

Atividade Semana 3

- Victor Azadinho Miranda
- RA: 171042191

GenBank Sequence Database

O GenBank é o banco de dados de sequências genéticas do NIH, uma coleção anotada de todas as sequências de DNA disponíveis ao público. O GenBank faz parte da Colaboração Internacional no Banco de Dados de Sequência de Nucleotídeos (INSDC), que inclui o DNA DataBank do Japão (DDBJ), o Arquivo Europeu de Nucleotídeos (ENA) e o GenBank no NCBI. Essas três organizações trocam dados diariamente.

O GenBank e seus colaboradores recebem sequências produzidas em laboratórios em todo o mundo de mais de 100.000 organismos distintos. O banco de dados foi iniciado em 1982 por Walter Goad e *Los Alamos National Laboratory*. O GenBank se tornou um banco de dados importante para pesquisas em campos biológicos e cresceu nos últimos anos a uma taxa exponencial, dobrando aproximadamente a cada 18 meses.

A versão 194, produzida em fevereiro de 2013, continha mais de 150 bilhões de bases de nucleotídeos em mais de 162 milhões de sequências. O GenBank é construído por envios diretos de laboratórios individuais, bem como envios em massa de grandes centros de sequenciamento.

EMBL Nucleotide Sequence Database

O *EMBL Nucleotide Sequence Database*, mantido no Instituto Europeu de Bioinformática (EBI) perto de Cambridge, Reino Unido, é uma coleção abrangente de sequências e anotações de nucleotídeos e anotações de fontes públicas disponíveis. O banco de dados faz parte de uma colaboração internacional com DDBJ (Japão) e GenBank (EUA). Os dados são trocados diariamente entre os institutos colaboradores para alcançar uma rápida sincronia.

Webin é a ferramenta preferida para envios individuais de sequências de nucleotídeos, incluindo anotações de terceiros (TPA) e alinhamentos. São fornecidos procedimentos automatizados para envios de projetos de sequenciamento em larga escala e dados do Escritório Europeu de Patentes. Registros de dados novos e atualizados são distribuídos diariamente e todo o banco de dados de sequências de nucleotídeos EMBL é lançado quatro vezes por ano. O acesso aos dados da sequência é fornecido via ftp e várias interfaces WWW.

Com o Sistema de Recuperação de Sequência (SRS) baseado na Web, também é possível vincular dados de nucleotídeos a outros bancos de dados especializados em biologia molecular mantidos no EBI. Outras ferramentas estão disponíveis para busca por similaridade de sequência (por exemplo, FASTA e BLAST).

Swiss-Prot Protein Sequence Data Bank

O SWISS-PROT é um banco de dados de sequência de proteínas anotado, estabelecido em 1986 e mantido em colaboração, desde 1988, pelo Departamento de Bioquímica Médica da Universidade de Genebra e pela Biblioteca de Dados EMBL. O banco de dados de sequência de proteínas SWISS-PROT consiste em entradas de sequência. As entradas de sequência são compostas por diferentes tipos de linhas, cada uma com seu próprio formato. Para fins de padronização, o formato do SWISS-PROT segue o mais próximo possível do banco de dados de sequências de nucleotídeos EMBL.

O banco de dados SWISS-PROT se diferencia de outros bancos de dados de sequência de proteínas por três critérios distintos.

Anotação

No SWISS-PROT, como na maioria dos outros bancos de dados de sequência, duas classes de dados podem ser distinguidas, os dados principais e a anotação. Para cada entrada de sequência, os dados principais consistem nos dados de sequência, nas informações de citação (referências bibliográficas) e nos dados taxonômicos (descrição da fonte biológica da proteína), enquanto a anotação consiste em uma descrição dos seguintes itens:

1. funções da proteína;
2. modificações pós-tradução, por exemplo carboidratos, fosforilação, acetilação, âncora GPI, etc.;
3. domínios e locais, por exemplo regiões de ligação de cálcio, locais de ligação de ATP, dedos de zinco, homeobox, kringle, etc.;

Redundância mínima

Muitos bancos de dados de sequência contêm, para uma dada sequência de proteínas, entradas separadas que correspondem a diferentes relatórios da literatura. No SWISS-PROT, busca-se o máximo possível mesclar todos esses dados, para minimizar a redundância do banco de dados. Se existirem conflitos entre vários relatórios de sequenciamento, eles serão indicados na tabela de recursos da entrada correspondente.

Integração com outros bancos de dados

É importante fornecer aos usuários de bancos de dados biomoleculares um grau de integração entre os três tipos de bancos de dados relacionados a sequências (sequências de ácidos nucleicos, sequências de proteínas e estruturas terciárias de proteínas), bem como com coletas de dados especializadas. Atualmente, o SWISS-PROT é referência cruzada com 24 bancos de dados diferentes. As referências cruzadas são fornecidas na forma de ponteiros para informações relacionadas às entradas do SWISS-PROT e encontradas em coleções de dados diferentes do SWISS-PROT.

DDBJ – DNA Data Bank of Japan

O Banco de Dados de DNA do Japão (DDBJ) é um banco de dados biológico que coleta sequências de DNA. Está localizado no Instituto Nacional de Genética (NIG), na prefeitura de Shizuoka, no Japão. Também é membro da *International Nucleotide Sequence Database Collaboration* ou INSDC. Ele troca seus dados com o Laboratório Europeu de Biologia Molecular do Instituto Europeu de Bioinformática e com o GenBank no Centro Nacional de Informações de Biotecnologia diariamente. Portanto, esses três bancos de dados contêm os mesmos dados a qualquer momento.

O DDBJ iniciou as atividades de banco de dados em 1986 no NIG e continua sendo o único banco de dados de sequência de nucleotídeos na Ásia. Embora o DDBJ receba principalmente dados de pesquisadores japoneses, ele pode aceitar dados de colaboradores de qualquer outro país. O DDBJ é financiado principalmente pelo Ministério da Educação, Cultura, Esportes, Ciência e Tecnologia do Japão (MEXT). O DDBJ possui um comitê consultivo internacional composto por nove membros, três membros da Europa, EUA e Japão. Esse comitê aconselha o DDBJ sobre sua manutenção, gerenciamento e planos futuros uma vez por ano. Além disso, o DDBJ também possui um comitê de colaboração internacional, que aconselha sobre várias questões técnicas relacionadas à colaboração internacional e é composto por participantes em nível de trabalho.