３．DBGET の利用法

目次

1 bget の使い方の基本 **2**

2 bfind を使ってエントリー名を調べる **3**

3 bfind のキーワード指定法：１つのまとまりとなる言葉はクオートで括る **4**

4 bfind のキーワード指定法：単語としての検索を明示する **4**

5 bget でホモロジー検索用に **FASTA** 形式の配列を得る **6**

6 bget の結果をファイルに保存する **6**

7 出力される結果を画面上で１頁ごとに止めて見る **7**

8 応用例：`bfind` と `bget` をパイプでつないでサブライブラリを作る 9

9 オンラインマニュアルの活用 **11**

10 メールサーバーの利用 **12**

11 `binfo` コマンド 13

12 `bman` コマンド 15

はじめに

bget, bfind は指定したデータベースより目的のエントリーを取り出したり、ごく単純なキーワード検索をするための最もシンプルなコマンドである。単純に作られているために手軽に使え、Unix の他のコマンド群などとも馴染みやすく、またこれらを組み合わせて使うことでサブライブラリー作成などにも使える。

講習内容

bget 指定されたデータベースより、指定されたエントリーを取り出す。

使い方： bget␣ データベース名 ␣ エントリー名...
-f を指定すると FASTA フォーマットで出力

bfind 指定されたデータベースのタイトル行から、与えられたキーワードにマッチするエントリーのリストを取る。

使い方： bfind␣ データベース名 ␣ キーワード...
キーワードは AND 結合
-w ワードオプション

一般的使用法

出力結果をファイルに保存	リダイレクション	> ファイル名
出力結果を次のコマンドに流し込む	パイプ	| コマンド名

応用例： bfind と bget の連結によるサブライブラリ作成

使用法の説明を見る　　man コマンド名

テキスト表記上の注意

通常の説明はこのような明朝体の字で書かれている。

本文中の説明で利用者が入力すべきコマンドなどは、枠で囲った中に

```
command␣strings␣you␣should␣type
```

のような強調タイプライター書体を用いて記す。

その結果表示される内容は、原則として、枠で囲った中に

```
Results of the command you've typed
```

のような小型タイプライター書体で示すものとする。

1 bget の使い方の基本

bget は指定されたデータベースより、指定されたエントリーを取り出す。使い方は、 bget の後にデータベース名とエントリー名 (複数でも可) を指定するだけ。

例： pir データベースより cchu という名前のエントリーを取る。

```
bget␣pir␣cchu
```

とタイプすると端末に以下のような結果が出る。

```
ENTRY            CCHU       #type complete
TITLE            cytochrome c - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             #sequence_revision 30-Sep-1991 #text_change 05-Aug-1994
ACCESSIONS       A31764; A05676; A00001
REFERENCE        A31764
   #authors      Evans, M.J.; Scarpulla, R.C.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1988) 85:9625-9629
   #title        The human somatic cytochrome c gene: two classes of processed
                   pseudogenes demarcate a period of rapid molecular
                   evolution.
   #cross-references MUID:89071748
   #accession    A31764
      ##molecule_type DNA
      ##residues      1-105 ##label EVA
      ##cross-references GB:M22877
REFERENCE        A05676
   #authors      Matsubara, H.; Smith, E.L.
   #journal      J. Biol. Chem. (1963) 238:2732-2753
   #title        Human heart cytochrome c. Chymotryptic peptides, tryptic
                   peptides, and the complete amino acid sequence.
   #accession    A05676
      ##molecule_type protein
      ##residues      2-28;29-46;47-100;101-105 ##label MATS
REFERENCE        A00001
   #authors      Matsubara, H.; Smith, E.L.
   #journal      J. Biol. Chem. (1962) 237:3575-3576
   #title        The amino acid sequence of human heart cytochrome c.
   #contents     annotation
   #note         66-Leu is found in 10% of the molecules in pooled protein
GENETICS
   #introns      57/1
CLASSIFICATION   #superfamily cytochrome c; cytochrome c homology
KEYWORDS         acetylated amino end; electron transfer; heme; mitochondrion;
                   oxidative phosphorylation; polymorphism; respiratory chain
FEATURE
   2-105             #protein cytochrome c #status experimental #label MAT\
   2                 #modified_site acetylated amino end (Gly) (in mature
                       form) #status experimental\
   15,18             #binding_site heme (Cys) (covalent) #status
                       experimental\
   19,81             #binding_site heme iron (His, Met) (axial ligands)
                       #status predicted
SUMMARY          #length 105  #molecular-weight 11749  #checksum 3247
SEQUENCE
               5        10        15        20        25        30
     1 M G D V E K G K K I F I M K C S Q C H T V E K G G K H K T G
    31 P N L H G L F G R K T G Q A P G Y S Y T A A N K N K G I I W
    61 G E D T L M E Y L E N P K K Y I P G T K M I F V G I K K K E
    91 E R A D L I A Y L K K A T N E
///
```

このようにデータベースのオリジナルフォーマットで結果が得られる。但し、このためにはエントリー名を知らないといけない。

2　bfind を使ってエントリー名を調べる

何々についてのエントリーを取り出したいというような場合、 bget で指定するためのエントリー名を調べるのに bfind が使える。 bfind の後にデータベース名を付け続けてキーワードを指定する。キーワードはスペースで区切って複数指定可能。この場合、各キーワードの AND 結合が検索される。

例：SWISS-PROT より human の interleukin に関するエントリーリストが欲しい。

```
bfind␣swissprot␣human␣interleukin
```

