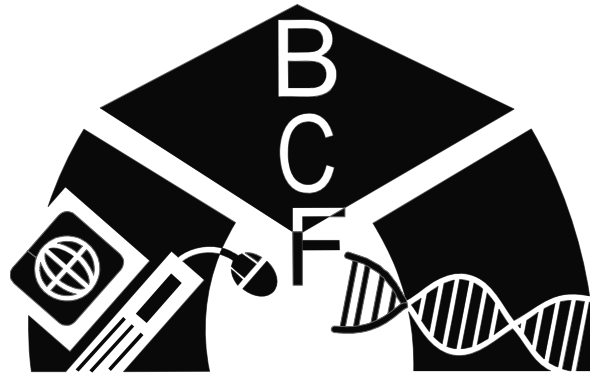


THE UNIVERSITY OF  
**ARIZONA**<sup>®</sup>  
TUCSON ARIZONA



**Biotechnology Computing Facility**

# **Introduction to GCG/SeqLab**

## **Part A**

**Preparations**  
**Introduction to SeqLab Interface**  
**Handling Input and Output**  
**Basic Operations**

Parts based on Documentation from Genetics Computer Group  
Customized for the UofA by the BCF

For assistance please contact:  
<http://bcf.arl.arizona.edu/>

Gavin Nelson – 621.3206 – [gmn@email.arizona.edu](mailto:gmn@email.arizona.edu)  
Susan Miller – 626.2597 – [sjmiller@email.arizona.edu](mailto:sjmiller@email.arizona.edu)  
Nirav Merchant – 621.8379 – [nirav@arl.arizona.edu](mailto:nirav@arl.arizona.edu)



# Contents

---

|  |           |
|--|-----------|
| <b>Contents</b> .....                    | <b>2</b>  |
| Overview .....                           | 3         |
| <b>What is GCG &amp; SeqLab?</b> .....   | <b>4</b>  |
| <b>Preparations</b> .....                | <b>5</b>  |
| What you need .....                      | 5         |
| <b>Handling Input and Output</b> .....   | <b>6</b>  |
| Importing Sequences into SeqLab .....    | 6         |
| Before You Begin .....                   | 7         |
| Starting SeqLab .....                    | 9         |
| Setting your Working Directory .....     | 10        |
| <b>Introduction</b>                      |           |
| <b>to the SeqLab Interface</b> .....     | <b>12</b> |
| Becoming Familiar                        |           |
| with the SeqLab Environment .....        | 12        |
| Icon bar .....                           | 15        |
| Online Help .....                        | 16        |
| <b>Building Project Lists</b> .....      | <b>17</b> |
| Overview .....                           | 17        |
| Introducing List Files .....             | 18        |
| Creating a List File .....               | 18        |
| Entering a New Sequence .....            | 20        |
| Adding Personal Sequence Files .....     | 26        |
| Adding Database Sequences .....          | 28        |
| Saving a List File .....                 | 30        |
| <b>Performing Basic Operations</b> ..... | <b>31</b> |
| Important concepts .....                 | 31        |
| Setting Output Preferences .....         | 32        |
| Simple Find Pattern .....                | 33        |
| Reverse & Compliment .....               | 34        |
| Translating .....                        | 34        |
| <b>Summary</b> .....                     | <b>37</b> |
| <b>Feedback</b> .....                    | <b>38</b> |

## *Overview*

### **In this tutorial you will learn:**

- ❖ What GCG/SeqLab is and how it may work for you.
- ❖ How to get setup to use GCG/SeqLab.
- ❖ Where to get help.
- ❖ How to launch SeqLab.
- ❖ About List Files.
- ❖ How to add sequences into SeqLab.
  - ◆ From Personal Files
  - ◆ Manually
  - ◆ From a Database
  - ◆ From NCBI (Net-Seq-Fetch Extension)
- ❖ How to edit sequences.
- ❖ How to perform basic operations on sequences.
  - ◆ Simple Find
  - ◆ Reverse \ Complement
  - ◆ Translation

# *What is GCG & SeqLab?*

---

GCG (The Wisconsin Package™) is an integrated package of over 130 programs that allows you to manipulate and analyze nucleic acid and protein sequences. Scientists from all over the world have collaborated to develop and refine the GCG, making it a flexible tool with which to analyze your sequence data.

## **Among its many features, GCG gives you access to:**

- **Comparison.** Compare two or more sequences. Create, edit, display, and analyze multiple sequence alignments.
- **Database Searching and Retrieval.** Search nucleic acid or protein sequence databases for sequences similar to your query sequence or sequence pattern.
- **DNA/RNA Secondary Structure.** Predict and display optimal and suboptimal DNA or RNA secondary structures.
- **Editing and Publication.** Enter sequences from a digitizer or a keyboard and edit them.
- **Evolution.** Determine and display evolutionary phylogenies from multiple sequence alignments.
- **Fragment Assembly.** Manage and assemble nucleotide sequence fragments in a sequencing project.
- **Gene Finding and Pattern Recognition.** Recognize terminators, repeats, protein coding regions, and other consensus patterns.
- **Mapping.** Calculate and display restriction digests and simulate RNA fingerprints.
- **Primer Prediction.** Predict optimal primers for PCR reactions.
- **Protein Analysis.** Identify sequence motifs in protein sequences and make predictions about peptide isolation.
- **Translation.** Translate nucleic acids into proteins and backtranslate proteins into nucleic acids.

## **Interfaces to GCG include:**

- **SeqLab.** Based on X Windows, SeqLab provides a point-and-click multiple sequence editor and annotation tool with access to all GCG programs.
- **Command-Line.** Each program steps you through the required parameters it needs to run. Add optional parameters to the command-line for customized analyses. Scriptable command lines provide the ability to do repetitive and numerous analyses.

# Preparations

---

## What you need

### User Account

### Software

- BioDesk
- SSH/SFTP

To use this tutorial you will need an account on Amadeus to access the GCG package. Use the online form at:

**<http://bcf.arl.arizona.edu/online-tools/new-account-form.html>**

You will then be sent a login and password to access the GCG server.

Programs in the GCG package can be run in two ways:

- Command line; UNIX prompt in a command shell
- Graphical User Interface; SeqLab via BioDesk.

Command-line GCG consists of initializing the GCG package by typing `gcg` and then typing subsequent commands.

**This tutorial does not cover command line procedures.**

The preferred method of using the SeqLab graphical interface requires a familiarity with BioDesk\*.

**Please complete our ‘Introduction to BioDesk’ tutorial before continuing.**

You may need a File Transfer program to allow you to transfer files to and from your GCG account folder. We recommend SSH (Secure Shell). SSH Clients generally include both a command line option and a file transfer feature.

Please see the campus software repository for a “no cost” SSH client:

**<https://sitelicense.arizona.edu/ssh/ssh.shtml>**



BioDesk is a complete graphical interface to a variety of UNIX based genomic analysis software. Please see our web site for more information:  
<http://bcf.arl.arizona.edu/biodesk/>

# *Handling Input and Output*

---

## *Importing Sequences into SeqLab*

Many databases and software packages have their own sequence format. To use sequences of a different format with SeqLab, try one of the following:

From the Editor, you can directly import sequences in:

- **FastA**
- **GenBank**
- **GDE**
- **ABI** - Electropherogram
- **SCF** - Electropherogram

If you have sequences in other formats, you must convert them to GCG format before loading them into SeqLab. To do so, GCG provides a number of reformatting programs that work with Staden, EMBL, and PIR files. For other file format conversions, use the readseq command line program on amadeus.

