４．IRx の利用法

目次

1 IRx の起動方法 1

2 GenInfo メインメニューの表示 2

3 IRx データベース選択画面 3

4 問い合わせキーワードの入力：単純なキーワードの例 3

5 問い合わせ結果のファイルへの書き出し 5

6 終了の方法 6

7 HGC 外部から無記名利用資格で IRx を利用する 8

8 応用−１ キーワードの入力：複数キーワード、同一フィールド内限定 11

9 応用−２ キーワードの入力： 検索フィールドの限定 12

10 応用−３ キーワードの入力： AND 条件の指定 13

11 応用−４ キーワードの入力： stemming の抑止 14

はじめに

IRx はアメリカ National Library of Medicine (NLM) で作成された情報検索ツール で、強力な indexing によってキーワード検索が極めて高速に行えるように なっている。また、フルスクリーン画面によって検索作業が対話的に使いよくなっている。

講習内容

- 起動から終了までの操作、ファイルへの保存法。
- HGC 外部からの無記名利用資格での使い方。
- キーワードの指定：単に列挙するだけ。 OR 結合
- [応用] 検索条件 (フィールド指定、同一フィールド内) 、
  論理演算子 (AND) 、特殊検索記号 (@,*) 等。

1 IRx の起動方法

IRx ではフルスクリーン制御を使ったメニュー形式で動くため、その端末がフルスクリーン対応していないといけない。ただ、ほとんどのディスプレイ端末は利用できる。例えば、 Sun ならば kterm, shelltool 等、 IRIS ならば xterm で利用する。 Macintosh 等の PC も対応しているものが多いが、この場合は端末形式として vt100 を指定するとよい。この設定は IRx を立ち上げた後で行うことができる。

起動方法： hgc にログインした状態で、

```
irx
```

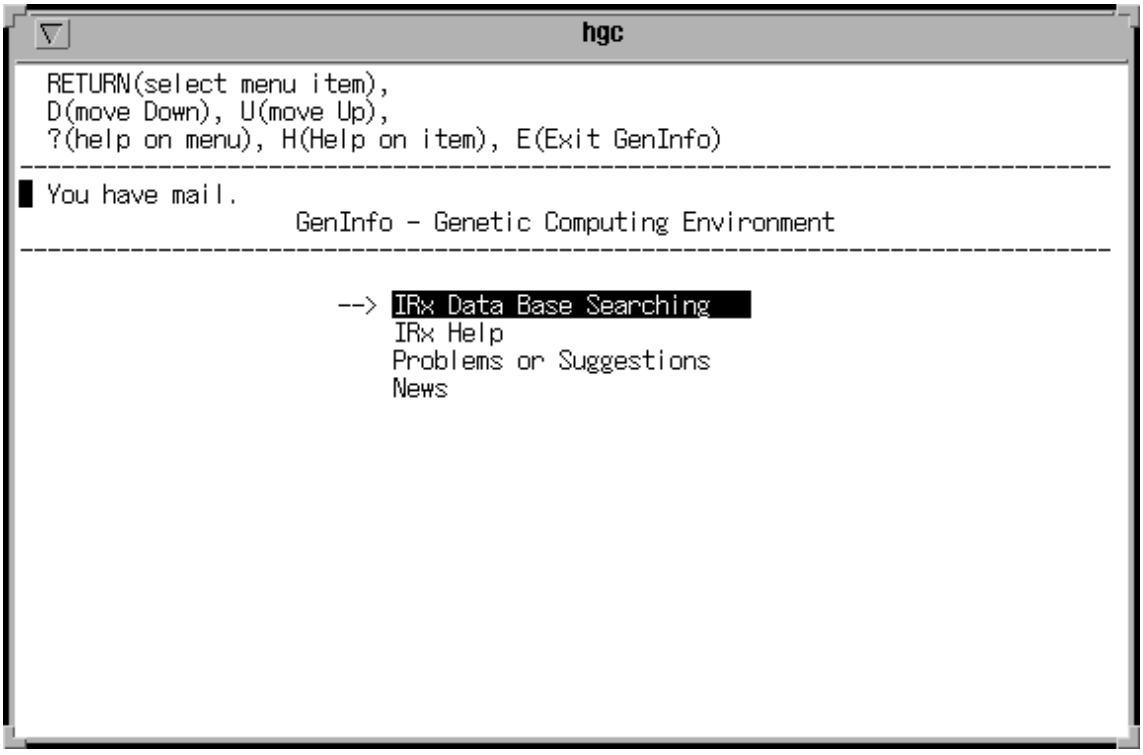
とだけタイプすると、次のようなオープニングメッセージが現われる。

```
                    *** Welcome to IRx Server ***

           IRx : Information Retrieval Experimental
                              Workbench

                            Version 3.0.37
                             6 March 1990

                   Lister Hill National Center for
                      Biomedical Communications

                    Served by Human Genome Center,
           Institute of Medical Science, The University of Tokyo

94/10/19        PMD updated. (Release 94-10, Oct 94)
94/10/28        SWISS-PROT updated. (Release 30.0, Oct 94)
94/11/07        PIR updated. (Release 42.0, Sep 94)
94/11/17        GenBank updated. (Release 85.0, Oct 94)
94/12/06        EMBL updated. (Release 40.0, Sep 94)
94/12/09        PRF updated. (Release 94-11, Nov 94)
94/12/26        LITDB updated. (Release 20-22, Dec 94)
95/01/05        LIGAND updated. (Release 11.0, 94/12/28)
95/01/09        PROSITE available. (Release 12.1, Oct 94)
95/01/22        OMIM updated. (Release MIM10+)
95/01/23        PDB updated. (Release , )


Current terminal type is: kterm

Type the name of terminal if you want to change >
```

この最後のところで確認を求めているように、現在設定されている端末タイプが表示されるので必要ならここで変更する。ただし大抵の場合は正しい端末タイプになっているので、その場合は単にリターンだけを打てばよい。