とタイプすると、以下のようなリストが打ち出される。

```
CYRG_HUMAN [P31785] CYTOKINE RECEPTOR COMMON GAMMA CHAIN PRECURSOR (GAMMA-C) (IN
TERLEUKIN-2 RECEPTOR GAMMA CHAIN) (IL-2R GAMMA CHAIN).
I12A_HUMAN [P29459] INTERLEUKIN-12 ALPHA CHAIN PRECURSOR (IL-12A) (CYTOTOXIC LYM
PHOCYTE MATURATION FACTOR 35 KD SUBUNIT) (CLMF P35) (NK CELL STIMULATORY FACTOR
CHAIN 1) (NKSF1).
I12B_HUMAN [P29460] INTERLEUKIN-12 BETA CHAIN PRECURSOR (IL-12B) (CYTOTOXIC LYMP
HOCYTE MATURATION FACTOR 40 KD SUBUNIT) (CLMF P40) (NK CELL STIMULATORY FACTOR C
HAIN 2) (NKSF2).
I1BC_HUMAN [P29466] INTERLEUKIN-1 BETA CONVERTASE PRECURSOR (IL-1BC) (EC 3.4.22.
36) (IL-1 BETA CONVERTING ENZYME) (ICE) (INTERLEUKIN-1 BETA CONVERTING ENZYME).
IL10_HUMAN [P22301] INTERLEUKIN-10 PRECURSOR (IL-10) (CYTOKINE SYNTHESIS INHIBIT
ORY FACTOR) (CSIF).
IL11_HUMAN [P20809] INTERLEUKIN-11 PRECURSOR (IL-11) (ADIPOGENESIS INHIBITORY FA
CTOR) (AGIF).
IL13_HUMAN [P35225] INTERLEUKIN-13 PRECURSOR (IL-13).
IL1A_HUMAN [P01583] INTERLEUKIN-1 ALPHA PRECURSOR (IL-1 ALPHA) (HEMATOPOIETIN-1).
IL1B_HUMAN [P01584] INTERLEUKIN-1 BETA PRECURSOR (IL-1 BETA).
IL1R_HUMAN [P14778] INTERLEUKIN-1 RECEPTOR, TYPE I PRECURSOR (IL-1R-1) (IL-1R-AL
PHA) (P80).
IL1S_HUMAN [P27930] INTERLEUKIN-1 RECEPTOR, TYPE II PRECURSOR (IL-1R-2) (IL-1R-B
ETA).
IL1X_HUMAN [P18510] INTERLEUKIN-1 RECEPTOR ANTAGONIST PROTEIN PRECURSOR (IL-1RA)
 (ICIL- 1RA) (IRAP).
IL2A_HUMAN [P01589] INTERLEUKIN-2 RECEPTOR ALPHA CHAIN PRECURSOR (IL-2 RECEPTOR
ALPHA SUBUNIT) (P55) (TAC ANTIGEN) (CD25).
IL2B_HUMAN [P14784] INTERLEUKIN-2 RECEPTOR BETA CHAIN PRECURSOR (IL-2 RECEPTOR)
(P70-75) (HIGH AFFINITY IL-2 RECEPTOR BETA SUBUNIT) (CD122).
IL2_HUMAN  [P01585] INTERLEUKIN-2 PRECURSOR (IL-2) (T-CELL GROWTH FACTOR) (TCGF).
IL3R_HUMAN [P26951] INTERLEUKIN-3 RECEPTOR ALPHA CHAIN PRECURSOR (IL-3R).
IL3_HUMAN  [P08700] INTERLEUKIN-3 PRECURSOR (IL-3) (MULTIPOTENTIAL COLONY-STIMUL
ATING FACTOR) (HEMATOPOIETIC GROWTH FACTOR) (P-CELL STIMULATING FACTOR) (MAST-CE
LL GROWTH FACTOR).
IL4R_HUMAN [P24394] INTERLEUKIN-4 RECEPTOR ALPHA CHAIN PRECURSOR (IL-4R) (CDW124
).
IL4_HUMAN  [P05112] INTERLEUKIN-4 PRECURSOR (IL-4) (B-CELL STIMULATORY FACTOR 1)
 (BSF-1) (LYMPHOCYTE STIMULATORY FACTOR 1).
IL5R_HUMAN [Q01344] INTERLEUKIN-5 RECEPTOR ALPHA CHAIN PRECURSOR (IL-5R-ALPHA).
IL5_HUMAN  [P05113] INTERLEUKIN-5 PRECURSOR (IL-5) (T-CELL REPLACING FACTOR) (TR
F) (EOSINOPHIL DIFFERENTIATION FACTOR).
IL6R_HUMAN [P08887] INTERLEUKIN-6 RECEPTOR PRECURSOR (IL-6R) (CD126).
IL6_HUMAN  [P05231] INTERLEUKIN-6 PRECURSOR (IL-6) (B-CELL STIMULATORY FACTOR 2)
 (BSF-2) (INTERFERON BETA-2) (HYBRIDOMA GROWTH FACTOR).
IL7R_HUMAN [P16871] INTERLEUKIN-7 RECEPTOR ALPHA CHAIN PRECURSOR (IL-7R) (CDW127).
IL7_HUMAN  [P13232] INTERLEUKIN-7 PRECURSOR (IL-7).
(以下略)
```