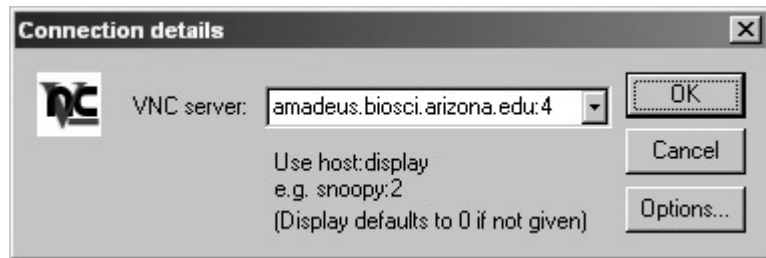
# Getting Started

## Before You Begin

Before you begin working with the tutorial, there are a few preparatory steps:

It is a good idea to create a separate directory into which you can collect sequences and output created from programs during the tutorial. When you are finished with the tutorial, you can simply delete the directory without getting it mixed in with any “real” project data you will create in the future.

1. **Start your BioDesk viewer, connect to your session, and enter your password.**

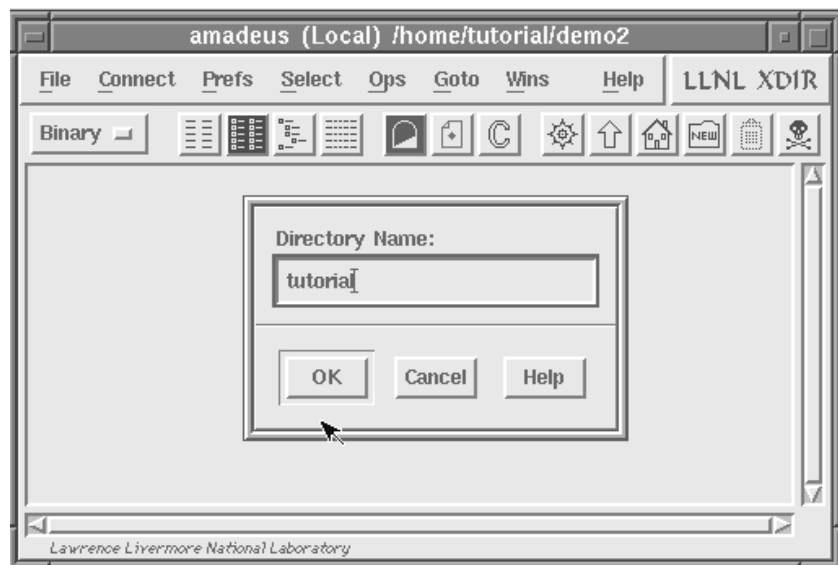


2. **To create a tutorial working directory:**

- a) **Open the BioDesk Menu and select File Manager.**

The FileManager splash-screen will be briefly displayed, then FileManager will appear.

- b) **Create a new directory called 'tutorial'.**



**3. Double-Click on the 'tutorial' folder to open it.**

You should see an empty folder.



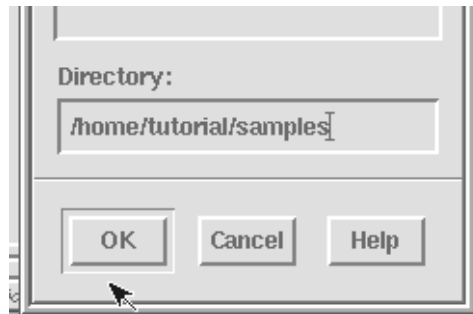
**4. De-Select (toggle) the 'Tunneling Mode' option by clicking on the tunneling mode button.**

This will open a new window when you ask to view a directory that's not already displayed.

**5. Click on 'Goto' from the menubar and select 'Directory...'**

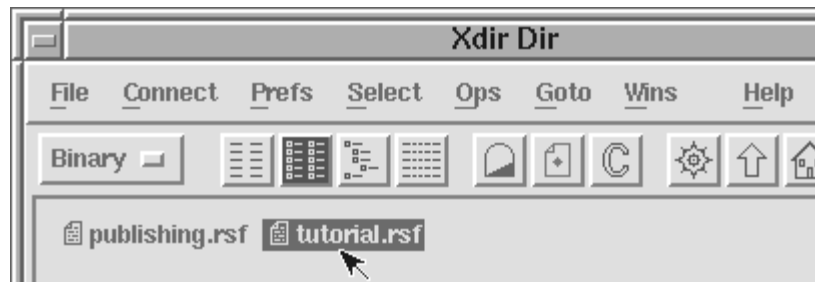


**6. Enter '/home/tutorial/samples'.**



**7. Click on OK.**

A second File Manger window will appear.



**8. Click-Drag the tutorial.rsfs file into the first window.**

This will copy the file into your tutorial folder.

**9. Close both File Manager windows.**

## Starting SeqLab

1. **Open the BioDesk Menu.**
2. **Click on 'GCG /SeqLab'.**

After you launch GCG/SeqLab, the SeqLab Main Window appears, along with a Welcome splash-screen.

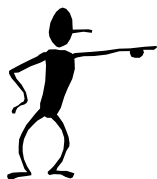


Note the contact information for reaching GCG Technical Support toward the bottom of this welcome window.

10. **Click OK to close the welcome window and reveal the SeqLab Main Window.**

## Setting your Working Directory

Your *working directory* is the directory where SeqLab automatically writes output files.



**Once you set this directory, it stays set from session to session until you change it.**

**No matter what directory you are in when you start SeqLab thereafter, your output files are written to the *working directory* last specified.**

This makes it easy to locate your output files.

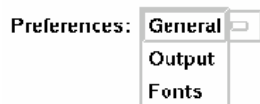
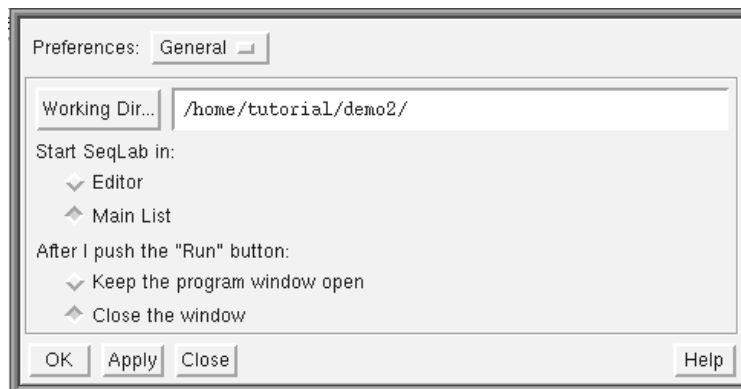
**Note:** If you are sharing a single account, verify that a previous user did not modify the working directory.

The first time you start SeqLab, it sets your working directory to the directory where you started SeqLab.

For this tutorial, let's set your working directory to the directory you created in the "Before You Begin" (pg. 5) section of this tutorial.

### 1. From the Options Menu, choose Preferences....

The User Preferences dialog box appears.

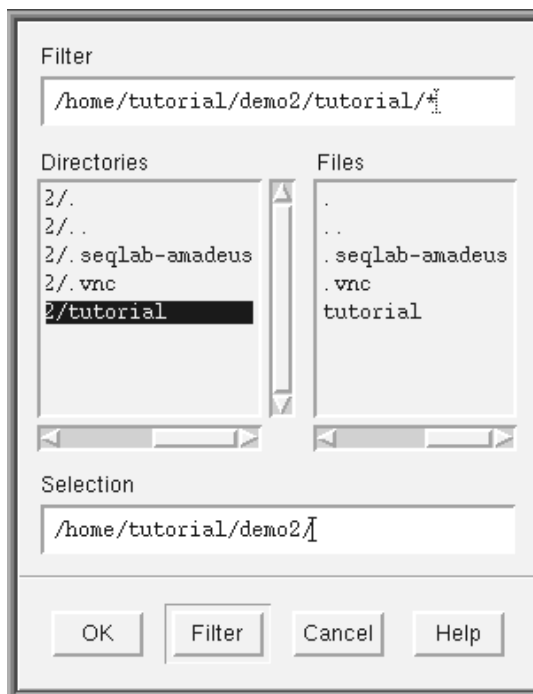


### 2. If necessary, select General from the 'Preferences:' option menu at the top of the dialog box.

### 3. Click the Working Dir... button.

The Change Working Directory dialog box appears.