2 GenInfo メインメニューの表示

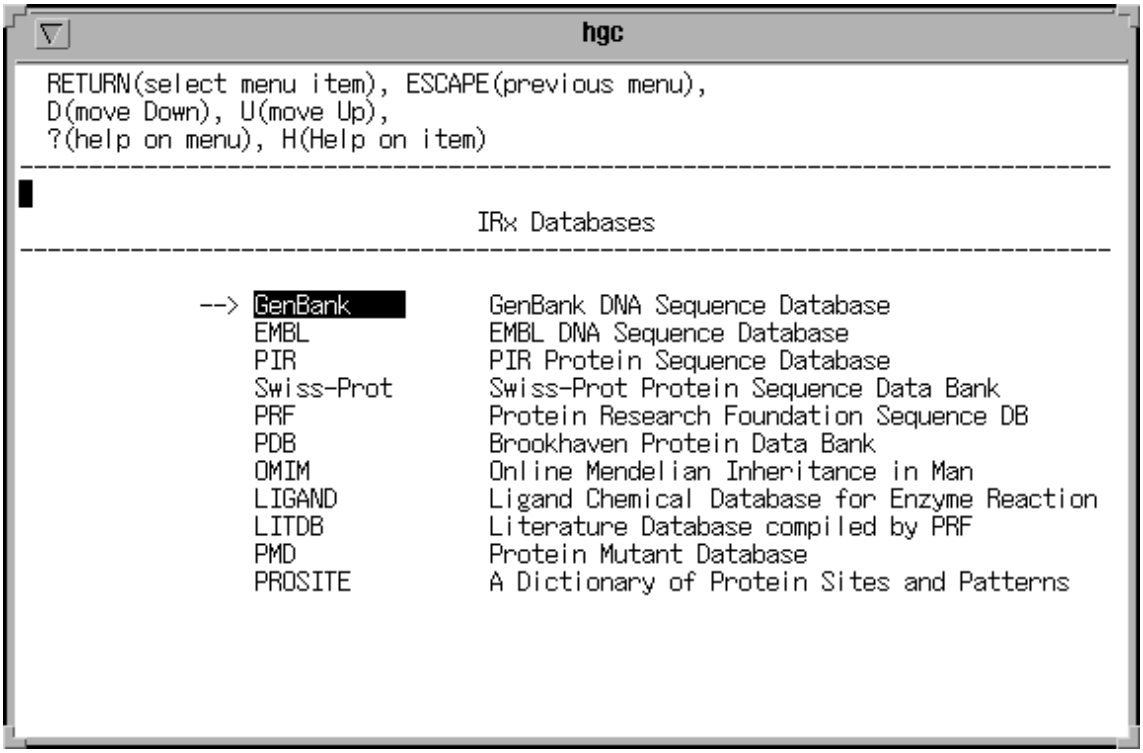
端末タイプがしかるべく設定されると画面が一旦消えて次のようなメニューが現われる。



ここから先は図のように画面上部に操作方法の簡単な説明が出ているので、これに従っていけばよいようになっている。この画面を最初に開いた状態では「IRx Data Base Searching」が反転した状態になっているので、そのままリターンを押す。もし別の項目が反転しているときは、**U** (Up) または **D** (Down) の各キーを押すことで、カーソルと反転項目が上下するので、目的のものが選ばれたらリターンを押す。

3 IRx データベース選択画面

IRx を選ぶとつぎのようなデータベース選択画面が出る。

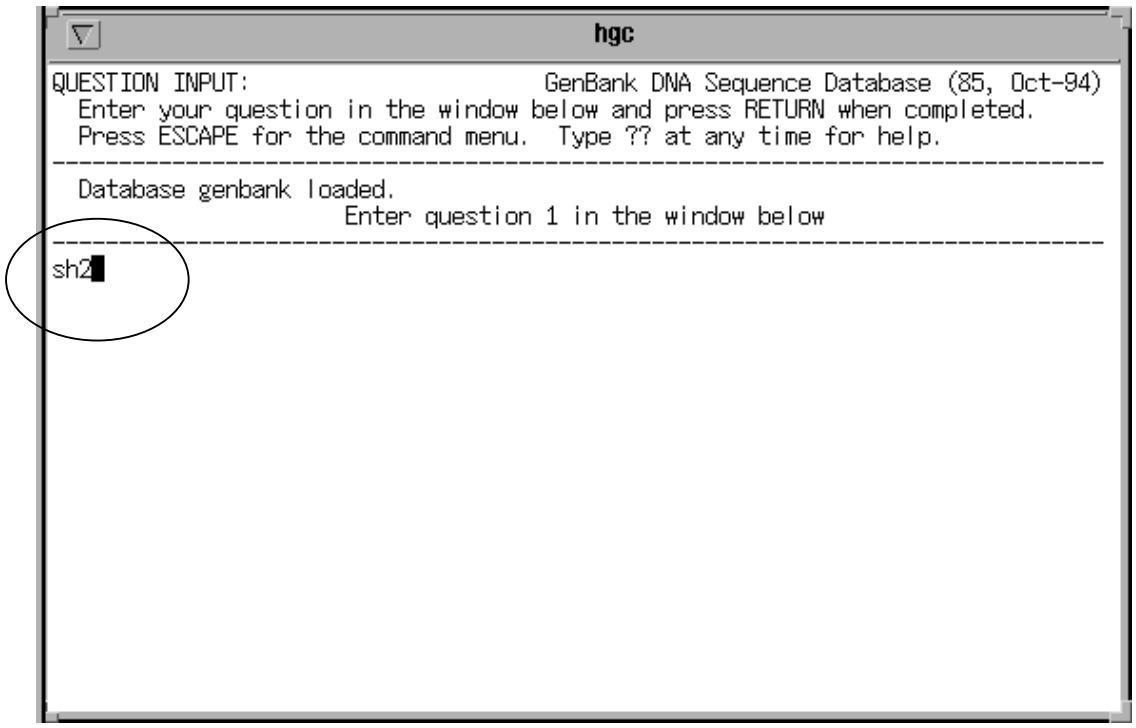


ここでのキー操作も画面上部にあるように、**D**, **U** でカーソルを上下させて目的のものを選んだらリターンを押す。まず例として GenBank を選ぶことにする。

