分子生物学用データベース一覧

インストール済みデータベース一覧

データベース名	種類	フォーマット	更新型
GenBank/GenBank-upd	核酸配列	テキストファイル, dbget用	定期/日々
EMBL/EMBL-upd	核酸配列	テキストファイル, dbget用	定期/日々
RefSeq/RefSeq-upd	核酸配列	テキストファイル, dbget用, BLAST用	定期/日々
EST_human/EST_mouse/EST_others	核酸配列	FASTA用, BLAST用	定期
NCBI nr-nt	核酸配列（非冗長）	FASTA用, BLAST用	定期
gss	核酸配列	FASTA用, BLAST用	定期
HTGS	核酸配列	FASTA用, BLAST用	定期
dbsts	核酸配列	FASTA用, BLAST用	定期
patnt	核酸配列	FASTA用, BLAST用	定期
NCBI nr-aa	アミノ酸配列（非冗長）	FASTA用, BLAST用, diamond用	定期
RefSeq-protin	アミノ酸配列	FASTA用, BLAST用, diamond用	定期
UniProt（TrEMBL, Swissprot）	アミノ酸配列	FASTA用, BLAST用, diamond用	定期
pataa	アミノ酸配列	FASTA用, BLAST用	定期
env_nr	アミノ酸配列	FASTA用, BLAST用	定期
pdbaa	アミノ酸配列	FASTA用, BLAST用	定期
PDB	アミノ酸配列	FASTA用, BLAST用	定期
kegg	遺伝子/ゲノム統合	テキストファイル, dbget用, FASTA用, BLAST用, diamond用	定期

データベースがあるディレクトリ

ディレクトリ **Directory**	内容 **Summary**
/apl/bio/ftp/(DB_name)/	FTPでダウンロードしたファイル　(/apl/bio/ftp/licenced/ (KEGG)はアクセス不可)
/apl/bio/db/ideas/(DB_name)/	テキストファイル　DBGET検索用インデックスファイル(.cdb, .tit)
/apl/bio/db/fasta/(DB_name)/	BLAST/FASTA検索用
/apl/bio/db/diamond/db/(DB name)/	DIAMOND検索用
/apl/bio/db/iproscan.bk/(DB name)/	InterProScan検索用
以下はシンボリックリンク	
/apl/bio/db/blast/db/	全BLAST/FASTA検索用DBファイルへのシンボリックリンク 環境変数 BLASTDB に設定済み
/apl/bio/db/diamond/db/	全DIAMOND検索用DBファイルへのシンボリックリンク
/apl/bio/db/igenomes/	イルミナゲノムズのDB (Illumina Genomes DB)

- コマンドの書き方例

diamond blastx --db /apl/bio/db/diamond/db/tremble --query exons.fa --out diamond.tab
blastx -db nr -query exons.fa -out blastout.tab -outfmt 6
環境変数BLASTDBが設定されているのでデータベースのパスは不要

DBGETコマンドの使い方

■ **DBGET**：分子生物学関連データベースの統合検索システム

- source /apl/bio/etc/bio.sh　すると使える
- コマンド名 -h でヘルプを表示する

- binfo：使えるデータベースの情報を取得

binfo
binfo (blast|fasta|diamond|dbget)
binfo diamond
diamondで使えるデータベースを表示する

- bfind：キーワード検索

bfind [option] (DB名) (keyword1) (keyword2) ...
bfind swissprot human interleukin
swissprotデータベースからキーワードhuman, interleukinを含むエントリを探す

- bget：配列データの取得

bget [option] (DB名):(ID1) (DB名):(ID2) ...
bget -f -n a swissprot:P05067
swissprotデータベースにあるP05067のアミノ酸配列のみを取得