This dialog box lets you navigate to and select a directory.



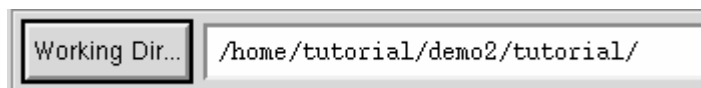
#### 4. To select the tutorial directory:

- Select the 'tutorial' directory in the 'Directories' scroll list and click 'Filter'\*.

You may need to use the vertical scroll bars to scroll down in the list.

- Ensure that the 'tutorial' directory name appears in the 'Selection' text box and click OK.

You return to the User Preferences dialog box, where the specification of your newly selected working directory appears to the right of the Working Dir... button.



#### 5. Click OK in the User Preferences dialog box.

SeqLab applies your changes to the working directory specification and closes the dialog box.



\* For an explanation on *Filters*, see the section on 'Adding Personal Sequence Files', pg 25.

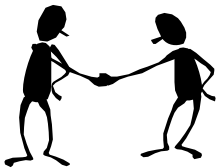
# *Introduction to the SeqLab Interface*

---

**Note:** *Since SeqLab allows you to change font settings, the SeqLab interface shown on your screen may look different than the interface shown in this tutorial. To change font settings, choose Preferences... from the Options menu, select Fonts from the 'Preferences:' option menu, and select from the available font types and sizes.*

## *Becoming Familiar with the SeqLab Environment*

The SeqLab Main Window is your workbench from which you will analyze sequences. SeqLab has two **modes**:



- **Main List**

Provides a good method of specifying the sequences with which you are working. You can use this mode for project management: You can have multiple projects, each with different or common sequences. You can consider each list as a workbench where you can collect selected sequences for further analysis.

- **Editor**

Provides a number of ways for you to visualize your sequences. From this mode you can view, create, or edit sequences; create or edit an alignment; and work with sequence features. You'll learn more about both of these modes as you work through the tutorial.

# 1. Take a few moments to familiarize yourself with elements in SeqLab's Main List mode.

Note that SeqLab will display the contents of the last list file you worked with. The default list file is named *working.list*. If this is the first time you're working with SeqLab, *working.list* will be empty.

In the illustration below, example data is shown to make the description more informative.

## Main List Mode

**Menu bar**  
Pull-down menus list the commands you use to work with sequences.

**Mode**  
The Mode: option menu lets you move easily between the Main List and the Editor.

**List file item attributes**  
For each sequence item listed in the currently loaded list file, SeqLab displays information about the sequence's attributes.

**Currently loaded list file**  
The List: specification displays the name of the list file currently loaded in the Main List.

**List file contents**  
Displays the items in your currently loaded list file.

The screenshot shows a window titled "SeqLab Main window on bangs" with a menu bar (File, Edit, Functions, Extensions, Options, Windows, Help) and a "List:" field containing "/home1/mcb/smith/tutorial/working.list". Below this is a "Mode:" dropdown menu set to "Main List". The main area displays a table with columns "Begin", "End", and "List Item".

|     | Begin | End   | List Item                                  |
|-----|-------|-------|--|
| N + | 1     | 11375 | /usr/users/smith/gamma.seq                 |
| P + |       | 50    | swissprot:hsp1_human                       |
| N + |       | 4379  | gb_pr1:hsp3                                |
| N + |       | 1097  | gb_pr1:hsp35r                              |
| N + |       | 164   | gb_pr1:hsp3mg                              |
| N + |       | 799   | gb_pr1:hsp3rna                             |
| * + |       |       | @/usr/users/smith/hs70.list                |
| * + |       |       | /usr/users/smith/tutorial/tutorial.rsfc(*) |
| * + |       |       | /usr/users/smith/hsp70.rsfc(*)             |

These attributes include:

Indicates the strand orientation: forward (+) or reverse (-) complement.

Specifies the location and name of each item in the currently loaded list file.

|     | Begin | End   | List Item                  |
|-----|-------|-------|----------------------------|
| N + | 1     | 11375 | /usr/users/smith/gamma.seq |

Displays the sequence type: nucleic (N), protein (P), or ambiguous/unknown (\*).

Shows the begin and end ranges of the sequence.



**2. Click on the 'Mode:' option menu and select Editor.**

The SeqLab Main Window switches to Editor mode.

**3. Take a few moments to familiarize yourself with the elements in SeqLab's Editor mode.**

Note that your Editor may be empty. In the Illustration below, we show example data to make the description more informative.

**Editor Mode**

**Display**  
This option menu lets you view sequence data as specially colored text or as a graphical display of features.

**Menu bar**  
Pull-down menus list the commands you use to work with sequences.

**Mode**  
The Mode: option menu button lets you move easily between the Main List and the Editor.

**Icon bar**  
See description on the next page.

**Currently loaded list file**  
The List: specification displays the name of the list file currently loaded in the Main List. When you save sequences in the Editor, they are saved in this list file.

**Horizontal scale**  
The horizontal scale allows you to change by factors of 2 the amount of data displayed within this window—providing a more global view (1:2, 1:4, etc.)

**Sequence names and data**  
SeqLab displays the name of each sequence you load into the Editor, as well as the sequence data.

**Status line**  
The status line helps you determine the location of the cursor within the loaded sequence data.

If you look closely within the sequence data, you can see an "L" highlighted where the cursor is positioned. The status line tells you that the cursor is at position 31 within the sequence hs70\_maize, column 31 within the alignment.

## Icon bar

On the icon bar, you see icons, an option menu, and toggle buttons.

The following icons provide quick access to SeqLab features.



Cuts entire sequences or regions within sequences and places them on the SeqLab clipboard.



Copies entire sequences or regions within sequences and places them on the SeqLab clipboard.



Pastes entire sequences or regions within sequences from the Clipboard. This option is available only after using Cut or Copy.



Lets you set/remove protections for editing sequences, such as allowing ambiguous characters, gaps, standard IUB-IUPAC characters, or reversals.



Displays sequence information, such as sequence name, accession number, description, authors, and reference, for the currently selected sequence.



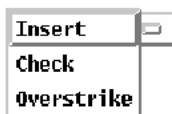
Groups selected sequences to make editing related sequences easier.



Ungroups selected sequences.

**Insert/Check/Overstrike.** The keyboard option menu lets you choose between the following:

- **Insert.** Inserts or deletes characters within a sequence to the left of the cursor position.
- **Check.** Allows you to check that a sequence was entered correctly. When you retype the sequence, SeqLab beeps if a mismatch occurs.
- **Overstrike.** Replaces or deletes characters at the cursor position.



Wrap

**Wrap.** When you select the Wrap toggle button, SeqLab wraps sequence data, along with the ruler, so you don't have to scroll horizontally to see the bases or residues in a sequence. (Note that you may still need to scroll vertically.) By default, the Wrap toggle button is not selected when you display the Editor.

Invert

**Invert.** When you select the Invert toggle button, SeqLab changes the characters to white or black on a colored background. By default, the Invert toggle button is selected when you display the Editor.



4. When finished reviewing the Editor, click on the 'Mode:' option menu and select MainList.

## Online Help

Online help is available from the Help menu on the Main Window. You can also find context sensitive Help buttons on most dialog boxes. Through on line help you have complete access to the Wisconsin Package Program Manual and the SeqLab Guide.

