

国立遺伝学研究所 DDBJ 利用者講習会 (6月17日) 資料
UWGC G 及び IDEAS 利用の手引

国立遺伝学研究所
分子遺伝研究部門
藤田 信之

FROMGENBANK	converts a file from GenBank format to UWCCG format
FROMSTADEN	converts a file from STADEN format to UWCCG format
TOSTADEN	converts a file from UWCCG format to STADEN format

Sequence Conversion

BACKTRANSLATE	translates a peptide into a nucleotide sequence
TRANSLATE	translates a nucleotide into a peptide sequence
REVERSE	reverses and/or complements a sequence
SHUFFLE	randomizes a sequence while maintaining its composition
SIMPLIFY	makes a simplification scheme for peptide sequence comparisons

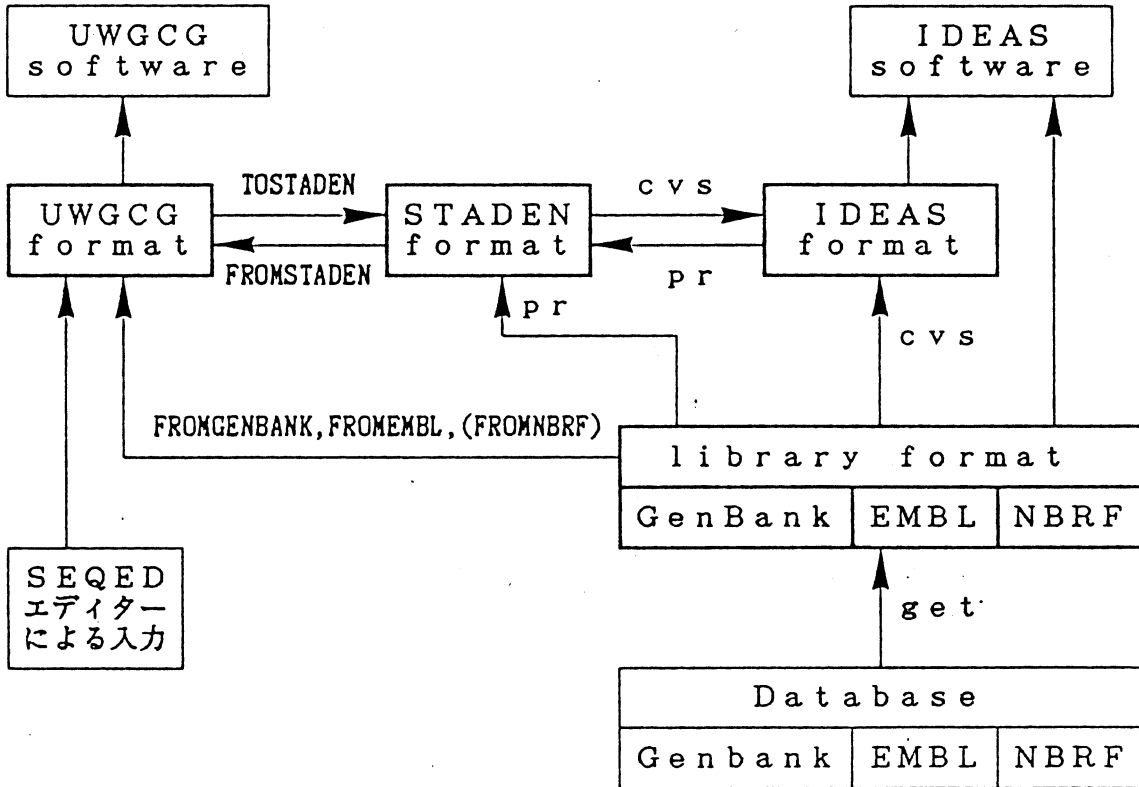
Printing and Publication

PUBLISH	arranges sequences for publication
POSTER	arranges text on figures for publication or for poster sessions
OVERPRINT	prints sequence figures using a daisy wheel printer

Database Search

STRINGS	finds files by character string (key word)
NAMES	finds files by names
SHOWFILES	writes a documented "file of file names"
FETCH	copies UWCCG data file(s) into your directory

個々のデータベースおよび解析ソフトウェアがそれぞれ異なったデータフォーマットを採用しているため、フォーマット変換（データの加工）をする必要がかなり頻繁に生じます。下に交換のおおまかな流れ図を示しますが、IDEASやUWGCGのプログラムを使いこなすには、この過程を十分に理解しておくことが重要です。