4 問い合わせキーワードの入力：単純なキーワードの例

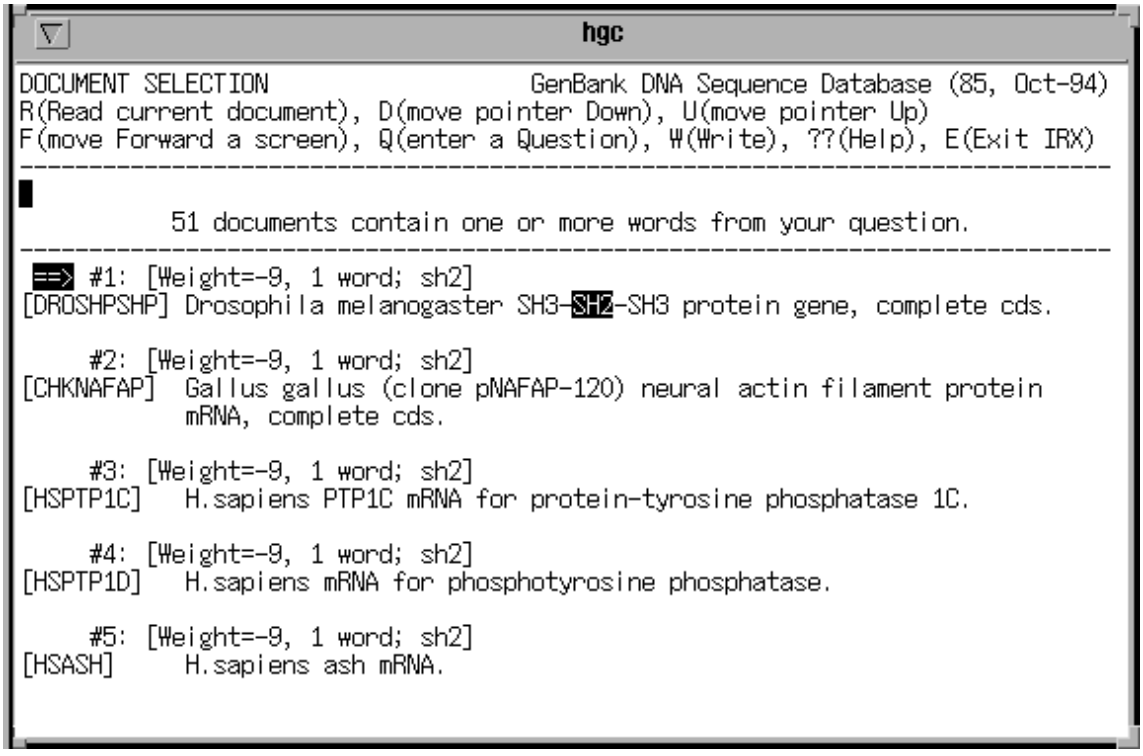
データベースを選ぶとまた画面が変わって問い合わせの字句を入力する画面になる。ここではまず例として SH2 (src homology 2) を検索してみよう。

（注意） IRx では 検索キーワード はたとえ大文字で書くものであっても
常に小文字で指定すること。



検索キーワードの入力は、上のような質問入力画面にそのまま言葉を小文字でタイプし、入力し終わったらリターンキーを押す。

すると、画面はつぎのようなリスト画面に変わる。

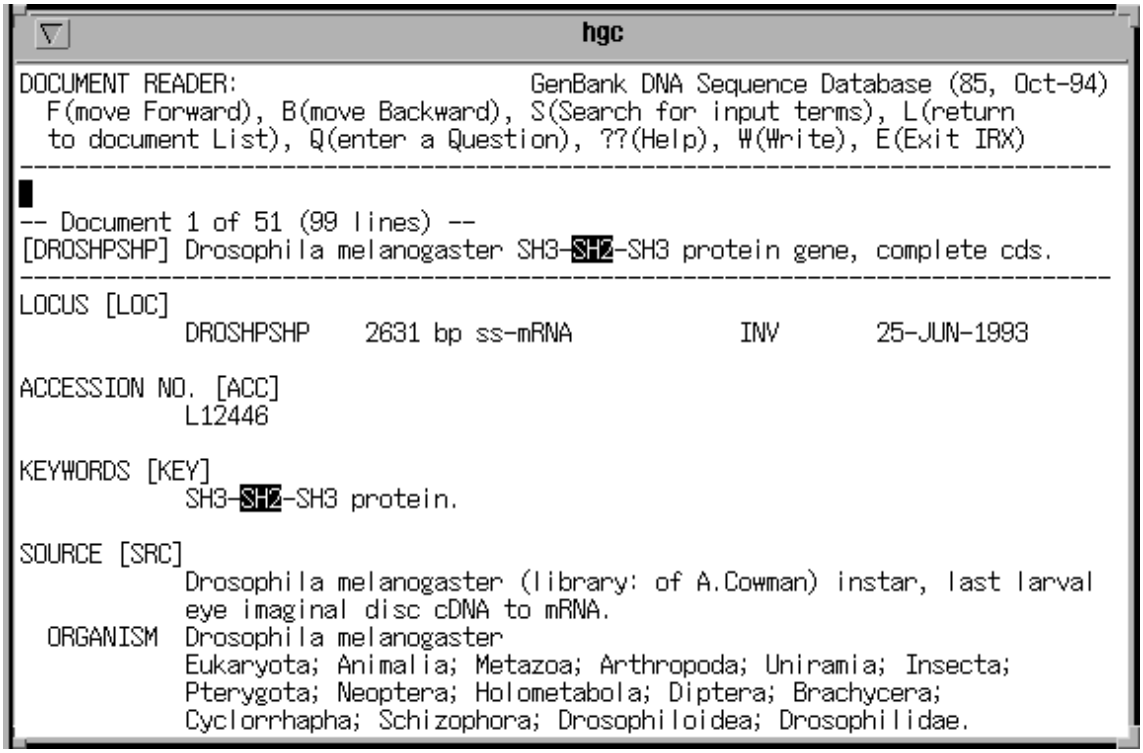


このリスト画面では、キーワードにマッチしたエントリーがキーワードによりよくマッチした順に並べて表示されている。

ここでも前画面と同様に **U**, **D** キーでカーソルを上下できる。また、 **F** (Forward), **B** (Backword) キーによって、 1 頁分前後に進める (戻す) こともできる。

