Sequence Analysis Spring 2000 Dr. Richard Friedman
(212)305-6901 (76901) friedman@cuccfa.ccc.columbia.edu 130BB

Lesson 4 Introducing GCG

In this lesson, you will initialize the GCG program and other programs used in the course, put programs in GCG format, and display and obtain sequences from the database.

Summary of commands:

Note: In this document different fonts have different meanings:

Times is used to explain commands.

`Courier is used to indicate commands and command options.`

`Courier italics are used to indicate command parameters, for example, filenames.`

`Courier bold is used to indicate commands that are not displayed.`

`Courier bold italics are used to indicate computer-generated output.`

Helvetica is used to indicate menu items.

Lab:

I. Initializing sequence analysis programs on cuccfa (If you are unfamiliar with Unix editors, let me know and I will do this for you).

1. Log onto cuccfa.
2. Make the following changes in your .login file:
3. Add the following entries to your path (i.e. between the parentheses where it says “set path=()”. Continue lines with a "\" .).

```
/usr2/seq/bin /usr/bin/X11 /usr/dt/bin
```

4. Enter the following command in your .login file.

```
source /usr2/seq/bin/.seqinit
```

5. Save your .login file and exit the editor.

6. Type

   ```
   class-setup
   ```

   (This program copies */usr2/seq/Xstuff/.Xdefaults* to your root directory, thereby configuring your Xwindows environment, and links your home directory to /usr2/seq/bin/.ncbirc, which enables you to run the ncbi version of Blast locally. I hope that this is clear).

7. Type

   ```
   source .login
   ```

   (You only have to make these changes once. Everything will initialize on its own whenever you logon after this.)

8. Make a directory called lab4 and go to it with the following commands:

   ```
   > mkdir lab4
   > cd lab4
   ```

II. Generating a sequence file in GCG format

1. Open pico or vi on cuccfa or Microsoft Word on your Mac or PC.
2. Make up a nucleic acid or protein sequence of at least 60 residues and type it in. Make up some documentation too, which you place at the front of the file. Place two periods between the documentation and the sequence.
   - A. pico or vi Users:
     - 1. Save the file.
   - B. Word users:
     - 1. Save the file as text only with line breaks.
     - 2a. Send the file over to cuccfa using Netscape, or WS-FTP or Fetch.
     - or
     - 2b. Copy and paste into a pico or vi file on cuccfa.
3. Use the reformat command as described in Chapter 3 of Tutorial Introduction to the GCG software package.

III. Read and work through GCG tutorial Chapters 1-3, 5.
Note: On page 1-12, the sequence designation should read “pr:hscolii?” not “pr:hscollii?”
Practice the commands in the "Summary of commands" below. You don't have to repeat the ones that you have already learned.

IV. GCG documentation

Open Netscape and familiarize yourself with:

http://www.ccc.columbia.edu/genhelp/

V. Practice the commands below that are not covered in the tutorial.

Summary of commands:

`gcg`	Initializes GCG program.
`reformat`	Reformats file into GCG-format.
`tofasta`	Translates a sequence file from GCG format to Fasta format.
`fromfasta`	Translates a sequence file from Fasta format to GCG format.
`fasta`	Runs the GCG program Fasta.
`fasta –fasta`	Runs the GCG program Fasta with query sequence in Fasta format.
`ssearch3`	Runs the Ssearch3 program from the Fasta3 package.
`genhelp`	Starts the online GCG manual. Includes Program manual, User Guide, and up-to-date databases.
`http://www.ccc.columbia.edu/genhelp/`	Address for the web version of the online GCG manual. Includes Program manual, User Guide, and up-to-date databases.
`http://www.ccc.columbia.edu/genhelp/databases.html`	Address for GCG databases.
`comcheck`	Has the GCG program print out all of the options for each gcg command.
`nocomcheck`	Has the GCG program **NOT** print out all of the options for each gcg command.
typedata *humak1*	Displays the sequence file of “humak1”.
typedata *humak1* | more	Displays the sequence file of “humak1” a page at a time.
fetch *humak1*	Copies the sequence of “humak1” into your directory.
fetch *hscolii?*	Copies the sequences of “hscolii?” (where “?” stands for any character) into your directory.
fetch *hscol**	Copies the sequences of “hscol*” (where “*” stands for any group of characters or no character) into your directory.