リストは 1 行に 1 エントリーで、行の先頭から最初の空白までがエントリー名、以下がタイトルの内容となっている。

3　bfind のキーワード指定法：１つのまとまりとなる言葉はクオートで括る

複数の単語で１つの概念となっているような成句は、それ全体をシングルクオート（'）で囲うことで、別々の単語の AND としてではなく、１つの言葉として検索される。

例：GenBank より Interleukin 6 についてのエントリーを探す。

```
bfind␣genbank␣'interleukin␣6'
```

結果

```
CEYIN6A    [L26032] Cercocebus torquatus interleukin 6 mRNA, complete cds.
HUMIFNB2   [M14584] Human interleukin 6 mRNA, complete cds.
HUMIFNB2A  [J03049] Human interleukin 6 (IL6) gene, 3' flank.
HUMIFNB2B  [M18403] Human hybridoma growth factor (interleukin 6) mRNA, complete
 cds.
HUMIL6     [M20566] Human interleukin 6 receptor mRNA, complete cds.
HUMIL6B    [M22111] Human interleukin 6 (IL 6) gene, 5' flank.
HUMIL6C    [M29150] Human interleukin 6 (B-cell stimulatory factor-2) mRNA, comp
lete cds.
HUMIL6CSF  [M54894] Human interleukin 6 mRNA, complete cds.
MACIN6A    [L26028] Macaca mulatta interleukin 6 mRNA, complete cds.
MM3IL6     [X51457] Mouse DNA for 3' IL-6 intracisternal A particle (IAP of IL-6
) and interleukin 6, exon 1.
MMIL65     [X51458] Mouse (MPC11) DNA for interleukin 6 (IL-6) 5' flanking regio
n.
MMIL6R     [X53802] Mouse mRNA for interleukin 6 receptor.
MMILHP1    [X06203] Mouse mRNA for interleukin-HP1 (homologue to interleukin 6).
MUSGP130A  [M83336] Mouse interleukin 6 signal transducer (gp130) mRNA, complete
 cds.
MUSIL6     [J03783] Mouse interleukin 6 mRNA, complete cds.
MUSIL6A    [M20572] Mouse interleukin 6 (IL-6) gene, complete cds.
RATGP130A  [M92340] Rat (clones rLG[08,14,25]) interleukin 6 signal transducer m
RNA sequence.
RATIL6     [M26744] Rat interleukin 6 (IL6) mRNA, complete cds.
RATIL6G    [M26745] Rat interleukin 6 (IL6) gene, complete cds.
RATIL6R    [M58587] Rat liver interleukin 6 receptor ligand binding chain mRNA,
complete cds.
BTINTLK6P  [Z11749] B.taurus DNA for interleukin 6 gene 5' flank (promoter).
CATIL6     [D13227] Cat mRMA for interleukin 6, complete cds.
CATINLK6A  [L16914] Felis catus interleukin 6 mRNA, complete cds.
OAINT6A    [X62501] O.aries mRNA for interleukin 6.
PIGIL6A    [M80258] Sus scrofa interleukin 6 (IL-6) mRNA, complete cds.
PIGPRO     [M86722] Sus scrofa prointerleukin 6 mRNA, complete cds.
A00343     [A00343] Artificial gene for human interleukin 6.
A00344     [A00344] Artificial gene for human interleukin 6, reverse complement.
```

注意 bfind では大文字小文字を区別しないので、キーワードとして指定するときは Interleukin でも interleukin でも構わない。（あえて大文字小文字を区別して検索したいときは、bfind のあとに -C を付けるとよい。）

4　bfind のキーワード指定法：単語としての検索を明示する

キーワードは通常はそれを部分文字列として含むもの全てにマッチする。部分文字列ではなく単語としてマッチしたものに限りたい場合には、-w を付ける。

例：fos についてのエントリーを得ようとして以下のようにすると...

```
bfind␣embl␣fos
```

結果：　もちろん目的の fos も拾われるのだが、

```
HS1514     [T17151] NIB457 Homo sapiens cDNA 3'end similar to Human v-fos transf
ormation effector protein (Fte-1).
HS5156     [T16515] NIB1418 Homo sapiens cDNA 3'end similar to Human cellular on
cogene c-fos.
T09381     [T09381] EST07274 Homo sapiens cDNA clone HIBBT54 5' end similar to c
-fos.
FOSTI31Y   [M33643] F.oxysporum STI35 protein mRNA, complete cds.
(途中略)
MFITS      [X07623] Loach (M. fossilis L.) rDNA for internal transcribed spacer
MFPMF2051  [X56641] M. fossilis 5S rRNA gene from clone pMf205.1.
MFPMF2052  [X56642] M. fossilis 5S rRNA gene from clone pMf205.2.
MFPMF206   [X56633] M. fossilis 5S rRNA gene from clone pMf206.
(以下略)
```

