2 . データベースサービスの概要

東京大学医科学研究所ヒトゲノム解析センターでは、コンピュータネットワーク「ゲノムネット」を通じて京都大学化学研究所スーパーコンピューターラボラトリーと協力しながら、以下のサービスを行なっています。

1　データベース検索

簡単なキーワード検索とエントリーの取得には DBGET が使えます。コマンドとしては bget と bfind です。 bget ではエントリー名またはアクセッション番号を指定して目的のエントリーを取れます。 bfind では簡単なキーワードを与えると、それにマッチするエントリーのリストを返します。詳しくは man コマンドでオンラインマニュアルを参照してください。また DBGET は電子メールでも利用できます。電子メール検索の場合のアドレスは

```
dbget@genome.ad.jp
```

です。この場合の検索コマンドは bget, bfind に代わって get, find です。利用法は本文に help とだけ書いたメールを送ると返送されます。また DBGET はネットワーク上から gopher, WWW によっても利用できます。詳しくは 8 章をご覧ください。

より複雑な検索には、米国 NLM (National Library of Medicine) で作成された検索システム IRx と、米国 NCBI (National Center for Biotechnology Information) で開発された検索システム Entrez が利用できます。 IRx を利用できる端末はフルスクリーン制御ができる必要があり、また Entrez を利用する端末は Xwindow というウィンドウシステムが動いている必要があります。コマンドはそれぞれ irx と entrez です。 IRx, Entrez は HGC 以外に、直接インターネットに接続されたマシンからも以下のようにして利用することもできます。

```
rlogin irx.genome.ad.jp -l irx
rlogin entrez.genome.ad.jp -l entrez
```

データベース検索で利用可能なプログラムとデータベースは以下のとおりです。

プログラム	データベース
DBGET (bget, bfind)	GenBank, GenBank-today, EMBL, SwissProt, PIR, PRF, PDB, PDBSTR, PROSITE, OMIM, LIGAND, AAINDEX, EPD, PMD, LITDB
IRx	GenBank, EMBL, SwissProt, PIR, PRF, PDB, OMIM, LITDB, LIGAND, PMD, PROSITE
Entrez	Medline subset, Genbank, SwissProt/PIR/PRF/GenPept combined

GenBank-today は毎日アップデートされています。

利用可能なデータベースは `binfo {dbget|irx}` によって確認できます。

2　ホモロジー検索、モチーフ検索

BLAST と FASTA によるホモロジー検索ができます。まず最初はより高速な BLAST を試されることをお勧めします。これらの検索は重い仕事なので一時に多数の検索を行なわないでください。これらは電子メールによる検索もできますので、たくさんの問い合わせ配列で検索したいときは電子メール検索または待ち行列による制御のある qfasta をお使いください。電子メール検索の場合のアドレスは

```
blast@genome.ad.jp
fasta@genome.ad.jp
```

です。なおホモロジー検索のために、同一配列のエントリーの重複を除いた non-redundant (nr) データベースが用意されています。

核酸配列 (nr-nt)

以下のデータベースから完全に同一な配列を除いて作成されています。

genbank	GenBank の最新リリース
genbank-upd	GenBank の daily updates
embl	EMBL の最新リリース

タンパク質配列 (nr-aa)

以下のデータベースから完全に同一な配列を除いて作成されています。

swissprot	Swiss-Prot の最新リリース
pir	PIR の最新リリース
prf	PRF の最新リリース
genpept	GenBank 最新リリースの翻訳
genpept-upd	GenBank daily updates の翻訳

MOTIF はタンパク質の問い合わせ配列に特徴的な配列パターンが存在するかモチーフ検索を行ない、もしあれば機能部位情報と立体構造情報を PostScript でグラフィック表示します。これは電子メールからのみ利用できます。電子メール検索ではいずれも本文に help とだけ書いたメールを送ると、利用方法を返送します。

