

# シーケンスライブラリ ダウンロード

(バージニア大学のHP [http://fasta.bioch.virginia.edu/fasta\\_www2/fasta\\_db\\_down.shtml](http://fasta.bioch.virginia.edu/fasta_www2/fasta_db_down.shtml) の和訳です。このドキュメントは、11/02/06 日の情報にもとづいて書かれています。このドキュメントは ecobioinfo.com で独自に訳したものでバージニア大学とは関係ありません。)

## シーケンスライブラリダウンロード

タンパク質と DNA のシーケンスライブラリファイルは、NCBI <ftp://ftp.ncbi.nih.gov/blast/db> や DMBL-EBL <ftp://ftp.ebi.ac.uk/pub/databases/> など、多くの異なる情報源からダウンロードすることができます。

## ライブラリフォーマット

FASTA プログラムは、多くの異なるライブラリフォーマットで動作します; FASTA でシーケンスライブラリを検索するためにファイル変換プログラムやフォーマットプログラムを使う必要は無いでしょう。しかし、FASTA プログラムは、ライブラリは FASTA フォーマット

[http://en.wikipedia.org/wiki/Fasta\\_format](http://en.wikipedia.org/wiki/Fasta_format) と仮定して動きます; 他のフォーマットのライブラリを検索するためには、ファイル名と一緒にフォーマットタイプを指定しなければなりません。例えば、

```
fasta34 -q mgstm1.aa "/slib/ncbi/refseq_protein 12"
```

は、NCBI/BLAST formatdb 形式の NCBI refseg\_protein ライブラリ を検索します。

サポートされている、よく知られたライブラリフォーマット:

フォーマット	説明
0	(デフォルト) FASTA フォーマット
1	Genbank flatfile
3	EMBL-EBI/Swissprot flatfile
5	GCG/PIR flatfile
6	GCG 圧縮バイナリ
12	NCBI BLAST formatdb version 2 (最新バージョン)
16	MySQL SQL query
17	PostgreSQL SQL query

## タンパク質とDNAのシーケンスデータベース

今日では、シーケンスデータベースを供給するサイトをひとつひとつ選ぶ理由はほとんどありません - 特に NCBI, EMBL-EBI, DDBJ 間で夜ごとに同期されている DNA シーケンスライブラリは。タンパク質シーケンスライブラリでは、NCBI と EMBL-EBI 両方で包括的な、しかし冗長なタンパク質シーケンスのコレクション、例えば NCBI NR と EMBL-EBI/PIR ユニポートを提供しています。しかし、両方のグループはさらにより質の高いデータベース、例えば NCBI refseg\_protein と EMBL-EBI SwissPort を提供しています。シーケンスの相同性検索は、より小さいデータベースを

使うときは、よりセンシティブになりますので、最初に小さい高品質のデータベースを探し、次に、重要な相同性が最初の検索で見つからなかった場合のみに包括的なデータベースを探すのが妥当です。

----- End of text -----