このように、 fossilis と余分のものも拾われてしまう。特にキーワードの文字列長が短いときなどは、こういった目的外の結果がかなり多くなりやすい。このようなときは -w （ワードオプション）を付ける。こうすると fos という単語を持つものだけを検索する。

```
bfind␣embl␣-w␣fos
```

```
HS1514     [T17151] NIB457 Homo sapiens cDNA 3'end similar to Human v-fos transf
ormation effector protein (Fte-1).
HS5156     [T16515] NIB1418 Homo sapiens cDNA 3'end similar to Human cellular on
cogene c-fos.
T09381     [T09381] EST07274 Homo sapiens cDNA clone HIBBT54 5' end similar to c
-fos.
DSDFRA     [X54143] Drosophila dFRA mRNA for Fos-related antigen
HSCFOS     [V01512] Human cellular oncogene c-fos (complete sequence)
HSCFOS01   [M15429] Human c-fos proto-oncogene regulatory elements.
HSCFOSRE   [X53723] Human c-fos upstream region with estrogen response element
HSFOS      [K00650] Human fos proto-oncogene (c-fos), complete cds.
HSFTE1A    [M84711] Human v-fos transformation effector protein (Fte-1), mRNA co
mplete cds.
S65138     [S65138] c-fos=FOS, p55 [human, mRNA Partial, 47 nt, segment 1 of 2]
S65140     [S65140] c-fos=FOS, p55 [human, mRNA Partial, 61 nt, segment 2 of 2]
TTFOS      [M85195] Transposon Tn2921 (from Serratia marcescens) fosfomycin-resi
stance protein (fos) gene, complete cds.
MMCFOS     [V00727] Mouse c-fos oncogene
MMCFOS01   [D12551] Mouse c-fos gene, 3'UTR.
MMCFOS5F   [M23768] Mouse c-fos gene, 5' flank.
MMFOS      [J00370] Mouse c-fos gene; cellular homolog to viral oncogene.
MMFOSA     [K03231] Mouse c-fos proto-oncogene, 5' flank.
MMFOSERE   [M65003] Mouse c-fos gene, 3' flanking region.
MMFOSR     [K02785] Mouse mRNA induced by PDGF with some homology to c-fos.
RNO2631    [U02631] Rattus norvegicus Sprague Dawley (c-fos) gene, 5' flanking r
egion.
RNCFOSR    [X06769] Rat c-fos mRNA
RNVFOSPR   [M84716] R.norvegicus putative v-fos transformation effector protein
(Fte-1) mRNA, complete cds.
RRFRA1     [M19651] Rat fos-related antigen (Fra-1) mRNA, complete cds.
RRWBFOS    [M34001] Rat wb-Fos protein gene, 5' end.
S60557     [S60557] c-fos gene homolog {polymorphic allele} [mice, C57BL/6J, Gen
omic, 132 nt]
M28857     [M28857] Figure 3. The 3' untranslated sequence of the (A) IFN-beta-c
-fos and (B) GM-CSF mRNAs.
M28858     [M28858] Figure 3. The 3' untranslated sequence of the (A) IFN-beta-c
-fos and (B) GM-CSF mRNAs.
M28859     [M28859] Figure 3. The 3' untranslated sequence of the (A) IFN-beta-c
-fos and (B) GM-CSF mRNAs.
M28860     [M28860] Figure 3. The 3' untranslated sequence of the (A) IFN-beta-c
-fos and (B) GM-CSF mRNAs.
REAC2NK2   [M18041] Avian transforming proviral retrovirus NK24 v-fos gene, part
ial cds.
REMSV5     [V01184] Provirus of a replication defective murine sarcoma virus (FB
J-MuSV) with c-fos(p55) and p15 E reading frames.
GDFOS      [Y00659] Chicken fos proto-oncogene mRNA
GGCFOS     [M18043] Chicken c-fos proto-oncogene, complete cds.
GGCFOSA    [M37000] Chicken c-fos proto-oncogene mRNA, complete cds.
```

5　bget でホモロジー検索用に FASTA 形式の配列を得る

bget そのままだとオリジナルの書式のエントリーが得られたが、この時 -f オプションを付けると、配列のみをホモロジー検索用に FASTA フォーマットで得ることができる。

例：PIR データベースより CBBO という配列 (ウシの Cytochrome b) を FASTA 形式で得る。

```
bget␣pir␣-f␣cbbo
```

結果：

```
>CBBO (pir)
MTNIRKSHPLMKIVNNAFIDLPAPSNISSWWNFGSLLGICLILQILTGLFLAMHYTSDTT
TAFSSVTHICRDVNYGWIIRYMHANGASMFFICLYMHVGRGLYYGSYTFLETWNIGVILL
LTVMATAFMGYVLPWGQMSFWGATVITNLLSAIPYIGTNLVEWIWGGFSVDKATLTRFFA
FHFILPFIIMAIAMVHLLFLHETGSNNPTGISSDVDKIPFHPYYTIKDILGALLLILALM
LLVLFAPDLLGDPDNYTPANPLNTPPHIKPEWYFLFAYAILRSIPNKLGGVLALAFSILI
LALIPLLHTSKQRSMMFRPLSQCLFWALVADLLTLTWIGGQPVEHPYITIGQLASVLYFL
LILVLMPTAGTIENKLLKW
```

このように、 FASTA 形式は先頭行が > で始まり、この行にはエントリーの名前とコメントが書かれる。以下の行が 1 文字コードで書かれた配列となっており、ここでのスペース、改行は任意である (配列の内容に影響しない) 。