The Help Menu provides help on the following topics:



- ❖ **On the Interface.** Provides a complete online version of the SeqLab Guide, which offers step-by-step instructions on using the SeqLab interface.
- ❖ **On Help.** Provides instructions on using SeqLab's online help and navigating through the online help topics.
- ❖ **On the Wisconsin Package.** Provides a detailed description of each program in the Wisconsin Package. This can also be accessed via web browser at: <http://bcf.arl.arizona.edu/gcg.html>
- ❖ **On SeqLab Version.** Displays the software version and copyright information.
- ❖ **On Database Versions.** Displays the latest versions of the databases installed at your site.



You can search the available Online GCG documentation at:

<http://bcf.arl.arizona.edu/>

by following the FAQs and HowTo's link

# Building Project Lists

## Overview

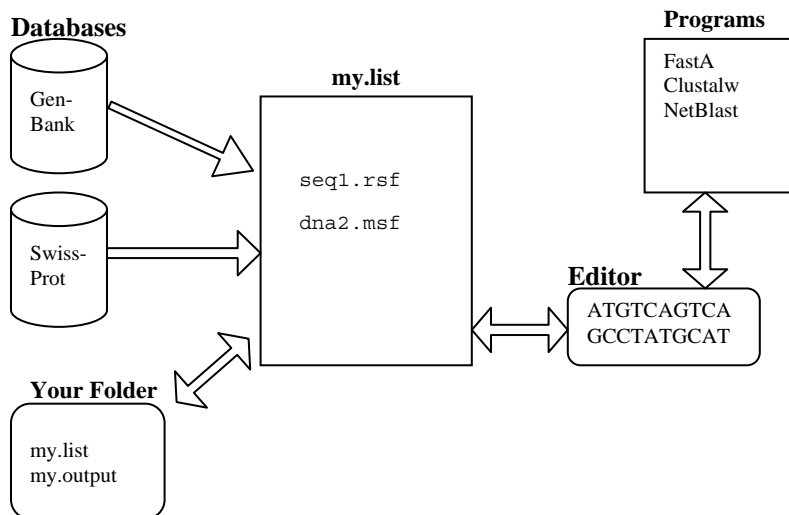
- **Creating List File**
- **Loading Sequences**
- **Set Editor preferences**
- **Methods for inputting a New Sequence**
- **Rename Sequence**
- **Protect a Sequence**
- **Saving Sequences**
- **Expanding RSF, list or MSF files**

SeqLab provides an easy way to specify sequence files and database entries on a project-by-project basis. You do this by adding entries to a *list file*, a concept you will be familiar with by the end of this lesson.

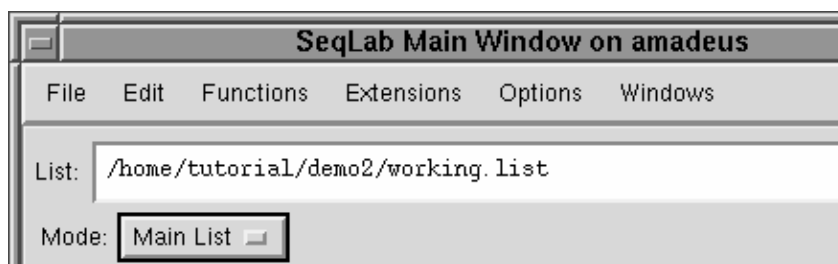
You will find that Main List mode serves as a handy project management tool, helping you specify related sequences.

For this lesson you will create an *Arabidopsis thaliana* project list file. You will learn how to add sequence file entries, as well as how to locate and work with public database entries. You will also enter and save a short segment of DNA you recently sequenced from *A. thaliana*.

The diagram below is a conceptual model of the data flow in SeqLab.



## Introducing List Files



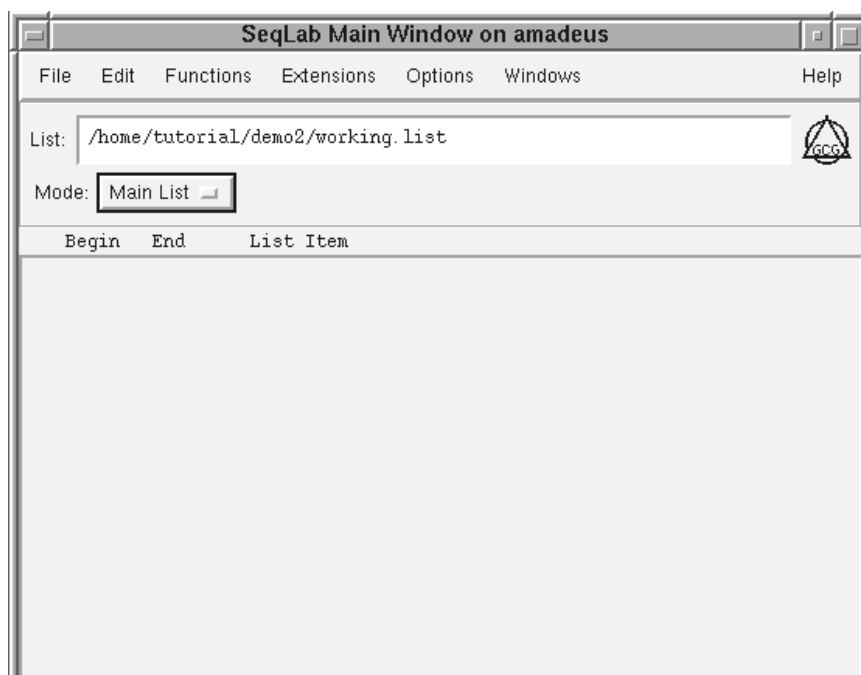
Note the 'List:' text box toward the top of the SeqLab Main Window; it shows the currently loaded list file.

*List files* allow you to specify sequences located in different directories and databases. For example, you can specify sequence entries from public or private databases, or sequence files from your own personal directories.

## Creating a List File



1. If necessary, click on the 'Mode:' option menu and select **Main List**.

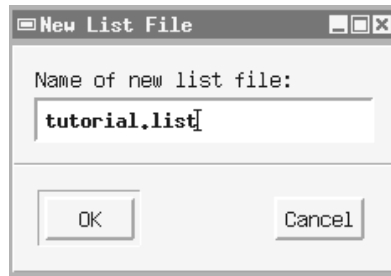


2. To create a new list file:

The first time you work with SeqLab, the program opens a new list file named 'working.list' that you can use to begin organizing your project sequences. For this tutorial, however, let's create a new list file named tutorial.list.

**a) From the File menu, choose New List....**

The New List File dialog box appears.



**b) Type tutorial.list in the text box.**

**c) Click OK.**

*Note: If you made changes to your previously loaded list, SeqLab prompts you to save any changes before opening up the new list file.*

The New List File dialog box disappears and the 'List:' file specification on the Main Window displays the name of your newly created list file.

```
List: /home/tutorial/demo2/tutorial/tutorial.list
```



**Make sure you name your list file with a .list extension (e.g. project16a.list).**

## Entering a New Sequence

Now let's enter a short sample sequence. In Editor mode, SeqLab allows you to enter new DNA, RNA, or protein sequences. For this lesson, you'll create a DNA sequence.

