

バイオアプリケーション・ バイオデータベース 利用法

BIAS5で利用可能なバイオ関連ソフトウェア一覧

次世代シーケンサデータ解析用	hisat2, stringtie, STAR, salmon, samtools, bamtools, BEDtools, bwa, RSEM, soap, bowtie, bowtie2, tophat, cufflinks, MACS2, GATK
次世代シーケンサデータ解析用: アセンブラ	Trinityrnaseq, canu, SPAdes, velvet, abyss, allpaths-lg
配列類似性検索	blast+, fasta, blat, diamond, MMseq2, vsearch
ペアワイズアライメント	lastz, MUMmer
マルチプルアライメント	clustalw2, clustal Omega, muscle, mafft, t_coffee, Gblocks
遺伝子情報取得	dbget
分子生物学用ソフトウェアパッケージ	EMBOSS
アセンブラー	CAP3, TGICL
遺伝子領域予測	augustus, genemark, genscan, glimmer
モチーフ検索	interproscan, hmmer, meme
系統樹解析	mrbayes, njplot, paup, phylip, phyml, tree-puzzle
言語など	R, python3, ruby

module whatis (module名) でアプリケーションの概要の表示が可能
module の一覧にないソフトを希望される場合はご連絡ください

module load

BIAS5のバイオアプリケーションは module コマンドで管理されています。

- 利用できるアプリケーションの module ファイルを表示

```
$ module avail
```

```
$ module avail bl
```

名前が bl から始まる module ファイルだけを表示

- アプリケーションの module ファイルを読み込む

```
$ module load (module名)
```

複数指定可

- 現在読み込んでいる module の確認

```
$ module list
```

- 読み込んでいる module を破棄

```
$ module unload (module名)
```

指定した module を破棄

```
$ module purge
```

読み込んだ module を全て破棄

module load

BIAS5のバイオアプリケーションは module コマンドで管理されています。

- アプリケーションの概要の表示

```
$ module whatis
```

```
$ module whatis (module名)
```

- module の設定内容の確認

```
$ module display (module名)
```

```
module display blast+/2.12.0
```

```
module-whatismodule whatis の内容
```

```
prepend-path # 環境変数 PATH の先頭に追加
```

実行中にコンフリクトが生じた場合、
module displayを実行することで
トラブル解決につながることも

BIAS5で利用可能なバイオ関連データベース一覧

項番	データベース	概要	フォーマット	更新型
1	GenBank/GenBank-upd	核酸塩基配列	フラット, DBGET	定期/日々
2	EMBL/EMBL-upd	核酸塩基配列	フラット, DBGET	定期/日々
3	RefSeq/RefSeq-upd	核酸塩基配列	フラット, DBGET, FASTA, BLAST	定期/日々
4	EST_human/EST_mouse/EST_others	核酸塩基配列	FASTA, BLAST	定期
5	NCBI nr-nt	非冗長核酸塩基配列	FASTA, BLAST	定期
6	gss	核酸塩基配列	FASTA, BLAST	定期
7	HTGS	核酸塩基配列	FASTA, BLAST	定期
8	dbsts	核酸塩基配列	FASTA, BLAST	定期
9	patnt	核酸塩基配列	FASTA, BLAST	定期
10	env_nt	核酸塩基配列	FASTA, BLAST	定期
11	pdbnt	核酸塩基配列	FASTA, BLAST	定期
12	NCBI nr-aa	非冗長アミノ酸配列	FASTA, BLAST, DIAMOND	定期
13	RefSeq-protein	タンパク質アミノ酸配列	フラット, DBGET, FASTA, BLAST, DIAMOND	定期
14	UniProt(TrEMBL, Swissprot)	タンパク質アミノ酸配列	フラット, DBGET, FASTA, BLAST, DIAMOND	日々
15	pataa	タンパク質アミノ酸配列	FASTA, BLAST	定期
16	env_nr	タンパク質アミノ酸配列	FASTA, BLAST	定期
17	pdbaa	タンパク質アミノ酸配列	FASTA, BLAST	定期
18	PDB	タンパク質立体構造	FASTA, BLAST	定期
19	kegg	遺伝子/ゲノム統合データベース	フラット, DBGET, FASTA, BLAST, DIAMOND	定期

DBGET基本コマンド binfo

binfo: データベースの情報を取得

- データベース全体の一覧を確認する

```
$ binfo
```

- 指定されたデータベースの情報を表示

```
$ binfo (DB名)
```

- 各検索ツールで利用できるデータベースを表示

```
$ binfo (dbget|fasta|blast|diamond)
```

binfoの実行例

- Blastで利用できるデータベースのリスト

```
$ binfo blast
```

- DBGETで利用できるデータベースのリスト

```
$ binfo dbget
```

DBGET基本コマンド bfind

bfind: キーワード検索

```
$ bfind [option] (DB名) (keyword1) (keyword2) ...
```

option: -C 大文字・小文字を区別して検索
-W パターンマッチではなく単語区切りで検索
-a エントリー名を ACCESSION [ID] で出力
-n 出力で DB名 を表示しない
-l (数字) 出力件数を制限

bfindの実行例

```
$ bfind swissprot human interleukin
```

swissport というDBからhumanとinterleukinの両方の情報を持つものを検索する

DBGET基本コマンド bget

bget: 配列データの取得

```
$ bget [option] (DB名):(ID1) ...
```

```
$ bget [option] (DB名) (ID1) (ID2) ...
```

option: -f FASTAフォーマットで配列を出力

 -n アミノ酸配列/塩基配列のみ出力する (-f オプションも利用する)

bgetの実行例

```
$ bget hsa:51341
```

```
$ bget -f hsa:51341       # 配列を取得
```

```
$ bget -f -n a hsa:51341   # アミノ酸配列のみを取得
```

```
$ bget -f -n n hsa:51341   # 塩基配列のみを取得
```


バイオデータベースの置き場所・フォーマット

ディレクトリ	内容
/bio/ftp/(DB名)/	FTPでダウンロードしたファイル (* /bio/ftp/licenced/ (KEGG)はアクセス不可)
/bio/db/ideas/(DB名)/	フラットファイル DBGET検索用インデックスファイル(.cdb, .tit) (* KEGG関係のDBはアクセス不可)
/bio/db/fasta/(DB名)/	BLAST/FASTA検索用DBファイル
/bio/db/diamond/(DB名)/	DIAMOND検索用DBファイル
/bio/db/iproscan/(DB名)/	InterProScan検索用DBファイル
/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
/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
/bio/db/igenomes/	イルミナゲノムズのDB

/bio/ftp にあるミラー

データベース	概要	ディレクトリ	URL
NCBI taxonomy	生物種分類	/bio/ftp/taxonomy/	ftp.ncbi.nih.gov/pub/taxonomy/
NCBI genomes	ゲノム	/bio/ftp/genomes/	ftp.ncbi.nih.gov/genomes/
NCBI Conserved Domain	タンパク質ドメイン構造	/bio/ftp/cdd/	ftp.ncbi.nih.gov/pub/mmdb/cdd/
InterProScan DB	InterProScan用	/bio/ftp/iprscan/	ftp.ebi.ac.uk//pub/databases/interpro/iprscan/
Ensemble	ゲノム	/bio/ftp/ensembl/	ftp.ensembl.org/pub/currnet_*/
Illumina iGenomes	ゲノム	/bio/ftp/Illuminalgenomes/	https://support.illumina.com/sequencing/sequencing_software/igenome.html