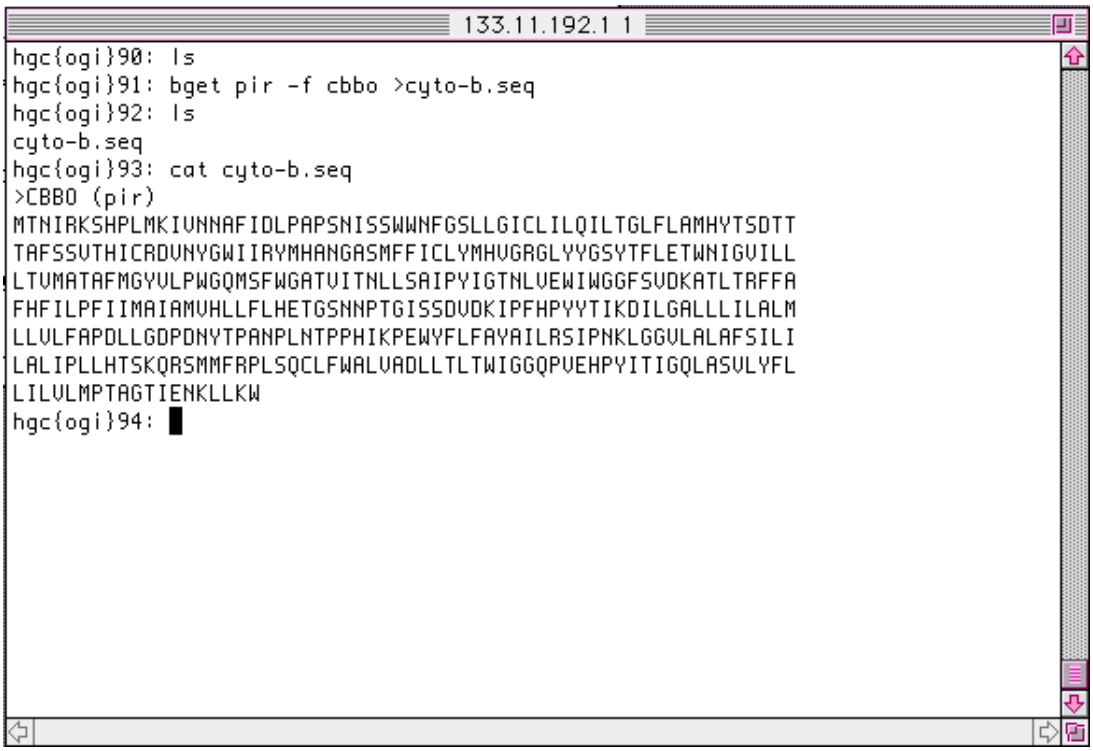
6　bget の結果をファイルに保存する

今までの例だと、実行結果は端末に出力されてそのままでは利用できなかった。多くの場合、結果はファイルに格納されて別の目的に利用される。ファイルに格納するには、「bget データベース名　エントリー名」に続けて、「 > 格納ファイル名」と指定する。格納用のファイルは与えられた名前で新規に作られて [*1]、端末に出てくる内容がそのままファイルにはいる [*2]。格納ファイルの名前は自由に決めてよいが、ファイル名の中に空白を含むことはできない。また一般に記号は意味を持つことがあるので使わないほうがよい [*3]。なお、このように通常端末に出力される結果をファイルに収めることをリダイレクションといい、 bget, bfind だけでなく通常の UNIX コマンドでも同様に行える。

例：直前の例の PIR CBBP エントリーの FASTA 形式配列を、ファイル cyto-b.seq に収める。

```
bget␣pir␣-f␣cbbo␣>␣cyto-b.seq
```

この結果は端末には現われない。ただ、ファイルリストを見るコマンド ls を打つと、確かに新しいファイル cyto-b.seq ができていることが分かる。ファイルの内容を確認するために cat cyto-b.seq と打つと、直前の例と同じ内容がファイルに入っていることが分かる。



注意

＊1　もし同名のファイルが既に存在する場合、 hgc の設定では不注意な上書き防止のために File exist のメッセージを出してこの操作は失敗する。この時はファイルを消去する rm コマンドでそのファイルを消してから再度実行する。

＊2　厳密には、端末に出てくる結果には 2 種類あり、 UNIX ではこれを「標準出力」と「標準エラー」という名前で区別する。大まかに言って前者がそのコマンドが正常に実行されたときの結果、後者がエラーメッセージに当たる。 > のリダイレクトでは標準出力だけが取り込まれる。

＊3　通常のファイル名は英文字と数字から構成され、必要に応じて．（ピリオド）で区切られた拡張子などを付けてもよい。このような拡張子は例えばC言語のプログラムならば .c というように意味を持つことが多い。また、/（スラッシュ）はディレクトリ階層を区切るために使う。~（ティルダ）を先頭に付けると、それはホームディレクトリを表すため、自分のホームにファイルを作ることを明示したい場合に便利である（例えば ~/myfile のように使う）。複数の単語からなるファイル名を付けたいとき、-（ハイフン）を言葉のつなぎに使ってもよいが（e.g.　cyto-b）ファイル名の先頭には使わないこと。 UNIX では大文字小文字を区別するので例えば HumanInterleukin というような付け方もある。_（アンダースコア）は余り意味を持たないので human_inerleukin のような名前も構わない。そのほかの記号は意味を持つことが多いので、使わないほうがよい。

