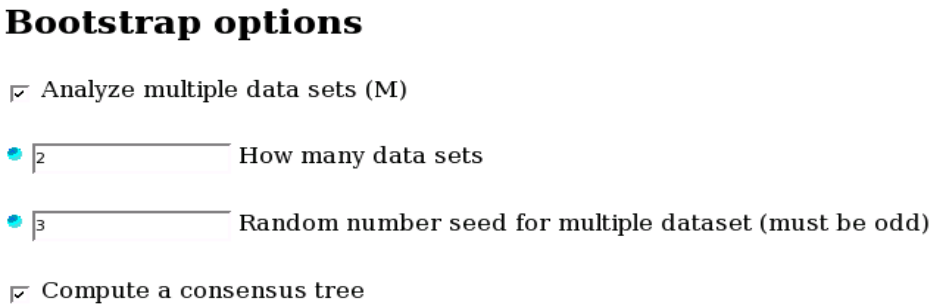
Presione el botón “**Run the selected program on outfile**”.

La anterior acción le llevará a una nueva página web.

Seleccione Neighbor-Joining como método de distancia.

De esta manera obtendremos un árbol NJ.

Siga el enlace que dice "**Bootstrap options**", y llene los campos de acuerdo a la siguiente imágen.



En la sección "**other options**", ingrese el número "**8**" en la casilla "**outgroup species**". ¿tiene alguna idea de por qué hemos escogido este número para el outgroup? **Pista**: está directamente relacionado con el alineamiento múltiple realizado.

Presione el botón "run neighbor" que se encuentra en la parte superior de la página.

Después de unos minutos apareceran los resultados en pantalla (o llegarán los enlaces correspondientes a su correo).

Results:

outfile.consense (1.93 Ko)

outtree.consense

consense | Run the selected program on outtree.consense

outfile (2.92 Ko)

outtree

consense | Run the selected program on outtree

params

consense.params

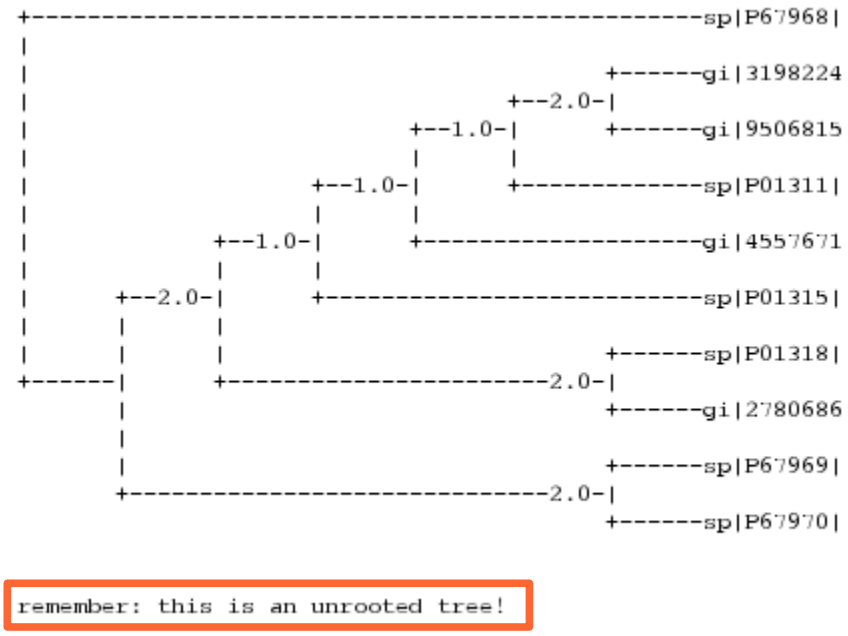
neighbor.out (5.65 Ko)

standard error file

From now, this files will remain accessible for 10 days at: http://bioweb.pasteur.fr/seqanal/tmp/neighbor/A24697011525708/
You can save them individually by the **Save file** function if needed.