目的のエントリーのところにカーソルが来たら、 Read の意味で **R** キーを押す。ここではリターンでは選択できないことに注意。

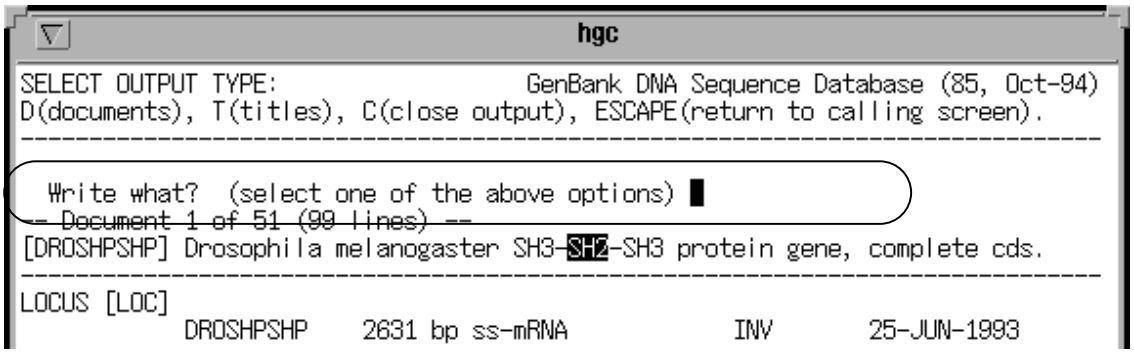
図の例では１番目の DROSHPSHP を選んでいる。



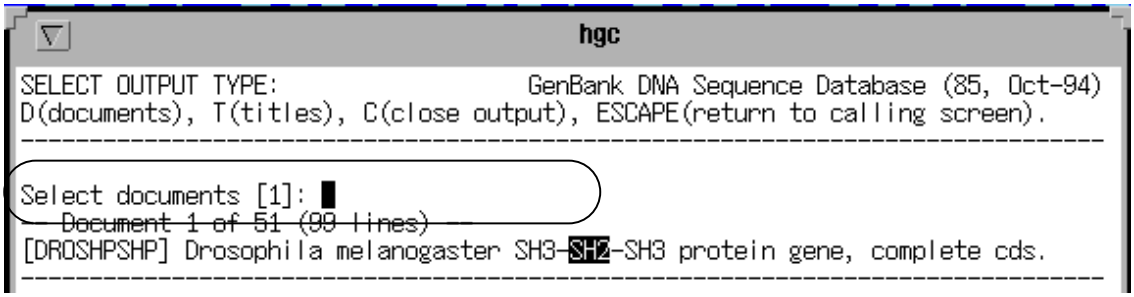
上の図はエントリーの内容を読む画面である。表示形式が GenBank オリジナルとは若干異なることに注意。ただ、内容的には同一である。ここで角括弧 [ ] に挟まれて書かれているフィールドアイテムは、後ほど述べる「フィールドを制限した検索」で使用される。

5　問い合わせ結果のファイルへの書き出し

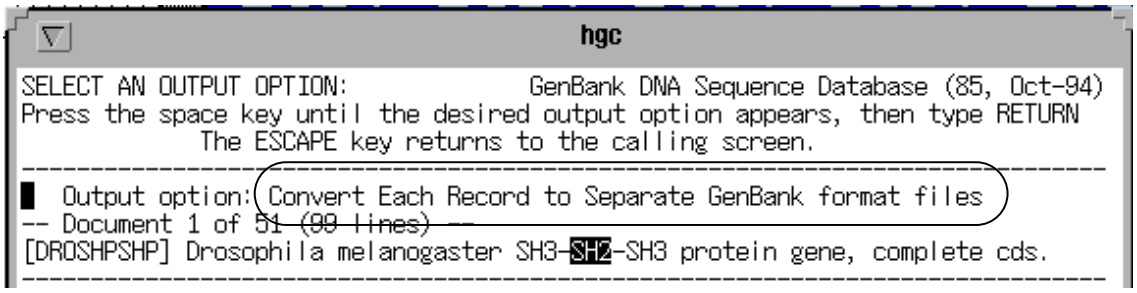
つぎに、ここで選んだ内容をファイルに保存する方法を述べる。 **W** (Write) コマンドによって、検索した結果を整形してファイルに納めることができる。つまり W キーを押すと画面がつぎのように変わり、何を書き出すかを尋ねる。



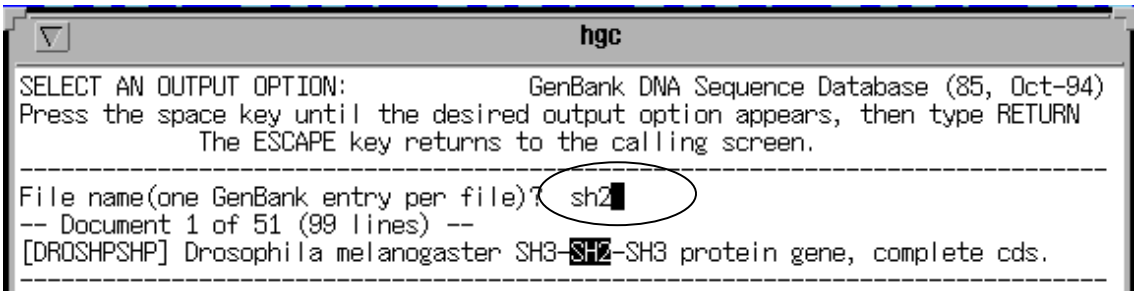
ここで聞いていることはエントリー（ここではドキュメントと呼んでいる）を書き出すのか、それとも１つ前の画面で見たようなタイトルリストを書き出すのかということである。通常はエントリーの内容を保存したいのであるから、 **D** (Documents) キーを押す。



つぎに聞いていることは、どのドキュメントを書くのかであるが、この例のようにドキュメント表示画面から Write コマンドを発行した状態では、今見ているのがどれかが (この場合 1 番目) 分かっているので、ここはリターンのみでよい。



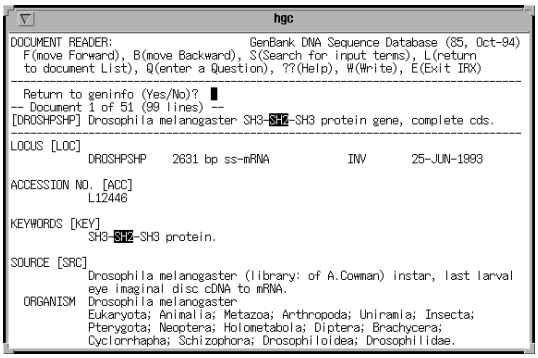
つぎに出力の形式をどうするかを聞いてくる。スペースキーを押すと出力形式が次々と変わるので、望みのものが出たところでリターンを押す。ここでは最初に出てくるオリジナルの GenBank 形式を選ぶこととする。