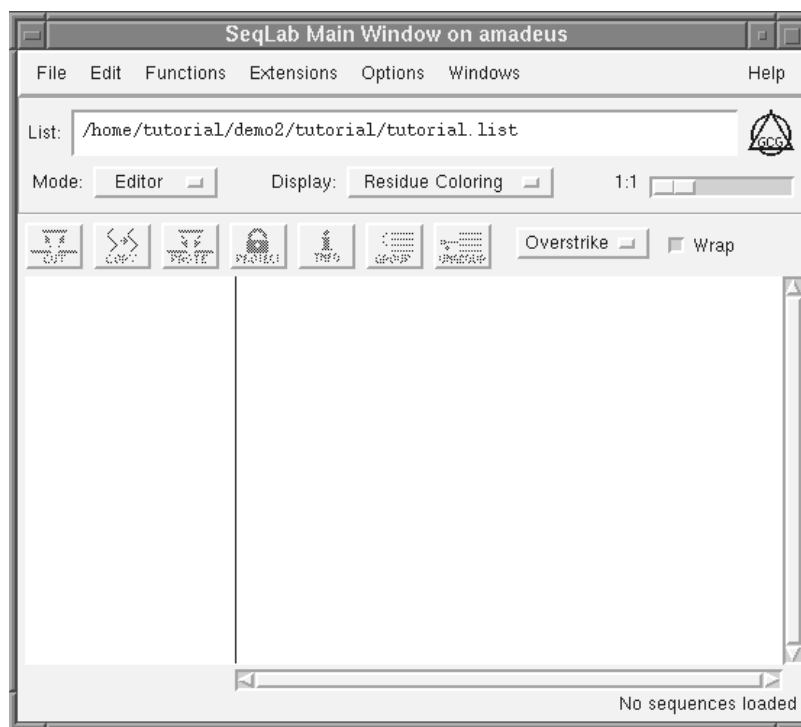
### 1. Make sure nothing is selected in your Main List.

If a sequence is selected, press and hold the <Ctrl> key and click your cursor on the selected sequence(s). The *control-clicking* should deselect a selected sequence.



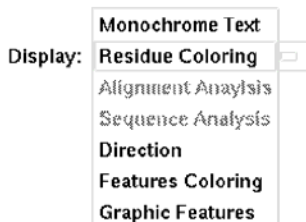
### 2. Click on the 'Mode:' option menu and select Editor.

The Editor window appears.



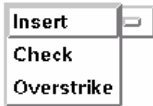
### 3. To set Editor option:

On the Editor window, there are several options you can set for displaying sequences and entering sequence information.



#### a) Click the 'Display:' option menu and select Residue Coloring.

This color mode displays all similar residues with same color (e.g. isoleucine, leucine, methioine, and valine are green).



Wrap

Invert

**b) Click the keyboard option and select Insert.**

This mode allows you to insert or delete sequence characters to the left of the cursor position.

**c) Note the Wrap toggle button.**

When you select the Wrap toggle button, SeqLab wraps sequence data, along with the ruler, so you don't have to scroll horizontally to see the bases or residues in a sequence.

Since you'll be typing a short sequence, you don't need to select the Wrap toggle button.

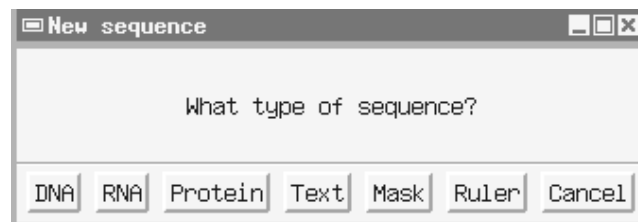
**d) Click then Invert toggle button a few times...**

This mode displays sequence characters in white or black on a colored background.

**4. To create a new sequence:**

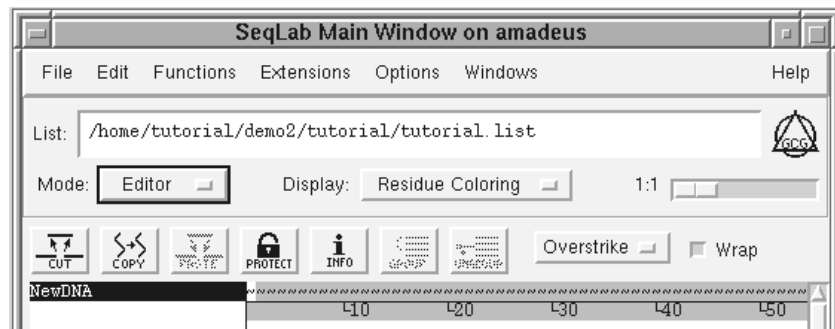
**a) From the File menu, choose New Sequence....**

The New Sequence dialog box appears prompting you "What type of sequence?"



**b) Click DNA.**

The dialog box disappears and you'll see a new sequence titled NewDNA in the Editor window, followed by a row of tildes (~). The tildes act as a space holder until they are replaced by "real" sequence data.



You can use "text" in the New Sequence dialog to leave notes/instructions long with your sequence data in the editor window.

**5. To rename a sequence:**

- a) **Make sure the sequence name NewDNA is selected.**
- b) **Click the Info icon.**



The Sequence Information dialog box appears

You can use this dialog box to rename a sequence, include a description of or comments about a sequence, or change the sequence attribute information, such as the sequence orientation or topology.

- c) **In the 'Name' text box, change NewDNA to manual\_dna.**

- d) **Click OK.**

The dialog box disappears and the new name appears in the Editor window.



You can use the Sequence Information windows to assign internal accession numbers, notes and detailed information specific to the sequence.

**6. Click your cursor on the first tilde (~) of the sequence data and type a sequence of about 20 bases:**

Just a random sequence, for this step it really doesn't matter.

You can use lowercase or uppercase text – GCG programs do not differentiate between them. If you make a mistake, just use the right and left arrow keys to move your cursor to the correct position, press <Delete> to remove the character to the left of the cursor, and type the correct character.

Because you set the “Display:” option menu to Residue Coloring, SeqLab displays the sequence characters as multi-colored text, with each nucleotide represented by a different color.

**7. Click your cursor on the first nucleotide, hold the shift key down and press the right arrow key.**

The entire sequence will shift to the right.

Take a moment to play with this feature; move the sequence, or parts of it back and forth. Manual alignments can be easily performed this way.

**8. Protect your sequence from accidental modification:**

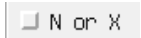
After entering a new sequence, you may want to define what sequence modifications are allowed. You can set sequence protections to allow unknown residues (N or X for nucleotides, or X for proteins), gaps, other changes, or reversals/complimenting.

**a) With your manual\_dna sequence name selected, click the Protect icon.**

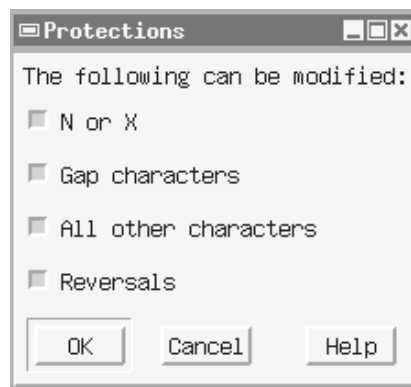
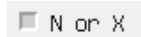


**Button In? Button Out?**

This button is *out*  
(Deselected):



This button is *in*  
(Selected):



The Protections dialog box appears.

When you create a new sequence, SeqLab allows *all* possible modifications. Now that you have entered your sequence and made any necessary changes, you may want to protect it from further modifications.

**b) Deselect all of the options in this dialog box and click OK.**

The protections are changed for the sequence and the dialog box closes.

- c) What happens if you try to modify a sequence you've protected?  
Click in the sequence data and try to insert or delete characters.**

SeqLab beeps and displays a "Cannot insert..." message box or text in the status line at the bottom of the window. Thus, your sequence is protected from modification. To remove the "Cannot insert" message box, click OK.

If you have trouble editing existing sequences, check the sequence protections as described earlier in this step.

## 9. To save your sequence:

Because this new sequence is part of the current project you've been working on, you'll want to add it to the list file loaded in the Main List.



- a) From the File menu, choose Save As....**

The Save As dialog box appears.