利用可能なプログラムとデータベースは以下のとおりです。

サーバー	プログラム	データベース
BLAST	blastp, blastn blastx, tblastn	nr-nt, GenBank, EMBL, nr-aa, SwissProt, PIR, PRF
FASTA	fasta, tfasta	nr-nt, GenBank, GenBank-upd, GenBank-subdivisions, EMBL, EMBL-subdivisions, nr-aa, SwissProt, PIR, PRF, PDBSTR
MOTIF	MotifFinder	PROSITE, MotifDic

nr-nt, nr-aa, GenBank-upd は毎日アップデートされています。

利用可能なデータベースは `binfo {fasta|blast}` で確認できます。

3 Anonymous FTP

インターネット上のマシンからは、 FTP により各種データベースやツールなどをダウンロードすることもできます。

```
ftp ftp.genome.ad.jp
```

ユーザー名を ftp または anonymous とし、パスワードに自分のメールアドレスを入力してください。現在、次のデータベースが利用可能です。

ディレクトリー	データベース
/db/hgc	(ヒトゲノム解析センター作成のデータベースシステム) Locus-In, ContigMaker, GNOME, Genomatica, HyperGenome, SMART, NetDBget 他
/db/genomenet	(ゲノムネットコミュニティ作成のデータベース) AAindex, Antigen, LIGAND, MotifDic, PMD, PRF 他
/db/ncbi	(NCBI repository に登録されたデータベース)
/db/ebi	(EBI archive に登録されたデータベース)
/db/gdb	(GDB が提供しているテキストファイル形式の GDB,OMIM データベース)
/pub/tools	(ソフトウェアツール) Macintosh をネットワーク接続する際に必要なツールを集めた HGC 配布キット、 Gopher, Mosaic, Network Entrez, GDB/accessor, BLAST, FASTA, Kinemage などのプログラム、他
/pub/doc	(利用方法などの文書) 利用の手引き、講習会資料、他
/pub/ncbi	(NCBI ツール群のコピー) entrez, toolbox (ASN.1 ToolKit) など

NCBI repository や EBI archive は研究者が提供するデータベースを集めたもので、以下のようなものがあります。

ACeDB, AAtDB, FlyBase, LISTA, ECD, dbEST, dbSTS などのゲノムデータベース
KABAT, hla, aids-db, pkinases, GCRDb など特定配列グループのデータベース
repbase, cpgisle, vector, TMbase, SRPDB など特殊領域の配列データベース
hamb, p53 などの変異データベース
EPD, TFD, Prosite, Blocks, Sbase, PRINTS, TransTerm などのモチーフデータベース
ENZYME, LIGAND 酵素データベース、 REBASE 制限酵素データベース
tRNA, BERLIN, rrna などの RNA 配列データベース
CarbBank 糖鎖データベース、 taxonomies 系統樹データベース
LiMB データベースのデータベース、他

NCBI repository および GDB, OMIM は毎週アップデートされています。

4 Gopher, WWW

前述の anonymous FTP およびデータベース検索プログラム DBGET, IRx, Entrez は Gopher を通しても利用できます。

```
gopher gopher.genome.ad.jp 70
```

また、DBGET は WWW からも利用できます。こちらは WWW のハイパーテキスト機能を利用して別のデータベースへのクロスリファレンスをたどったり、データベース間のリンク関係の模式図より検索したいデータベースを選べるなど、統合化データベースシステムになっています。

```
http://www.genome.ad.jp
```

5 データベースサービスとサポートレベル一覧

	BLAST	FASTA	MOTIF	DBGET	IRx	Entrez	FTP
GenBank	○	◎		◎	○	○	
EMBL	○	○		○	○		
nr-nt	◎	◎					
SwissProt	○	○	△	○	○	△	○
PIR	○	○	△	○	○	△	○
PRF	○	○		○	○	△	
nr-aa	◎	◎					
PDB			△	○	○		
PDBSTR		○		○			
PROSITE			○	○			○
MotifDic			○				○
LIGAND				○	○		○
PMD				○	○		○
AAINDEX				○			○
EPD				○			○
Medline				△		○	
LITDB				○	○		
OMIM				◎	◎		◎
GDB				△			◎
LinkDB				◎			

サポートレベル ◎　daily/weekly updates も含めて利用可能
サポートレベル ○　latest release のみが利用可能
サポートレベル △　アプリケーション中で統合化されて利用可能

注：DBGET には Gopher, WWW を含む。

（登録ユーザーで直接データベースにアクセスしたい方へ）
データベースは基本的に「/bio/db/ データベース名」というディレクトリ下にあります。