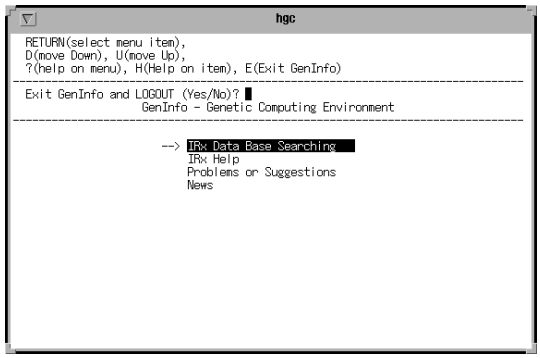


最後に聞いてくるのがファイル名、つまり何という名前で書き出すかである。希望する名前をタイプすればよい。ここでは sh2 とした。これで今いるディレクトリに sh2 と言う名前の GenBank オリジナル形式で書かれたファイルが生成された。

6 終了の方法

では取敢ずここで終わる方法を述べる。終了のためのコマンドは **E** (Exit) で、この状態で **E** と押すと、終了してよいかどうかの確認を Yes / No で求めてくるので、 **Y** を押す。すると 3 の IRx データベース選択画面に戻る。ここでは画面上部に指示されているように **ESCAPE** キーを押す。キーボードによっては **ESC** とだけ書かれているものもある。 ESCAPE キーを押すと 2 の GenInfo メインメニューに戻る。ここで **E** を押すと再度終了確認を求めるので、 **Y** を押すと IRx は全て終了し、起動前の状態に戻る。





なお、終了の方法としては上記のように Exit コマンドで一気に初期メニューに戻る他に、**ESCAPE** キーを押して行くことで 1 ステップずつ前の画面に戻ることもできる。この場合も最後に GenInfo メニューを終了させるときは **E** を押さないといけない。

(参考 1) ヘルプ

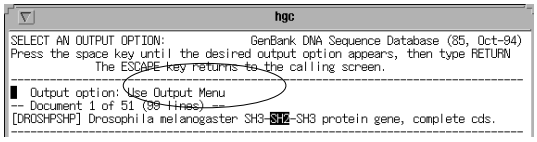
このように IRx は操作が易しく、しかもメニュー画面上部にコマンドキーの簡単な説明が表示されているので初心者にも使いやすいが、さらに詳細な使い方を知りたいときや、つぎにどうするのか分からないときなどは、画面上部に出ているように ? (場合によっては ??) を押すと、その状況に応じたヘルプ画面が表示される (図は省略)。

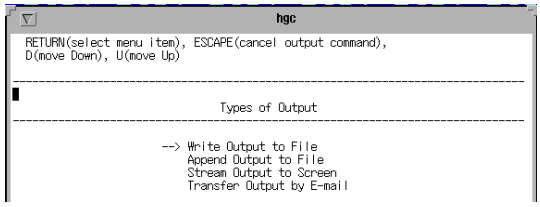
(参考 2) ファイル保存方法の補足：別形式 (**FASTA** 形式) で保存する

ファイルに保存するときにはオリジナルフォーマットばかりではなく、配列だけを FASTA 形式にして保存することも可能である。

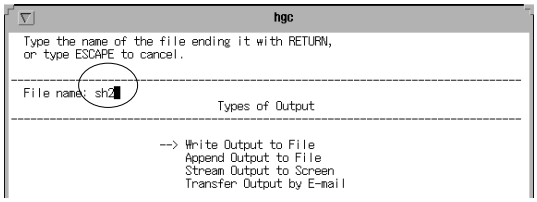
先ほどのファイル保存の例で FASTA 形式を選ぶ方法を示す。

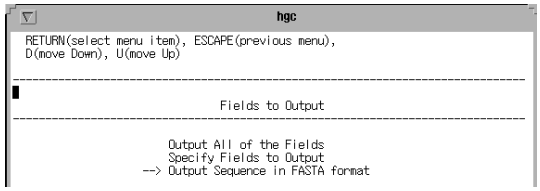
前項「ファイルへの書き出し」の 3 番目のところで、スペースキーを押して行き、図のように `Use Output Menu` が出たところでリターンキーを押す。画面が変わったところで Write Output to File を選び、





ついで問い合わせに応じてファイル名を入力してリターンキーをを打つと、最後にフィールド選択メニューが出るのでそこで `Output Sequence In FASTA format` を選ぶとよい。





7　HGC 外部から無記名利用資格で IRx を利用する

IRx は HGC に利用資格をもっている人ばかりではなく、インターネットに接続されたマシンからならば HGC 利用資格を申請していなくても誰でも利用できる。

利用法：
ネットワークを通して遠隔ログインするコマンド telnet 等を利用して `irx.genome.ad.jp` にアクセスし、ユーザー名として `irx` と指定するとパスワードフリーで irx だけが利用できる。

```
hgc{ogi}85: telnet␣irx.genome.ad.jp
Trying 133.11.192.1 ...
Connected to hgc.tokyo-center.genome.ad.jp.
Escape character is '^]'.


SunOS UNIX (hgc)

login: irx
Last login: Sat May 29 19:36:33 from imsvx1.ims.u-tok
SunOS Release 4.1.2-JLE1.1.2 (GENERIC_BAC_ULTRA) #1: Thu Feb 25 22:39:33 JST 1993

You have mail.



*** Welcome to IRx Server ***
```

ここから先は 1 の起動方法と全く同じ。

ただし、ここで Write コマンドを用いてファイル保存を行った場合、後処理操作がつけ加わる。即ち、HGC 内部で利用している場合は、IRx を起動したディレクトリに保存ファイルが作られるが、無記名利用資格の場合は自分のディレクトリというものがないため、保存ファイルは一時的に作られるが、終了したときにそれをネットワーク経由で利用者の手元に送るかどうかを問い合わせてくる。

この時、電子メールによる転送と、FTP（ファイル転送プロトコール）による転送が選択できる。いずれもインターネット接続されているところでは利用可能なので、好みの方を選択するとよい。