- b) Change the 'File name:' specification to manual.rsff.**

- c) Click OK.**

SeqLab saves the sequence as an RSF (Rich Sequence Format) file and adds it to the list file loaded in the Main List.



**It's a good idea to save all your sequences in RSF format. You can convert from RSF to other formats without losing information like comments and annotations.**



## 10. Click on the 'Mode:' option menu and select Main List.

Note the new sequence file `manual.rsff` that appears at the top of the list.

## 11. Double-click on the filename `manual.rsff`.

```
/home/tutorial/demo2/tutorial/manual.rsff{*}  
/home/tutorial/demo2/tutorial/manual.rsff{MANUAL_DNA}
```

The entry expands to display the one sequence, `manual_dna`, contained within the RSF file. (Since RSF & MSF can store multiple sequences as a single file, you can expand RSF, list, and MSF files in this way to view a list of the sequences they contain.)

RSF files can contain one or more sequences as well as descriptive information such as author, sequence weight, creation date, sequence description, and known sequence features. RSF files get their name from the "rich" annotation of sequences features which they may contain.

Collapse the expanded RSF file by double-clicking on the filename `manual.rsff` again.



## 12. From the File menu, choose Save List.



### Navigation Tip:

Click on the sequence name, type a number and hit enter; it will position the Editor cursor to that base pair number.



An `{*}` after a file name suggests that it can contain more than one sequence (i.e. multi-sequence format).

## Adding Personal Sequence Files

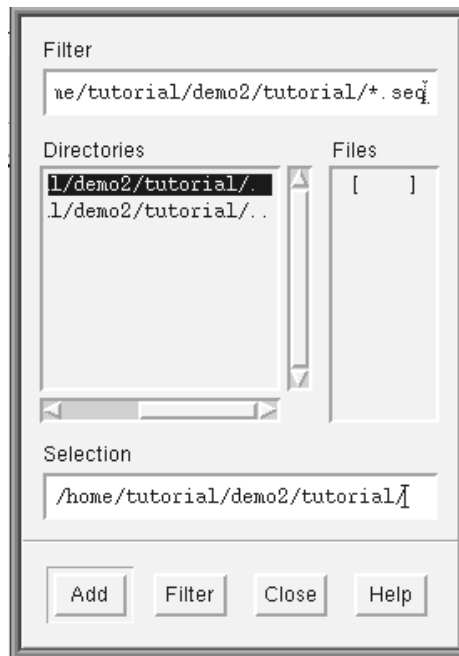


One of the first things you will want to do in SeqLab is add sequences. For example, you can add sequence files from your directories.

**The files must be in GCG format.** (See “Handling Input and Output” later in this tutorial.)

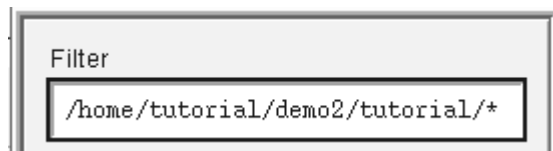
1. From the File menu, choose ‘Add Sequences From’ → ‘Sequence Files...’

The Add Sequences dialog box appears.



2. To change the filter SeqLab is using to display files:

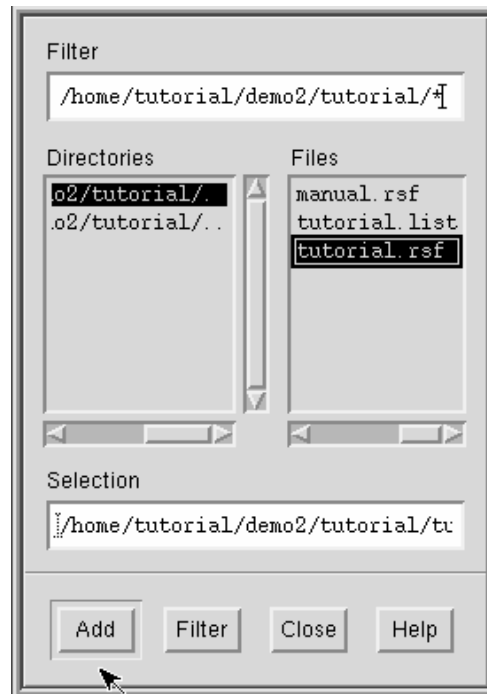
The Filter text box at the top of the dialog box helps you narrow the sequences displayed in the Files scroll list. Your current filter should end with \*.seq. The \* is a wildcard that means “anything or nothing”. Therefore, the \*.seq filter will display all files in your working directory that end with .seq.



- a. Click the cursor in the ‘Filter’ text box at the top of the dialog box and change \*.seq to \*

**b. Click the Filter button.**

The 'Files' list displays all files in your working directory.



**3. In the 'Files' list, select tutorial.rsf.**

**4. Click Add.**

**5. Click Close.**

The dialog box disappears and the Main List displays your newly added sequence file.



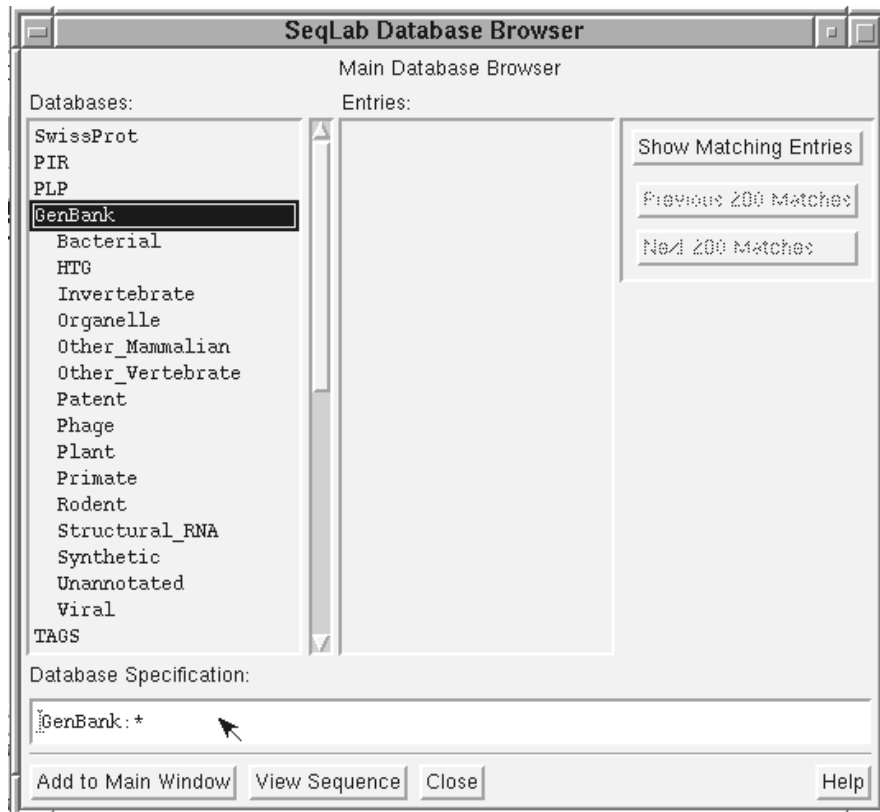
When you add sequences from public databases or import them into GCG, the default protections disallow modifications.

## Adding Database Sequences

GCG includes local copies of the GenBank, PIR, SWISS-PROT, and SP-TREMBL databases. However, more up-to-date versions of these databases are available on the web.

1. From the File menu, choose Add Sequences From → Databases....

The SeqLab Database Browser appears.



The Database Browser lets you view individual entries within the databases. More importantly, the Database Browser lets you add sequences to the list file you have loaded in the Main List.

To use the Database Browser effectively, you must know either the name or accession number of a database entry.