上の図の中で、“get”, “cvs”, “pr”はIDEAS/SEQMANの中のコマンドを表わし、“TOSTADEN”, “FROMSTADEN”, “FROMGENBANK”, “FROMEMBL”, “FROMNBRF”はUWGCGの中のプログラムを表わす。

getコマンド (IDEAS/SEQMAN) の使い方

```

genbank
%get  embl      sequenceID >outputfile
      nbrf
      genbank
%get  embl      <listfile >outputfile
      nbrf
    
```

cvsコマンド (IDEAS/SEQMAN) の使い方

```
%cvs inputfile >outputfile
```

prコマンド (IDEAS/SEQMAN) の使い方

```
%pr inputfile >outputfile
```

1. データベースを検索するためNAQを呼び出す。検索は GenBank DNAデータベースを対象とする。
2. Definition行に"E.", "coli", "heat", "shock"を含むエントリーを検索する。
3. NAQから抜け出る。
4. 必要なデータを取り出すためIDEAS中のSEQMANを呼び出す。単に"IDEAS"とのみ入力するとメニューが表示されるので、その中からSEQMANを指定することも可能。
5. GETコマンドにより、GenBankデータベース中のエントリー"ECOHTPR"を取り出し、自分のディレクトリに格納する。ファイル名は"ECOHTPR."とする。
6. 確かにコピーできたかどうかをTYPEコマンドで確認する。
7. SEQMANから抜け出る。同時にIDEASからも抜け出る。
8. UWGCGの使用を宣言する。(最初に一度のみ必要)
9. GenBankフォーマットからUWGCGフォーマットに変換するためFROMGENBANKを呼び出す。
10. フォーマット変換するファイルの名前を入力する。
11. 確かにフォーマット変換できたかどうかをTYPEコマンドで確認する。

12. 制限酵素の切断点を求めるためMAPを呼び出す。

13. ファイル名を入力する。

14. データの始まりを指定する。

15. データの終わりを指定する。

16. 制限酵素名を入力する。"*"を指定すると登録されている酵素すべてが対象となる。

17. アミノ酸配列への翻訳を3つのフレームについて行なう。ただしオープンフレームのみを表示する。

18. 結果を端末(TERM)に出力する。ファイル名を指定してファイルに書き出すことも可能。

1. UWGCGの使用を宣言する（最初に一度だけ必要）。
2. 配列編集用の画面エディター"SEQED"を呼び出す。
（SEQEDを用いるためには、VTエミュレーターが必要）
3. 作成するファイルの名前を入力する。
4. コメントを入力する。コメントは5行まで入力できるが、他のフォーマットに変換する際失われることに注意。
5. 配列を大文字で入力する。画面の範囲を越えると自動的にスクロールされる。

主な編集コマンド

→, ←	一文字右へ, 一文字左へ
n→, n←	n文字右へ, n文字左へ
>, <	50文字右へ, 50文字左へ
:n [CR]	n番目の文字へ
/文字列 [CR]	文字列の検索
[DEL]	一文字削除
:n1, n2 delete [CR]	n1番目からn2番目まで削除

6. コロン ":"に続けて "check [CR]"と入力し、再入力モードに入る。
7. 入力の誤りをさがすため、配列を再度入力する。一致しない箇所には "-"が表示されブザーが鳴るので修正する。行の間の移動は "↓"および "↑"で行なう。
8. 確認が終わったらコロン ":"に続けて "exit [CR]"と入力し、エディターから抜け出す。
9. ファイルが出来ていることをTYPEコマンドで確認する。

10. UWGCGフォーマットからSTADENフォーマット（中間フォーマット）に変換するためTOSTADENを呼び出す。
11. 変換するファイルの名前を入力する。
12. データの始まりを指定する。
13. データの終わりを指定する。
14. 変換後のデータを格納するファイルの名前を入力する。
15. フォーマット変換できていることをTYPEコマンドで確認する。
16. IDEASを呼び出す。
17. IDEASの中のSEQMANを呼び出す。
18. CVSコマンドにより中間フォーマットの"TEST. SEQ"をIDEASフォーマットに変換する。変換後のデータは"TEST. SEQ"の名前でファイルに書き出す。
19. フォーマット変換できていることをTYPEコマンドで確認する。
20. SEQMANから抜け出る。
21. ホモロジー検索を行なうためIDEASの中のSEQFPを呼び出す。

22. 検索するファイルの名前を入力する。
 23. 検索の対象となる配列集合を指定する。
 24. 検索結果の出力先を指定する。
 25. 検索のパラメーターを変更する。
 26. 最初のパラメーター (MAXD) を-70に変更する。2番目以降のパラメーターは初期設定のままとする。
 27. データのIDを入力する。
 28. データの始まりと終わりを指定する。
 29. 以上の設定をバッチジョブとして実行させる。
-
30. IDEASから抜け出る。
この段階で一旦LOGOUTするか、別の作業を行なう。
 31. バッチジョブが終わったら結果を出力させる。