`fetch m81385`	Copies the sequence whose accession number is “m81385” from our local database into your directory.
`fetch gb:m81385`	Copies the sequence whose accession number is “m81385” from our local genbank database into your directory. Specifying the database is useful if the same accession number is used in both Genbank and Genpept.
`netfetch m81385`	Copies the sequence whose accession number is “m81385” from the NCBI database into your directory. Saves as an RSF file.
`reformat m81385.msf{*}`	Extracts file in standard GCG format from RSF file.
`datalist`	Lists the databases available to the local versions of Gapped BLAST and PSIBLAST.
`fastacmd -d nr -s 4019221`	Displays sequence with accession number "*4019221*" from the local nonredundant protein database in Fasta format.
`fastacmd -d nr -s 4019221 > 4019221.tfa`	Copies sequence with accession number "*4019221*" from the local nonredundant protein database in Fasta format to file "4019221.tfa".
`fastacmd -`	Gives options for fastacmd.
`tk`	Sets graphics output to be displayed on the terminal.
`lw`	Sets graphics output to be written to a Postscript file which can be downloaded and printed on your Mac or PC printer. The Postscript file is called plot.ps .
`lz`	Sets graphics output to be printed on cuccfa’s printer.
`g`	Sets graphics output to be written to a gif file which can be downloaded to your Mac or PC and viewed and printed with Netscape. The GIF file is called GCG_GRAPHICS. Change its name to a *.gif file before downloading and download as a binary file.
`x`	Sets graphics to output to an X-windows graphics screen. Must be activated from an X terminal.
`gs name.ps`	Displays Postscript output in an X-window. Must be executed from an X-terminal.

pspdf *name* ***GS*> <ctrl D>**	Converts a postscript file, "name.ps" to a PDF file, "name.pdf". Do not include the ".ps" as part of the name. Type <ctrl D> in response to the "GS>" prompt. PDF files can either be downloaded to a Mac or a PC and displayed with Adobe Acrobat or displayed on cuccfa by typing:
acroread *name.pdf*	Displays a PDFoutput in an X-window on cuccfa. Must be typed from an X-terminal.
http://www.ghostscript.com/	Where to get programs to view Postscript files on your Mac or PC.
pepplot -figure	Sets graphics output to be printed into a Figure file which can be downloaded to a Mac, and displayed, manipulated, and printed with the Macintosh program GCG-Figure.
figure *filename.figure*	Displays data in whatever figure format that was previously indicated.
blastpgp -i hba_hobbit.tfa -d sw	Runs the local version of Gapped BLAST and PSIBLAST.
http://www.ncbi.nlm.nih.gov/	Accesses the NCBI web-page which contains the BLAST2 server.
Select Sequence Analysis Course from the Windows program menu.	Activates the common desktop environment. Requires an X-windows client program, such as eXodus for the Mac or Exceed for PC.
Select Terminal from the CDE Notepad menu.	Opens an X-terminal window.
seqlab	Starts the X-Windows Version of GCG, if typed in an X-terminal. Or click Seqlab icon.

For SeqLab users (optional):

Work through the SeqLab tutorial, Introduction and Lesson 1.

Web sites:

http://dot.imgen.bcm.tmc.edu:9331/seq-util/readseq.html
Translates sequence files between different formats.

http://www.ncbi.nlm.nih.gov/
Accesses the NCBI web-page which contains the BLAST server.

http://www.ccc.columbia.edu/~friedman/isaac.html
A list of useful and instructive sequence analysis web-sites organized according to the Lessons in this course.

http://www.ghostscript.com/
Where to download (free) software that displays and prints Postscript files on a Mac or PC.

http://www.adobe.com/products/acrobat/readstep.html
Where to download Acrobat Reader (free) which displays and prints PDF files on a Mac or PC.