### To add an entry from the Database Browser:

A journal article of interest referred to a nucleic acid sequence *A. thaliana* by accession number Z12136.

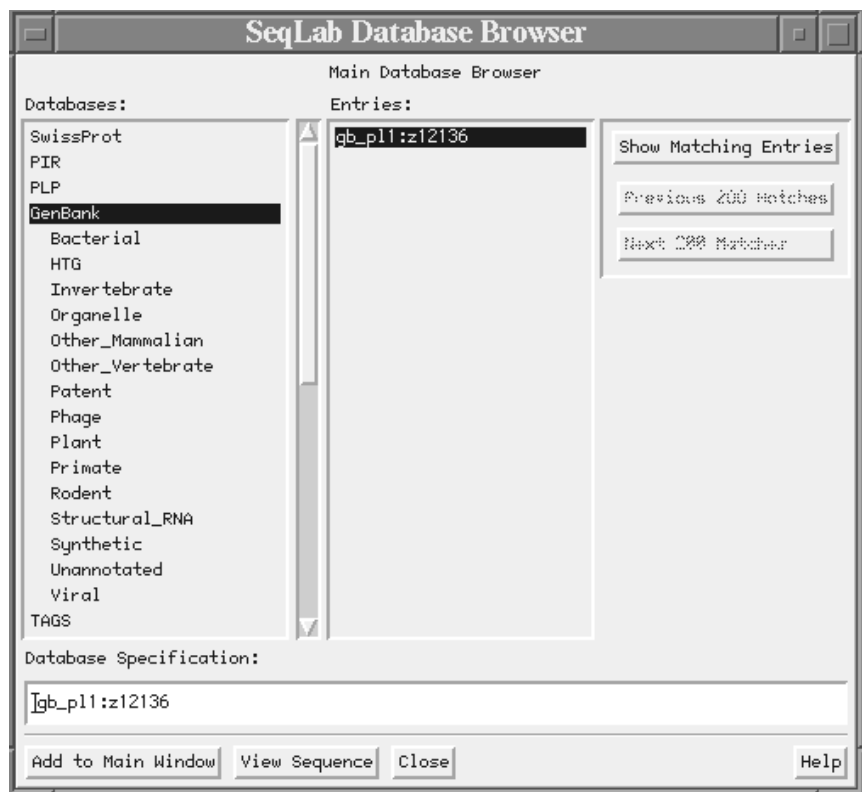
**a) Scroll down the ‘Databases’ list and select GenBank.**

GenBank is the complete nucleic acid database provided by GCG that includes all EST, STS, and GSS sequences.

**b) In the ‘Database Specification:’ text box, replace GenBank:\* with GenBank:z12136.**

**c) Click Show Matching Entries.**

The matching entry appears in the ‘Entries:’ list.



**d) Select the Z12136 entry in the ‘Entries’ list.**

**e) Click Add to Main Window**

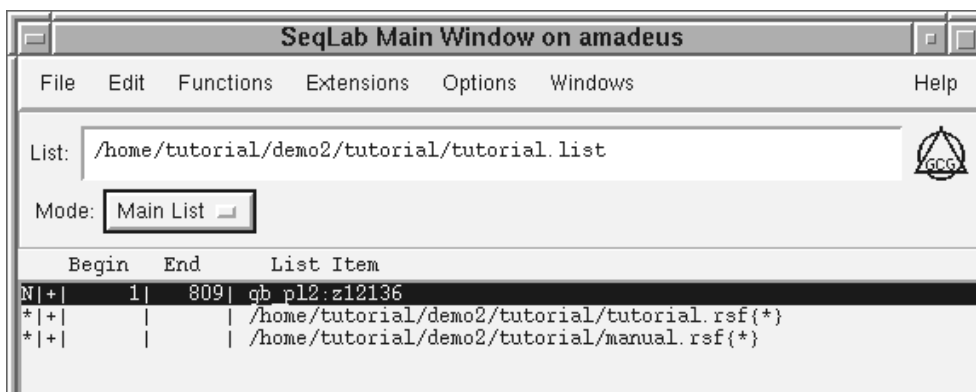
**Note:** Entries are not case insensitive.



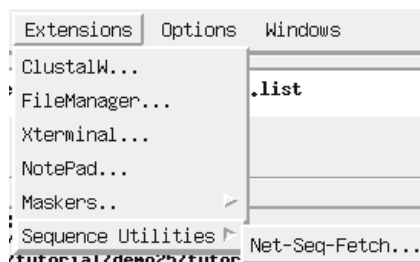
If you want to browse all sequences starting with ‘ze’, you can enter GenBank:ze\* and click ‘Show Matching Entries’.

**f) Click Close.**

The dialog box disappears and the Main Window displays your newly added sequence files.



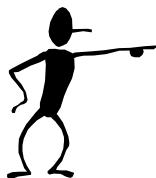
To retrieve sequences directly from NCBI you can use the Net-Seq-Fetch utility that BCF has incorporated into SeqLab. It can be found on the Extensions menu under Sequence Utilities.



However, this requires some familiarity with the Output Manager, which we have not yet covered.

### *Saving a List File*

It is a good idea to save your work every so often in case there is a power shortage or you lose your connection to the network.



► **From the File menu, choose Save List.**

SeqLab updates tutorial.list to include your newly added sequences.

# Performing Basic Operations

## Overview

### Programs

- Output Preferences
- Finding simple patterns
- Reverse & Compliment
- Translating

The GCG (Wisconsin Package), upon which SeqLab is built has over 130 programs you can use to do different types of sequence analysis. In SeqLab, many of these programs are grouped by function (for example database searching, mapping, or fragment assembly) and appear under the Functions menu.

A few of the most common operations, such as pattern finding, translating, reversing/complementing, and creating a consensus sequence, are available from the Edit menu within the Editor.

This lesson teaches you how to select and run operations in SeqLab. Although each program has a different analysis and a unique set of options for customizing that analysis, the basics of selecting input sequences, running a program, and viewing the output are the same.

For this lesson you will use the tutorial.rsf sequence you copied to your directory in the Getting Started section.

If you would like to browse GCG programs by category, select the individual category in the Function menu to view the program. For program descriptions follow the FAQs and HowTo's links at:

<http://bcf.arl.arizona.edu/>

## Important concepts



The **Job Manager** keeps track of the programs started by SeqLab and displays their status, so you know if a program is running, completed, waiting in the batch queue, or stopped because of an error. You will often keep the Job Manager window open in the background as you work with SeqLab.

The **Output Manager** keeps track of the output files created in your current session with SeqLab. From this window you can display and print text and graphic files, as well as save files with a new name, delete files, and add output files to the Main List and Editor. The Output Manager also has options that let you load text and graphic files created in previous sessions. You often will keep the Output Manager window open in the background as you work with SeqLab.



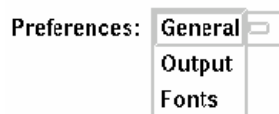
If you know the program name, but are unsure of the category, select the 'Alphabetical' listing from the Function menu.

## Setting Output Preferences

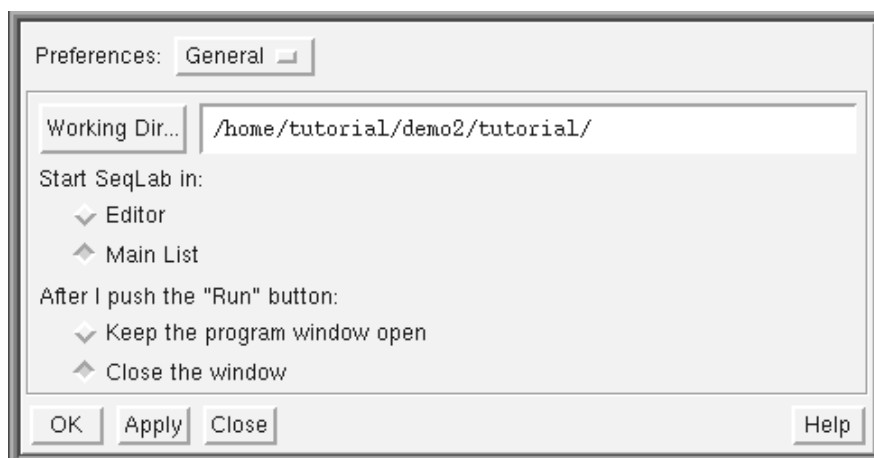
There are a number of preferences you can set for running programs and working with output files.

