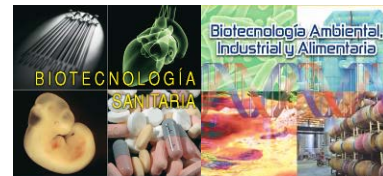

Póster

Implementación y puesta a punto de un algoritmo de búsqueda automatizada de secuencias homólogas



Pablo Mier¹ y Antonio J. Pérez-Pulido^{1,*}

¹Centro Andaluz de Biología del Desarrollo (CABD), CSIC-UPO, Facultad de Ciencias Experimentales (Área de Genética), Universidad Pablo de Olavide, 41013 Sevilla, España.

Palabras clave: bioinformática, homología de secuencia; aplicación web

RESUMEN

Motivación: El término “homología” hace referencia al origen evolutivo común entre dos proteínas que han divergido, presentando características comunes a nivel de secuencia y estructura (Fitch, 1970). Generalmente, la búsqueda de proteínas homólogas a una proteína de interés se realiza de forma manual, mediante el uso del algoritmo PSI-BLAST, ya sea en un ordenador propio o en un portal web como el del NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). La búsqueda de homólogos de forma manual es tediosa y requiere de mucho tiempo y conocimientos básicos de bioinformática. Una alternativa para la búsqueda de homólogos es el uso de algoritmos de acceso abierto disponibles en la web. Adicionalmente, existen muchas bases de datos que recogen conjuntos de proteínas homólogas previamente formados. El problema principal, tanto de los algoritmos como de las bases de datos es su dificultad de uso, ya que por lo general ninguno de estos recursos está destinado ni preparado para investigadores de laboratorio.

Resultados: Se ha desarrollado una herramienta escrita en lenguaje de programación Perl para la búsqueda automatizada de proteínas homólogas a una secuencia inicial de interés. El funcionamiento del algoritmo se ha probado con distintos conjuntos de secuencias, a partir de cuyos resultados se han optimizado los parámetros que rigen su funcionamiento. El algoritmo desarrollado es eficiente (sensible, específico y rápido), exhaustivo a la hora de buscar secuencias homólogas y admite el uso de bases de datos del propio usuario. La modificación de los parámetros por defecto permite que el usuario pueda primar un mejor dato de sensibilidad en detrimento del valor de especificidad, o viceversa.

Adicionalmente, se ha diseñado una aplicación web de libre acceso para cualquier usuario (<http://www.bioinfocabd.upo.es/cgi-bin/orthoFinder/ortholog2.cgi>) para el uso del algoritmo. Sólo es necesario subir un archivo con la secuencia de interés para iniciar la ejecución básica del programa, usando los parámetros óptimos por defecto. La simplicidad de la interfaz facilita la ejecución del algoritmo por parte del usuario. Los resultados de una búsqueda están disponibles generalmente en menos de dos minutos, y quedan allí almacenados durante una semana para posibles consultas recurrentes. La facilidad de uso del programa por medio de la aplicación web, así como el método de presentación de los resultados, permiten generar una interfaz de uso del programa por y para distintos tipos de usuarios.

BIBLIOGRAFIA

Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Syst.Zool.*, **19**, 99–113.