まず、メールによる転送を示す。

```
There are some files that may be created by 'Write' command.
Total amount of files is 6 KB

You can send them either by E-mail or by FTP
Do you want to send them ? (y/n)  y
Okay...
Select Mail or FTP  (m/f)  m

Enter your Mail Address : user@ims.u-tokyo.ac.jp

Send a mail to ogi@ims.u-tokyo.ac.jp containing 1 files of
IRX results   -- OK ?  (y/n) y
Mail was sent to user@ims.u-tokyo.ac.jp
Bye
Connection closed by foreign host.
hgc{user}84:
```

最初に送り返すかどうかには y と答える。つぎにメールか FTP かを聞いてくるので、メールの意味で m とうつ。そうすると送り先のメールアドレスを聞いてくるので、自分のよく使っている電子メールアドレスを入力する。最後に確認を求めるので、y と答えればよい。

つぎに、FTP による転送方法を示す。

```
There are some files that may be created by 'Write' command.
Total amount of files is 3 KB

You can send them either by E-mail or by FTP
Do you want to send them ? (y/n)  y
Okay...
Select Mail or FTP  (m/f)  f
Enter your host machine name > my_host.ims.u-tokyo.ac.jp
Enter your login name > myname
Enter password for myname @ my_host.ims.u-tokyo.ac.jp >
Please re-enter the password >

What do you wish to name the IRX output file in your remote macine?
N.B. All your results will be archived by tar and compressd
before sending, so .tar.Z will be automatically added.
Name of result file [irx-result.930529.2026.tar.Z] >

FTP host name   : my_host.ims.u-tokyo.ac.jp
login name      : myname
remote file name: irx-result.930529.2026.tar.Z

OK ? (y/n) y
[1] 6201

Your results will be send within a few minutes
Bye
Connection closed by foreign host.
hgc{ogi}86:
```

y によって送り返しを指示した後、Mail or FTP では F を選ぶ。つぎにどこへ送るかの送り先マシン名を聞いてくるので、図の例のように Full Qualified Domain Name (何とか . 組織機関名 . {ac | or | go | co }. jp という形式の名前) 又は直接 IP-address を (例えば 133.11.196.1 の様に) 指定する。つぎにそのマシン名での利用者名を聞いてくるのでそこでの自分のログイン名を指定する。するとその名前でのパスワードを聞いてくるが、これは端末には打った内容が表示されない。パスワードはもう 1 度確認のために同じことを打つように求められる。つぎに、送るファイルの名前をどうしたいか聞いてくるが、とくに差しつかえない場合リターンだけ打てばシステムがかってにつけた名前 (irx-result....tar.Z という形式) でファイルが送られる。

このファイルはネットワーク転送の都合上、tar という形式に 1 まとめにした後、圧縮されているので、それをもとに戻すには、図の例ならば、

```
uncompress irx-result.930529.2026.tar.Z
tar xvf irx-result.930529.2026.tar
```

とすると、今いるディレクトリに保存したファイルがもとの名前で展開される。

8　応用-１ キーワードの入力：複数キーワード、同一フィールド内限定

今の例は単一の単語によるキーワード指定であったが、複数のキーワードを指定することができる。それは、単に複数の単語をスペースで区切って並べるだけでよい。この場合 (bfind とは異なり) 、各キーワードのうちどれかをもつもの、すなわち OR 結合がとられる。

次の例では、ヒト遺伝病データベース OMIM より cystic fibrosis に関するものを検索しようとしている。まずデータベース選択画面まで行き、そこで OMIM を選んで質問入力画面を出す。

```
QUESTION INPUT:            Online Mendelian Inheritance in Man (MIM10+, Jan-95)
  Enter your question in the window below and press RETURN when completed.
  Press ESCAPE for the command menu.  Type ?? at any time for help.
--------------------------------------------------------------------------------
  Database omim loaded.
                      Enter question 1 in the window below
--------------------------------------------------------------------------------
cystic fibrosis
```

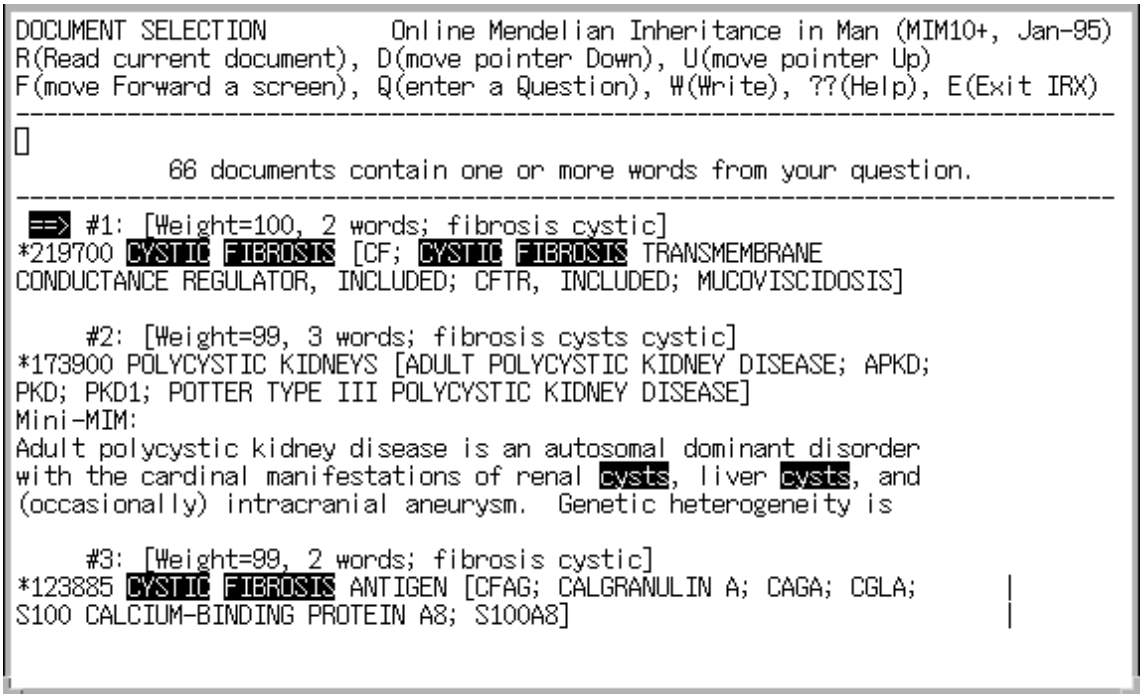
先ずはそこに直接 cystic fibrosis と打ってみると、次の図のように確かに cystic fibrosis のエントリーが先頭にあるのだが、リストの下の方には関係ないものも拾われている (表示は略す)。それは、 OR 検索であるためにどちらか一方の言葉があればいいのと、さらに IRx では語形変化も考慮にいれたバリエーションも含めて検索してしまうためである。(リストの順番のところで、キーワードの似ている順と言った意味は、厳密一致ではなくこのように変化しているものも拾っていることに由来する。) 次と比較のため、検出された件数に注意。

```
DOCUMENT SELECTION         Online Mendelian Inheritance in Man (MIM10+, Jan-95)
R(Read current document), D(move pointer Down), U(move pointer Up)
F(move Forward a screen), Q(enter a Question), W(Write), ??(Help), E(Exit IRX)
--------------------------------------------------------------------------------

          290 documents contain one or more words from your question.
--------------------------------------------------------------------------------
 ==> #1: [Weight=100, 2 words; fibrosis cystic]
*219700 CYSTIC FIBROSIS [CF; CYSTIC FIBROSIS TRANSMEMBRANE
CONDUCTANCE REGULATOR, INCLUDED; CFTR, INCLUDED; MUCOVISCIDOSIS]

     #2: [Weight=99, 3 words; fibrosis cysts cystic]
*173900 POLYCYSTIC KIDNEYS [ADULT POLYCYSTIC KIDNEY DISEASE; APKD;
PKD; PKD1; POTTER TYPE III POLYCYSTIC KIDNEY DISEASE]
Mini-MIM:
Adult polycystic kidney disease is an autosomal dominant disorder
with the cardinal manifestations of renal cysts, liver cysts, and
(occasionally) intracranial aneurysm.  Genetic heterogeneity is

     #3: [Weight=99, 2 words; fibrosis cystic]
*123885 CYSTIC FIBROSIS ANTIGEN [CFAG; CALGRANULIN A; CAGA; CGLA;
S100 CALCIUM-BINDING PROTEIN A8; S100A8]
```