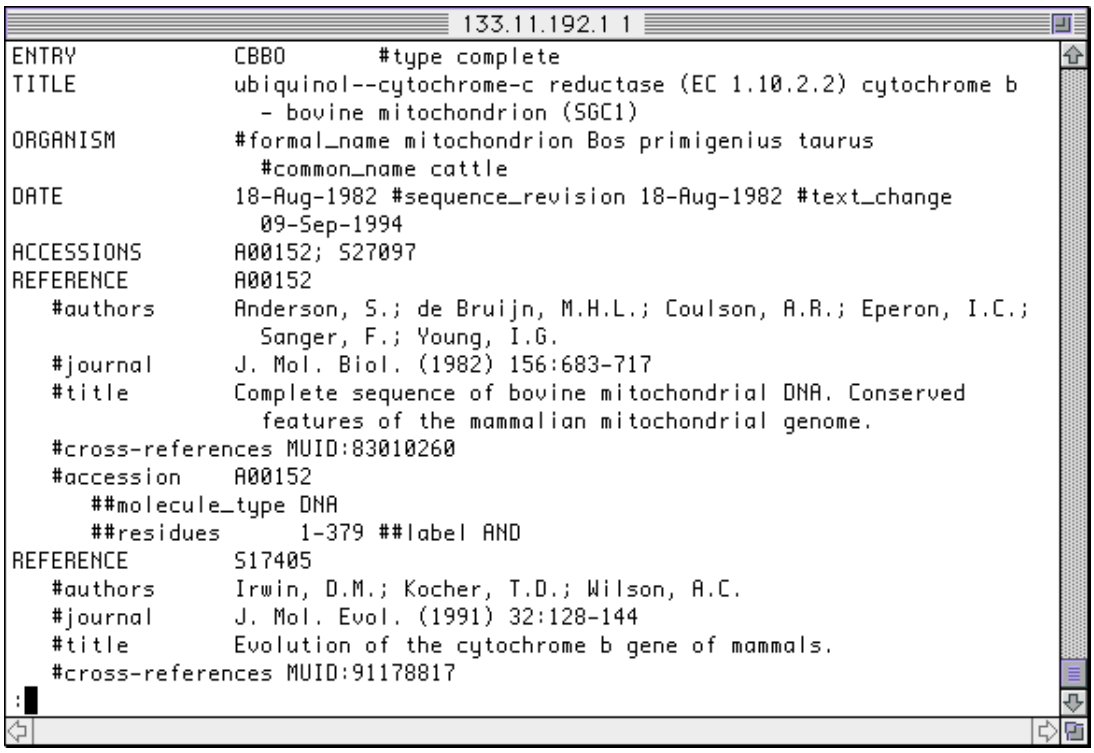
7　出力される結果を画面上で 1 頁ごとに止めて見る

わざわざファイルにいれる必要はないが、多量の結果が画面に流れてしまい目で追うことができないというときには、 less と言うコマンドにその結果を入れてやると、 1 頁表示したところで止って次の動作を利用者が指示するのを待つ。

例： bget pir cbbo の結果を 1 頁ごとに止めて表示させる。

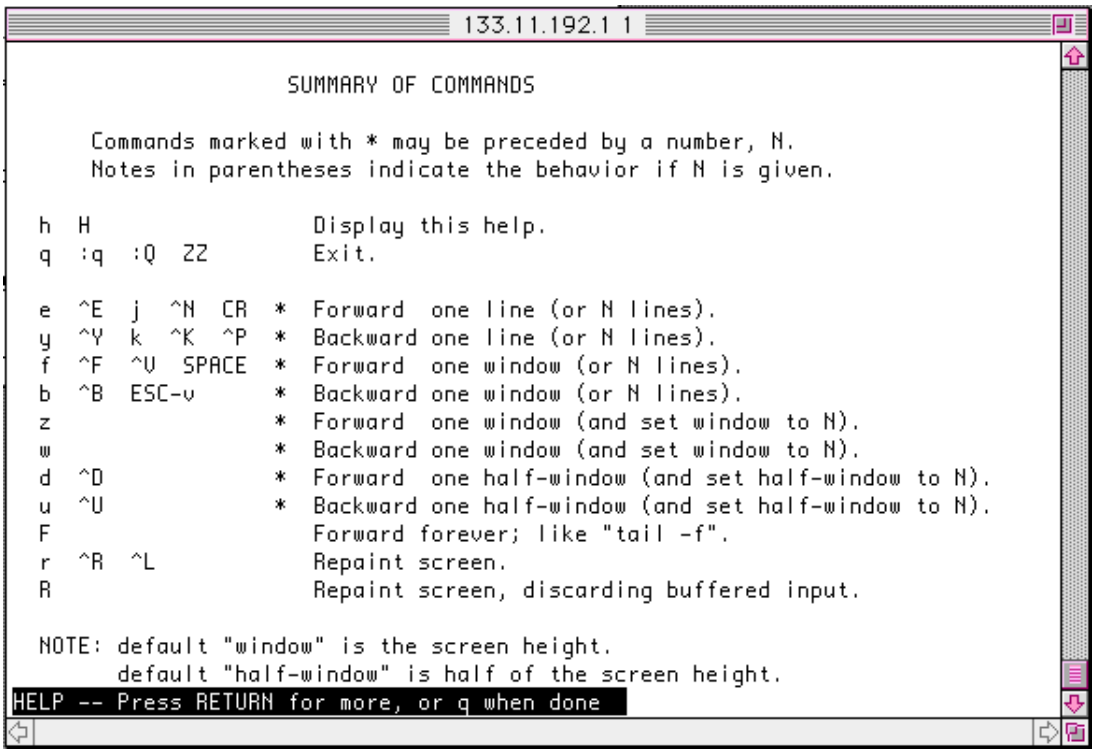
```
bget␣pir␣cbbo␣|␣less
```

結果：



このように先頭から 1 頁表示したところで停止し、最下行に：を表示してユーザーの入力を待つ。ここで スペースバー を叩くと (リターン不要) 、次の 1 頁を表示する。また b を押すと直前の頁に戻る。 G で最終頁に行き、 g で先頭頁に戻る。 less から抜ける場合には q を押す。このほかの説明は h キーを押すとヘルプメッセージが表示される。 less の中ではさらに文字列検索やコマンド呼び出しも可能。

なお、ここの例のように、前のコマンド（bget）の出力結果を次のコマンド（less）の入力に入れることを「パイプ」と言い、 | (縦棒) で前後のコマンドの区切れを指示する。この機能もリダイレクトと同様、一般の UNIX コマンドで行える。



less で h キーを押してヘルプを表示させたところ

8 応用例：bfind と bget をパイプでつないでサブライブラリを作る

今までの例では bget は常にデータベース名と１個以上のエントリー名を付けて呼び出されていたが、bget はエントリー名を付けずに呼び出すことも可能で、その場合は端末（標準入力）からエントリー名を読み込もうとする。この場合１行に１エントリーずつで、各行の先頭から１０文字までがエントリー名と見なされる。これを利用して bfind の結果を直接 bget に入れることもできる。つまり、あるキーワードを持つエントリーだけを抜き出したサブライブラリが作れる。

例：GenBank より growth factor と言う言葉を持つエントリーを抜き出して FASTA 形式のサブライブラリーファイル growth-factor を作る。

```
bfind␣genbank␣growth␣factor␣|␣bget␣genbank␣-f␣>␣growth-factor
```

これをホモロジーサーチプログラム fasta （詳細は別講で述べる）の検索対象ライブラリとすることが可能で、特定のグループに限って検索を何度も行いたいような場合には検索時間が大幅に節約できる。

