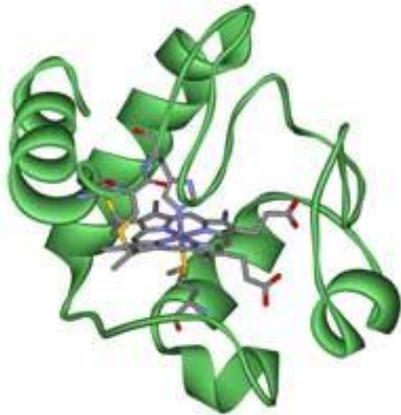


BASES DE DATOS Y ÁRBOLES FILOGENÉTICOS

El objetivo de esta actividad (**Tenorio G.** 2014. *Uso de bases de datos biológicas para enseñar evolución y bioquímica*) es conocer y utilizar las bases de datos biológicas para comparar las secuencias de proteínas y entender diferencias en la evolución de organismos.



Estructura cristalina del citocromo c

Imagen cortesía de Klaus

Hoffmeier/Wikimedia commons

Aunque se siguen utilizando análisis fenotípicos para entender la evolución de los seres vivos, es decir análisis basados en las características morfológicas de las especies, lo más frecuente hoy en día es realizar análisis moleculares.

Si se comparan genes homólogos o secuencias de proteínas de organismos con un supuesto ancestro común, se puede saber cuánto tiempo ha pasado desde que dichos organismos divergieron, pues se supone que cuanto más tiempo pasa, tanto más diferentes serán sus ADN o las secuencias de sus proteínas. Dicha divergencia se suele visualizar en un árbol filogenético.

Si queremos examinar la similitud de sus genes, lo primero es conseguir las secuencias de ADN y después alinearlas correctamente. Gracias a Internet las bases de datos biológicas son fácilmente accesibles.

En esta actividad vamos a utilizar dos recursos. Uno es del Centro Nacional de Información Biotecnológica (NCBI) de USA, que ofrece acceso libre a información biomédica y genómica. El otro es del Instituto Europeo de Bioinformática (EBI) de GB, que también ofrece datos gratuitamente de diversos experimentos y realiza investigación básica en biología computacional. De la base de datos del NCBI vamos a obtener las secuencias de los genes y proteína que necesitamos, y después vamos a utilizar las herramientas disponibles en el EBI para alinear dichas secuencias y analizarlas.

Método

Cuando se investigan relaciones evolutivas entre organismos diferentes, es importante elegir con cuidado qué gen o proteína se estudiará. Lo más habitual es escoger genes presentes en una amplia variedad de seres vivos, como puede ser los que determinan los citocromos; en este caso elegimos el citocromo c.

El citocromo c es una pequeña hemoproteína que forma parte de la cadena de transporte de electrones en la mitocondria. Si partimos de la base de que todos los seres vivos aeróbicos han evolucionado de un ancestro común que utilizó el primer citocromo c, podemos esperar que sea una elección acertada en este caso.

La actividad consta de tres partes:

- buscar la secuencia de aminoácidos del citocromo c en diferentes organismos,
- alinearlos
- construir el árbol filogenético.

En esta actividad se utilizarán las siguientes especies: *Anabaena variabilis*, *Arabidopsis thaliana*, *Canis lupus*, *Chlamydomonas reinhardtii*, *Drosophila melanogaster*, *Escherichia coli*, *Pica pica*, *Zea mays*, *Lemur catta*, *Tarsius bancanus*, *Cebus albifrons*, *Macaca sylvanus*, *Hylobates lar*, *Pongo pygmaeus*, *Gorilla gorilla*, *Homo sapiens*, *Pan troglodytes*, *Pan paniscus*

Se pueden añadir más especies, todas las que se quiera, pero hay que tener en cuenta que al aumentar el número de especies, se tarda más en buscar las secuencias y el programa también tarda más en ofrecer la solución.

I. Buscando la secuencia de proteínas

1. Ir a la página web de NCBI.
2. En la zona de búsqueda, en la parte superior de la página, seleccionar "protein" en el menú desplegable.
3. Escribe el nombre de una especie, ej: *Homo sapiens* y cytochrome c o cytochrome c oxidase.
4. Hacer clic en el botón de búsqueda.
5. En una nueva página aparecerá la lista de resultados. La mayor parte de ellos tienen la misma secuencia, a partir de fuentes diferentes, pero otros pueden ser secuencias parciales o pertenecer a una especie o proteína diferente. Elije con cuidado la proteína correcta de interés y haz clic en el enlace marcado como 'FASTA'.
6. De la nueva página que se cargue, copie la cadena de letras mayúsculas que representa la secuencia de aminoácidos. Pega las letras en un documento Word, recuerda etiquetar la secuencia con el nombre del organismo del que se obtuvo, de la forma siguiente: >Homo sapiens. Si no colocas el signo > delante del nombre el programa no lo leerá.
7. Repetir estos pasos con todos los organismos de la lista. Puedes incluir más organismos de los distintos reinos, para ver cómo evolucionó la vida en general.

II. Alineando secuencias

8. Ir al sitio web del Instituto Europeo de Bioinformática (EBI) y haz clic en "Services". Después selecciona "proteins".
9. Haz clic en Clustal Omega. Copia el texto del documento Word y pégalo en el cuadro de texto etiquetado como "PASO 1" (STEP 1), no olvides poner el nombre de cada especie con el símbolo > delante. Asegúrate de que las letras de aminoácidos de cada secuencia queden seguidas, no pueden haber huecos.