それでは cystic fibrosis という１つの言葉として検索するにはどうすればよいかを次に示す。

```
QUESTION INPUT:            Online Mendelian Inheritance in Man (MIM10+, Jan-95)
  Enter your question in the window below and press RETURN when completed.
  Press ESCAPE for the command menu.  Type ?? at any time for help.
--------------------------------------------------------------------------------

                      Enter question 2 in the window below
--------------------------------------------------------------------------------
"cystic fibrosis"
```

図の例で示すように、その言葉の前後をダブルクオート（"）で囲って "cystic fibrosis" のように書いてやればよい。

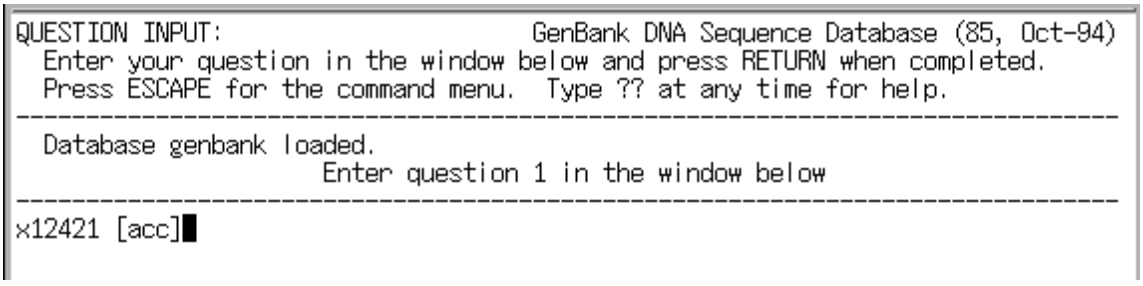


先ほどの例と比べると、だいぶ件数が絞り込まれたことが分かる。

（参考）IRx では厳密にはダブルクオートでの囲い込みは、その中を１まとまりの言葉とするという意味ではない。このように指定されると、それらは共に同一フィールドにあることが条件となる。場合によっては単語の順番で１つながりとなって現われないものも拾うことがありうるが、ただリスト画面ではキーワードのスコアの高い順に並んでいるので、リストの上位から見ていけば目的のものがあるかどうかは容易に分かるだろう。従って、近似的にダブルクオートでの囲い込みは１まとまりの言葉として検索すると見做してもよい。

9 応用-2 キーワードの入力： 検索フィールドの限定

特定のフィールドだけを検索対象としたい場合は、キーワードの後ろにそのフィールド名を角括弧 [ ] で囲って指定すればよい。



この例では Accession 番号 X12421 のものを探したのだが、このように、目的のエントリーそのものが見つかっている。

```
DOCUMENT SELECTION                    GenBank DNA Sequence Database (85, Oct-94)
R(Read current document), D(move pointer Down), U(move pointer Up)
F(move Forward a screen), Q(enter a Question), W(Write), ??(Help), E(Exit IRX)
--------------------------------------------------------------------------------
[]
            1 document contain one or more words from your question.
--------------------------------------------------------------------------------
 ==> #1: [Weight=-9, 1 word; x12421]
[HSEIF2AP]  Human eIF-2alpha gene for translation factor eIF-2 alpha subunit,
            promoter region.
```

10　応用-3 キーワードの入力： AND 条件の指定

IRx では特にことわりがないと、複数指定されたキーワードは OR 結合として (つまり、どれか 1 つでもあればよい) 検索された。逆に、指定したキーワード全てを備えること、即ち AND 結合を指定するには、それらのキーワードの間に大文字で AND と書いてやる。

