

BIOINFORMÁTICA

PARTE 2: SERVICIOS WEB EN BIOINFORMÁTICA

Resumen. Este documento es el segundo de una serie donde hablaremos de la Bioinformática y los usos que puede tener en la investigación médica.

Una aplicación primordial de la Bioinformática es la consulta de información a través de bases de datos disponible en la web.

Existen bases de datos primarias, que contienen información directa de la secuencia, estructura o patrón de expresión de ADN o proteína, y secundarias, que contienen datos e hipótesis derivados del análisis de las bases de datos primarias, como mutaciones, relaciones evolutivas, agrupación por familias o funciones, implicación en enfermedades, etc.

Bases de Datos de nucleótidos

La colaboración de las tres bases de datos más importantes hace posible acceder a casi toda la información de secuencias de ADN desde cualquiera de sus tres sedes:

EMBL-BANK en el Instituto europeo de Bioinformática (EBI)

DNA Data Bank of Japan (DDBJ) en el Centro de Información Biológica (CIB)

GenBank en el Centro Nacional de Información Biotecnológica (NCBI)

Si bien son mantenidas por distintos organismos en distintos países, existe una coordinación entre las distintas bases. Una secuencia enviada a cualquiera de las bases se verá reflejada en las otras dos en aproximadamente una semana, ya que esa es la frecuencia de actualización entre las distintas bases genéticas. Por este motivo es indistinto que base se use para enviar nuevas secuencias, aunque normalmente los europeos utilizan EMBL-BANK y los americanos GenBank.

Bases de Datos de Proteínas

Bases de datos de secuencias de aminoácidos.

Swissprot contiene secuencias anotadas o comentadas, es decir, cada secuencia ha sido revisada, documentada y enlazada a otras bases de datos.

TrEMBL por Translation of EMBL Nucleotide Sequence Database incluye la traducción de todas las secuencias codificantes derivadas del (EMBL-BANK) y que todavía no han podido ser anotadas en Swissprot.

PIR por Protein Information Resource está dividida en cuatro sub-bases que tienen un nivel de anotación decreciente.

ENZYME enlaza la clasificación de actividades enzimáticas com-

pleta a las secuencias de Swissprot.

Otras son Prosite, Interpro, PDB.

Bases de Datos de Genomas

Ensembl integra genomas eucariotas grandes, entre otros contiene genoma humano, ratón, rata, fugu, zebrafish, mosquito, Drosophila, C. elegans, y C. briggsae.

Genomes server y TIGR son portales con información o enlaces de todos los genomas secuenciados por el momento, desde virus a humanos.

En el siguiente artículo comentaremos más sobre las aplicaciones de la bioinformática en la investigación médica.

Ing. Jaime Rodríguez Gaona

Consultor Asociado, BioEstadística, S.C.

jaime.rodriguez@bioestadistica.com