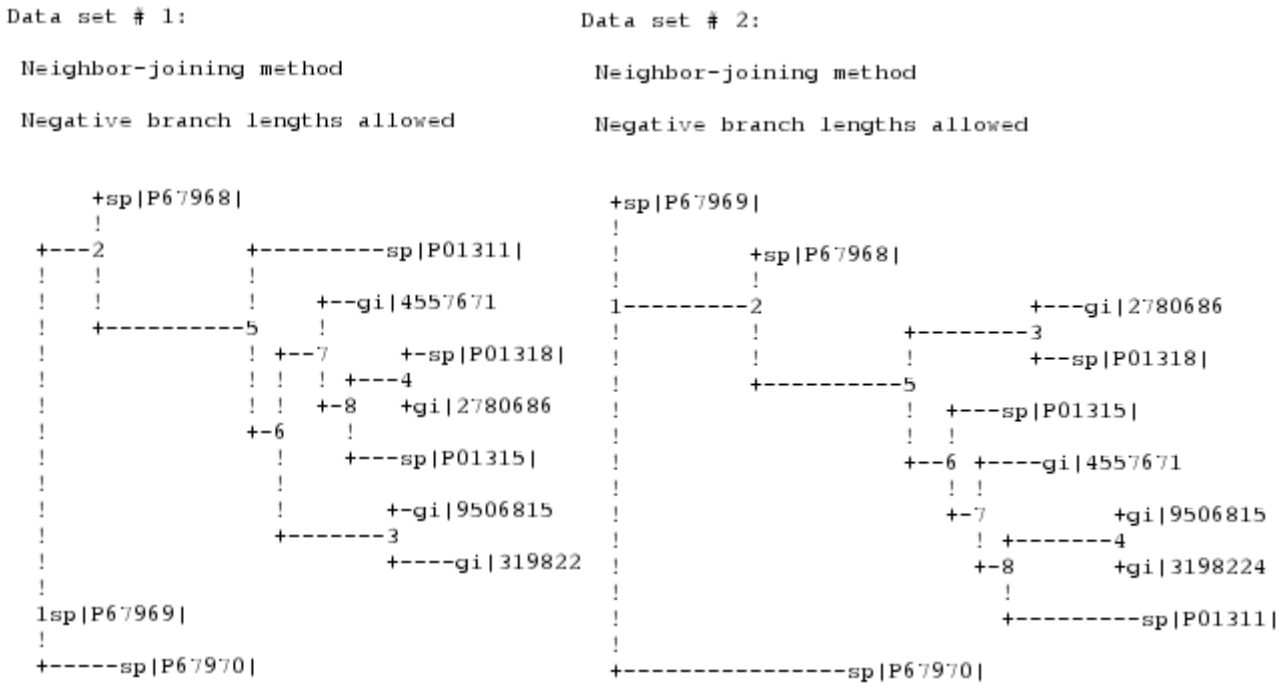
Job summary | default format

Los primeros 2 enlaces llevan a los archivos que contienen los árboles de los consensos. "**Outfile.consense**" muestra una versión en texto de nuestro árbol, así como datos generales acerca del proceso de "bootstrapping".



"**Outtree.consense**", lleva a un archivo en formato "Newick".

Los siguientes dos enlaces nos muestran los árboles que fueron utilizados para crear el árbol consenso. De igual manera que para el árbol consenso, el enlace "**outtree**" es una versión en formato Newick de dichos árboles.



Esta vez encontramos únicamente dos árboles debido a que este fue el número de réplicas que pedimos, generalmente deben ser alrededor de 100!

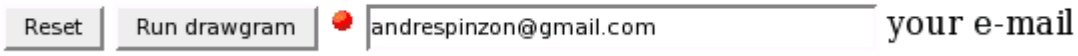
Phylip, ofrece también herramientas que permiten visualizar nuestro árbol de una mejor manera.

Seleccione la opción “drawtree” del menú desplegable, justo debajo de los dos primeros enlaces y presione el botón ubicado a la derecha de este.

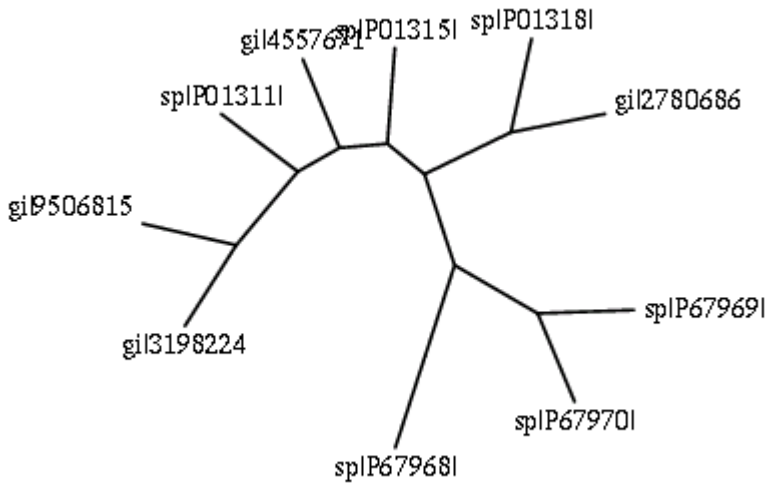
Esto le llevará a una nueva página en la que puede establecer el formato de la gráfica en la que será exportado nuestro árbol. Además de si queremos ver nuestro árbol como un fenograma o un cladograma.

Esta vez utilizaremos las opciones por defecto (archivo en formato postscript y cladograma).

Presione el botón “**Run drawtree**”. Asegúrese de que su dirección de correo electrónico se encuentra en la casilla.



Obtendrá una nueva página web con un enlace al archivo postscript.



Bogotá Colombia - Julio de 2006.

Cualquier sugerencia o inquietud dirigirla a:

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