次の例では yeast の第 3 染色体に関するものを検索しているが、 chromosome III はさらに応用 1 の例で示したように 1 まとまりの言葉と見なすためにダブルクオートで囲い、これと yeast と言う言葉を AND 条件で検索している。

```
QUESTION INPUT:                       GenBank DNA Sequence Database (85, Oct-94)
  Enter your question in the window below and press RETURN when completed.
  Press ESCAPE for the command menu.  Type ?? at any time for help.
--------------------------------------------------------------------------------
  Database genbank loaded.
                        Enter question 1 in the window below
--------------------------------------------------------------------------------
yeast AND "chromosome iii"[]
```

```
DOCUMENT SELECTION                    GenBank DNA Sequence Database (85, Oct-94)
R(Read current document), D(move pointer Down), U(move pointer Up)
F(move Forward a screen), Q(enter a Question), W(Write), ??(Help), E(Exit IRX)
--------------------------------------------------------------------------------
[]
          103 documents contain one or more words from your question.
--------------------------------------------------------------------------------
 ==> #1: [Weight=3, 3 words; iii yeast chromosome]
[SCCHRIII]  S.cerevisiae chromosome III complete DNA sequence.

     #2: [Weight=3, 3 words; iii yeast chromosome]
[YSCCHR1RAA]Saccharomyces cerevisiae chromosome I right arm.

     #3: [Weight=3, 3 words; iii yeast chromosome]
[SCTY117X]  Yeast transposon Ty1-17 and flanking tRNA-Glu 3/Leu 3,LEU2 genes
            from small ring of chromosome III (LEU2 coding for
            beta-isopropylmalate dehydrogenase).

     #4: [Weight=3, 3 words; iii yeast chromosome]
[SCRAHS1]   Yeast CN31C chromosome III RAHS DNA (right arm transcription
            hot-spot).
```

(参考) AND 演算子の他に、 NOT, OR といった論理演算子を使うことができる。ただし、 OR は通常キーワードの区切れにこれが仮定されているので特にこれを指定する必要はない。

（参考）AND 結合とダブルクオート囲い込みの違いは、いずれもキーワードとして与えられた単語全てが出現しなくてはいけないが、 AND 結合では必ずしも同一フィールドでなくてもよいのに対し、クオート囲い込みでは必ず同一フィールドにないといけない。

11 応用-4 キーワードの入力： stemming の抑止

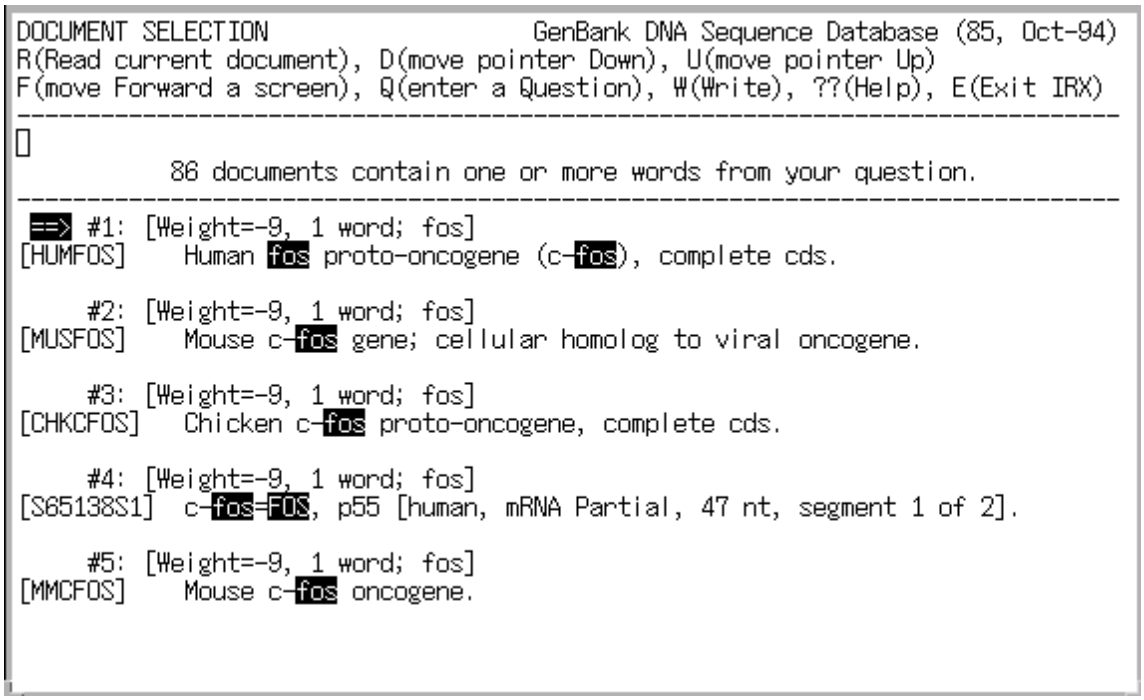
IRx ではより柔軟なキーワード検索を行うために、通常では stemming という、単語の語幹部と語尾部を推定し語尾変化も考慮した検索をしている。ただしこれが逆に余計なものを拾ってしまう可能性もあり、 stemming を抑止したいこともある。そのようなときはキーワードに続けてアットマーク（＠）を付ける。

例：癌遺伝子 fos を検索したい。そこでただ fos とだけ指定すると、語幹 fo と語尾 s のような可能性も同時に検索されてしまうので、次のように極めて多量に見つかってしまい、 IRx の方で問い合わせを修正して再度試行するかどうか問い合わせてくる。

```
QUESTION INPUT:                        GenBank DNA Sequence Database (85, Oct-94)
  Enter your question in the window below and press RETURN when completed.
  Press ESCAPE for the command menu.  Type ?? at any time for help.
--------------------------------------------------------------------------------
  2605 documents found.  Do you want to modify your question? 
                       Enter question 1 in the window below
--------------------------------------------------------------------------------
fos
```

こういうときには次のように stemming 抑止指定の@を付けると、大幅に絞り込むことができる。

```
QUESTION INPUT:                        GenBank DNA Sequence Database (85, Oct-94)
  Enter your question in the window below and press RETURN when completed.
  Press ESCAPE for the command menu.  Type ?? at any time for help.
--------------------------------------------------------------------------------
  Processing ...
                       Enter question 1 in the window below
--------------------------------------------------------------------------------
fos@
```



（参考）このほかにも、特殊記号として

　#　任意の１文字
　$　任意の１文字だがなくてもよい といったものが使える。これらを指定すると
　*　任意の文字列（なくてもよい）
それにマッチするキーワード一覧が示され、その中から再度キーワードを選ぶことになる。ただし IRx では基本的に OR 検索がなされ、しかもマッチしたスコア順に表示されるので、余りこれらを使う必要はないと思われる。