10. En el PASO 2 (STEP 2), elije un formato 'Clustal w/ numbers' que mostrará la longitud de cada secuencia. Finalmente, haz clic en "entregar" para completar el PASO 3 (STEP 3).
11. Las alineaciones para las múltiples secuencias aparecerán en una nueva ventana. Lo primero que es hacer clic en *show colors*. Esta opción da el mismo color a cada aminoácido, y así es más fácil identificarlos. Los colores agrupan los aminoácidos por sus características. Los rojos son pequeños, hidrofóbicos y aromáticos; los azules son ácidos; los magenta son básicos; los verdes son hidroxilos, aminos, amidas y básicos; y los grises son el resto.
12. Para analizar la alineación, ten en cuenta los siguientes símbolos:
 - un asterisco (*) significa que las secuencias son idénticas en esa posición
 - dos puntos (:) indica sustituciones conservadas (mismo color de grupo)
 - un punto y seguido (.) se refiere a una sustitución semi- conservada (formas similares).
13. Si hace clic en la opción "*Result summary*", puede ver el porcentaje de identidad (*percent identity matrix*) conservado entre los diferentes organismos tras la alineación. Hacer clic y analiza en esta matriz el porcentaje de identidad de dos organismos para la secuencia de la proteína citocromo c. Además, si está instalado Java™, se puede usar Jalview, un programa gratuito para editar múltiples alineaciones de secuencias, visualización y análisis. Con Jalview, se podrá ver la secuencia consensuada del citocromo c y el nivel de conservación de los diferentes aminoácidos.

El software Clustal Omega tiene otras muchas opciones diferentes con las que operar

III. Construyendo un árbol filogenético

14. En los resultados de Clustal Omega, haga clic en "*Arbol filogenético*" en la parte inferior (se necesita tener instalado Java™).
15. Puede obtener el árbol filogenético o el cladograma. En un cladograma, las longitudes de las ramas del árbol son arbitrarias, mientras que en un árbol filogenético, las longitudes de las ramas indican cuánto ha evolucionado la proteína en el tiempo.

Cuestiones

1. *Indica tus resultados con un corta-pegar de la secuencia alineada con colores y con el árbol filogenético*
2. *Las moléculas homólogas son un ejemplo de evolución divergente. ¿Cómo se puede explicar la evolución divergente utilizando el citocromo c?*
3. *Las alineaciones se pueden hacer utilizando secuencias de nucleótidos (genes) o aminoácidos (proteínas). ¿Por qué es más útil usar proteínas que ADN para analizar relaciones de evolución?*
4. *En un árbol filogenético, un "clado" queda formado por todos los organismos que tienen un ancestro común. Da un ejemplo de clado en tu cladograma.*

5. ¿Qué organismo(s) ha(n) sufrido un evento de especiación más recientemente, según los análisis filogenéticos del citocromo c? ¿Cuál es el número de eventos de especiación?
6. Indica alguna causa que explique por qué algunos aminoácidos han cambiado debido a mutación y otros no. Los aminoácidos conservados ¿no han cambiado porque sus codones no han sufrido ninguna mutación en absoluto?
7. Busca alguno de los aminoácidos conservados en la alineación. Investiga su función en el citocromo c en Internet.

Glosario

Secuencia de consenso: Un conjunto conocido de secuencias conservadas, o el orden calculado de los aminoácidos más frecuentes encontrados en cada posición en una alineación de secuencias.

Aminoácido conservado: Una secuencia de aminoácidos en un polipéptido que es similar en múltiples organismos.

FASTA: un formato basado en texto para representar secuencias de nucleótidos o péptidos usando códigos de una sola letra. Una secuencia en formato FASTA comienza con una descripción en una sola línea, seguida por líneas de datos de secuencias. La línea de descripción se distingue de los datos de secuencia a través del símbolo mayor que (>).

Proteína homóloga: Aquellas proteínas compartidas por algunos organismos que derivan de un ancestro común.

Evento de Especiación: El momento en que especies ancestrales divergen hacia nuevas especies.

Referencias

- Bromhan L. Penny D. (2003) [The modern molecular clock](#). *Nature Reviews Genetics* 5: 216-224
- Kozlowski C. (2010) [La bioinformática con papel y lápiz: construcción de un árbol filogenético](#). *Science in School* 17: 28-33.
- Sievers F. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539

Recursos en la red

- **Tenorio G.** 2014. Uso de bases de datos biológicas para enseñar evolución y bioquímica <http://www.scienceinschool.org/node/4653>
- Centro Nacional De EEUU para la Información Biotecnológica (US National Center for Biotechnology Information) <http://www.ncbi.nlm.nih.gov/> ofrece acceso a información biomédica y genómica.
- El Instituto Europeo de Bioinformática (European Bioinformatics Institute) <http://www.ebi.ac.uk/> ofrece datos accesibles gratuitamente

- Kimball J. ha escrito un libro de biología disponible online denominado 'Taxonomía: Clasificando la Vida' que incluye un capítulo sobre '[Árboles filogenéticos](#)'.

Recursos

- Para aprender más sobre el uso del [citocromo c en árboles fologenéticos](#).
- El sitio web [Protein Data Bank](#) (banco de datos de proteínas) en Europa pertenece al Instituto Europeo de Bioinformática y puede utilizarse para buscar y ver la estructura 3D del citocromo c.