注意 サブライブラリを作るときはキーワードの選択に注意しないと極めて大きなものができてしまって利用者用ファイルとして格納できなくなることもある。

サブライブラリを使ったホモロジー検索の例：ヒトの *fos* 遺伝子配列を収めた HUMFOS エントリーを GenBank より取り出し、さきほどの growth-facotr サブライブラリに対してホモロジーサーチをかける。

```
hgc{ogi}82: bget␣genbank␣-f␣humfos␣>fos
hgc{ogi}83: fasta␣fos␣growth-factor
 >HUMFOS (genbank) : 6210 nt
 vs  library
 searching growth-factor library


     initn   init1
<  4     1     1:=
   8     0     0:
  12     0     0:
  16     0     0:
  20     0     0:
  24     0     0:
  28     1     1:=
  32     9     9:=====
  36    24    24:============
  40    58    58:============================
  44    81    81:========================================
  48   110   110:==================================================
  52   132   132:==================================================
  56    98    98:================================================
  60    96    96:================================================
  64    62    62:==============================
  68    30    30:===============
  72    22    22:===========
  76    16    16:========
  80    15    15:========
  84    12    12:======
  88     1     1:=
  92     3     3:==
（途中略）
>160    12    12:======
1147715 residues in   783 sequences
 statistics exclude scores greater than 170
 mean initn score:  53.1 (10.88)
 mean init1 score:  53.1 (10.88)
  783 scores better than 1 saved, ktup: 6, fact: 6
 joining threshold: 95  scan time:  0:00:12
 Enter filename for results : out
 How many scores would you like to see? [20]
The best scores are:                                      initn init1   opt
HUMTDGF1A (genbank)                                         817   257   735
HUMPDGFAA (genbank)                                         668   293   604
HSBNGF (genbank)                                            569   323   653
HUMBCGF (genbank)                                           477   356   733
HSBFR (genbank)                                             438   232   757
HUMFGF4H (genbank)                                          391   299   561
HUMGFB (genbank)                                            373   240   618
HUMFGFB (genbank)                                           359   233   604
（途中略）
 More scores? [0]
 Display alignments also? y
 number of alignments [20]?
Library scan:  0:00:12  total CPU time:  0:00:20
hgc{ogi}84:
```

9　オンラインマニュアルの活用

UNIX ではほとんどのコマンドのリファレンスマニュアルはオンラインマニュアル参照コマンド man を使って読むことができる。bget, bfind なども詳しい使用法はオンラインマニュアルにあるので参照されたい。

例：

```
man␣bfind
```

結果：(less が呼び出される)

```
BFIND(1L)                USER COMMANDS                BFIND(1L)



NAME
     bfind - find entries in a database with  matching  character
     patterns in the definition line

SYNOPSIS
     bfind dbname [ -Cnow ] pattern ...

DESCRIPTION
     bfind searches for sequence entries of a given database with
     matching  character  patterns  in  the definition line.  The
     definition line usually contains a brief description of  the
     content  of  each  entry.   The  argument  dbname  may  be a
     reserved keyword for the database.  Currently available pub-
     lic databases are:

          dbname        database
/tmp/man5572
```

(参考) bget, bfind で利用可能なデータベース

dbname	database
genbank	GenBank nucleic acid sequence database
genbank-today	Current GenBank including daily updates
embl	EMBL nucleic acid sequence database
pir	PIR protein sequence database
nbrf	(Same as above)
swissprot	SWISS-PROT protein sequence database
prf	PRF protein sequence database
pdb	Brookhaven Protein Data Bank
pdbstr	Re-organized Protein Data Bank
prosite	Dictionary of protein sites and patterns
omim	Online Mendelian Inheritance in Man
ligand	Ligand Chemical Database for Enzyme Reaction
litdb	Literature Database compiled by PRF
pmd	Protein Mutant Database
aaindex	Amino Acid Index Database

10　メールサーバーの利用

以上説明した bget, bfind は HGC 内部のみで使えるものであるが、これと同一の機能のものを電子メール経由で HGC 利用資格をもっていない不特定利用者に対して開放しているものもある。

まず、以下のアドレスに対して本文に

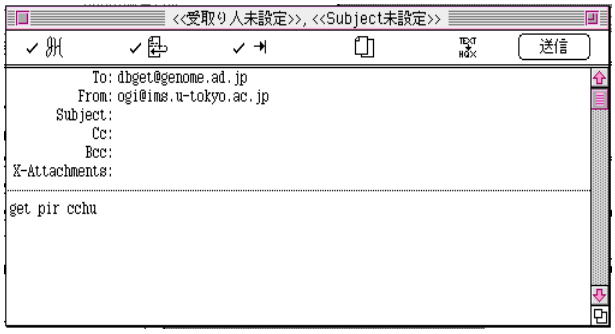
help

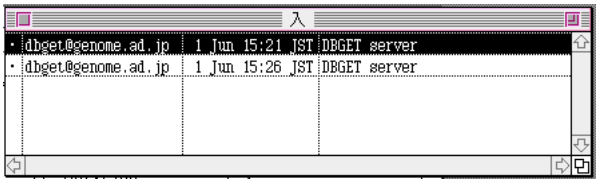
とだけ書いたメールを出してみること（電子メールの使い方は別講で説明する）。そうするとメールサーバーの利用法の説明がメールを出したところに戻ってくるようになっている。

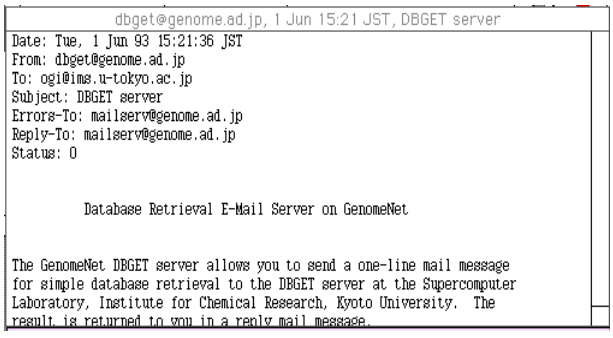
dbget メールサーバーのアドレス：　　dbget@genome.ad.jp

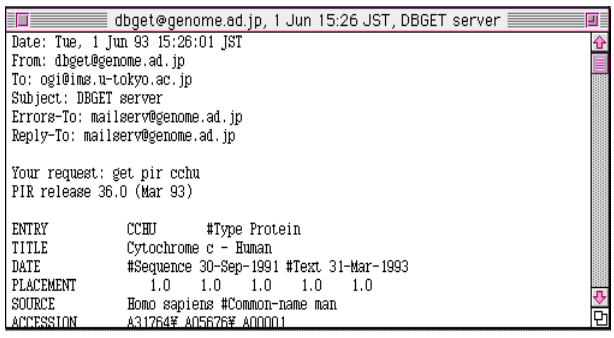
利用法は、基本的にコマンドの bget, bfind と同様である。ただし、コマンドの方は bget / bfind のように先頭に b が付いたのに対し、メールサーバーでの問い合わせでは b を付けずに get / find とだけ書くことに注意。

Macintosh Eudora 電子メールツールによる dbget 利用の例









11 binfo コマンド

binfo コマンドを使用すると、現在サービスされているデータベースの一覧とそのバージョンなどに関する情報が参照できる。

```
binfo
```

```
                *** Last database updates ***
   Date     Database          Release          #Entries  #Residues
   ----     ------------   -----------------   -------   ----------
95/01/19   nr-nt          95-01-19 (Jan 95)  210,385  215,451,895
95/01/19   nr-aa          95-01-19 (Jan 95)  149,971   43,312,391
94/11/09   genbank             85.0 (Oct 94)  215,273  217,102,462
95/01/19   genbank-today      85.0+ (Jan 95)  191,683  191,131,147
95/01/19   genbank-upd        85.0+ (Jan 95)
94/12/03   embl                40.0 (Sep 94)  212,744  212,015,482
94/11/02   pir                 42.0 (Sep 94)   71,995   21,244,777
94/10/27   swissprot           30.0 (Oct 94)   40,292   14,147,368
94/12/07   prf                94-11 (Nov 94)   57,526   19,781,002
94/11/15   pdb                 69.0 (Jul 94)    2,785      727,546
94/11/15   pdbstr              69.0 (Jul 94)    4,302      803,892
94/10/29   prosite             12.1 (Oct 94)    1,029
94/12/29   ligand              11.0 (Dec 94)    3,801
94/12/25   litdb              20-22 (Dec 94)  216,795
94/07/04   omim              MIM10+ (Jul 94)    6,763
94/10/19   pmd                94-10 (Oct 94)    5,597       89,144
95/01/13   aaindex              2.1 (Jan 95)      402
```

binfo に続けて特定のデータベース名を指定すると (例： binfo genbank) 、そのデータベースだけに関する情報が参照できる。また binfo に続けてサービスの種類を表すキーワード、 dbget, fasta, blast または irx を付けると、そのサービスの対象となっているデータベースだけに関する情報が得られる。

また -d オプションをつけると次のような書式で表示される。

```
nr-nt           Non-redundant nucleic acid sequence database
                Release 95-01-19, Jan 95
                210,385 entries, 215,451,895 bases
                Last update:  95/01/19
                <fasta> <blast>

nr-aa           Non-redundant protein sequence database
                Release 95-01-19, Jan 95
                149,971 entries, 43,312,391 residues
                Last update:  95/01/19
                <fasta> <blast>

genbank         GenBank nucleic acid sequence database
                Release 85.0, Oct 94
                National Center for Biotechnology Information
                215,273 entries, 217,102,462 bases
                Last update:  94/11/09
                <dbget> <fasta> <blast> <irx>

genbank-today   Latest GenBank quarterly release plus cumulative daily updates
                Release 85.0+, Jan 95
                National Center for Biotechnology Information
                191,683 entries, 191,131,147 bases
                Last update:  95/01/19
                <dbget>

genbank-upd     Cumulative daily updates of GenBank
                Release 85.0+, Jan 95
                National Center for Biotechnology Information
                Last update:  95/01/19
                <fasta>

（以下略）
```

12 bman コマンド

bman コマンドを使用すると指定したデータベースのリリースノートを見ることができる。使用方法は bman のあとにデータベース名を指定する。

```
bman␣genbank
```

```
GBREL.TXT          Genetic Sequence Data Bank
                         15 October 1994

               NCBI-GenBank Flat File Release 85.0

                Distribution CD-ROM Release Notes

  215273 loci, 217102462 bases, from 215273 reported sequences

  This document describes the data written on GenBank flat file
distribution CD-ROMs. If you have any questions or comments about the
data bank, the CD-ROM, or this document, please contact NCBI via email at
info@ncbi.nlm.nih.gov or:

   GenBank
   National Center for Biotechnology Information
   National Library of Medicine, 38A, 8N805
   8600 Rockville Pike
   Bethesda, MD  20894
   USA
   Phone:  (301) 496-2475
   Fax:    (301) 480-9241


(以下略)
```

現在、 bman コマンドの対象となるデータベースには以下のものがある。

genbank, embl, swissprot, pir, prf, pdb, pdbstr,
prosite, omim, aaindex, pmd, litdb, ligand