### 1. From the Options menu, choose Preferences....

The User Preferences dialog box appears

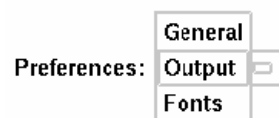


### 2. Select General from the 'Preferences:' option menu at the top of the dialog box.



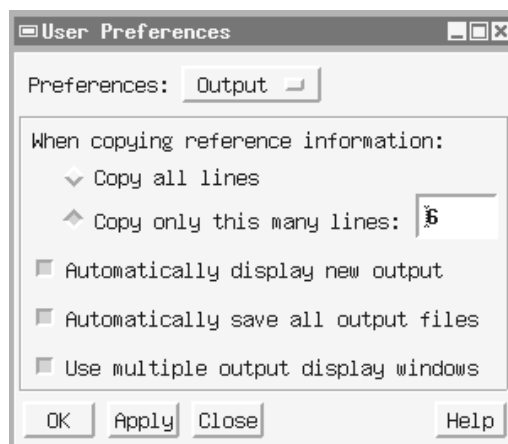
### 3. Set the 'After I push the 'Run' button' option to 'Close the window'.

This preference automatically closes a program window after a program begins to run.



### 4. Click Apply.

### 5. Select Output from the 'Preferences:' option menu at the top of the dialog box.



## 6. Select 'Automatically display new output.'

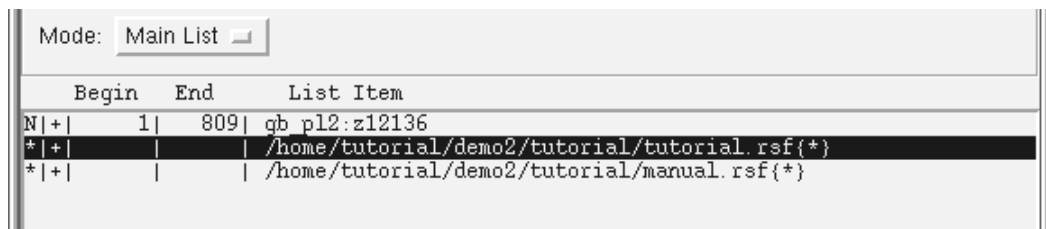
This option tells SeqLab to automatically display output when a program finishes running.

## 7. Click OK.

SeqLab applies the changes and closes the dialog box.

## Simple Find Pattern

Let's use the Find operation to locate some likely patterns in our sample sequence.



The screenshot shows a window titled 'Mode: Main List'. It contains a table with the following data:

|     | Begin | End | List Item                                    |
|-----|-------|-----|--|
| N + | 1     | 809 | qb pl2:z12136                                |
| * + |       |     | /home/tutorial/demo2/tutorial/tutorial.rsfl* |
| * + |       |     | /home/tutorial/demo2/tutorial/manual.rsfl*   |

1. Select tutorial.rsfl from the Main List.
2. Click on the 'Mode:' option button and select Editor.
3. Select the sample\_dna sequence.
4. Click on 'Edit' from the menu bar and select 'Find...'

The 'Find' dialog box appears.



5. Enter atg in the 'Pattern:' window and click on 'Find All', then click on 'Cancel'.

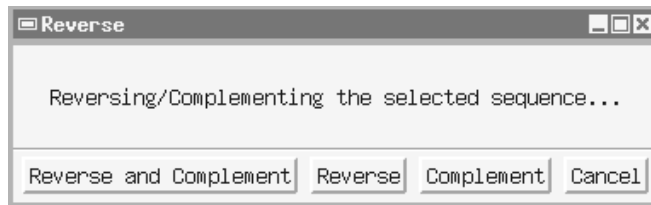
The sequence coloring changes to highlight all instances of the given pattern.

6. Click on the Display menu change it from 'Find' back to 'Residue Coloring'

## *Reverse & Compliment*

Now, let's find the reverse and compliment of our sample sequence.

1. Select our sample\_dna sequence in the Editor.
2. Click on 'Edit' from the menu bar and select 'Reverse...'



The Reverse dialog box appears.

3. Click on 'Reverse and Complement'.

The Editor now displays the Reverse and Compliment of our sample sequence.



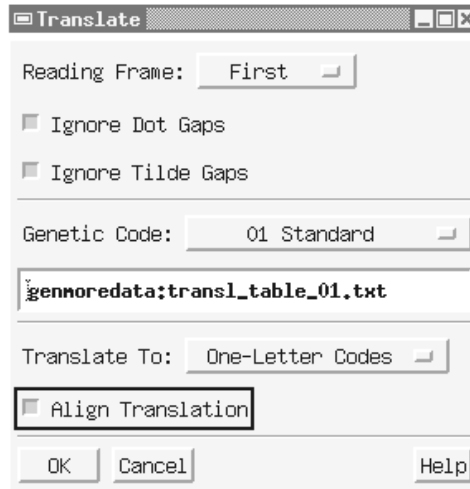
4. We don't really need the reverse of this sequence; you can easily reverse the process by going through the last three steps again.

## *Translating*

Let's translate the sample DNA sequence to a protein sequence. To do so, we will use the Translate operation.

1. In the Editor, select the sequence named sample\_dna.

**2. From the 'Edit' menu, select 'Translate...'**



The Translate dialog box appears.

**3. Set the appropriate translation parameters and run the program.**

Let assume that the first open reading frame is the most likely (the Frames program can find and display ORFs for you). Verify the following translation parameters to:

'Reading Frame:' First  
'Genetic Code:' 01 Standard  
'Translate to:' One-Letter Codes  
'Align Translation' Selected



**4. Click OK.**

The translated sequence appears below the original DNA sequence. The new sequence is named manual\_dna\_frame1.

**5. To rename the translated sequence:**

**a) Select the sequence name sample\_dna\_frame1.**

**b) Click the Info icon.**



The Sequence Information dialog box appears.

**c) In the 'Name:' text box, replace sample\_dna\_frame1 with sample\_protein.**

**d) Click OK.**

The Sequence Information dialog box disappears and the Editor displays two sequences; one DNA, one protein.

```

sample_dna  ATGGCGGACCAAGCTCACCGACGAACAGATCGCCGAGTTCAAGGAGGCCITCAACCTCTTCGACAAAGGACGGCGCACGG
sample_protein M A D Q L T T D E Q I I A E F K E A F S L F D K D G D G
L10 L20 L30 L40 L50 L60 L70
sample_dna  CTGCATCAACCAACAAAGGAACCTTGGCACTGTGATGCCGCTCATTGGGGCAGAACCCCTACTCAGGCTCAGCTTCAGGACA
sample_protein C T T T T K E L G T V M R S L G Q N E T E A E L Q D M
L80 L90 L100 L110 L120 L130 L140 L150
sample_dna  TCATCAATCAGGTTGATGCTCATGGCAATGGAACCAATCCACTTTCCTCAGTTTCTCAACCTCATGGCAAGCAACATG
sample_protein I N E V D A D G N G T I I D F F E F L N L K A R K N
L160 L170 L180 L190 L200 L210 L220 L230
sample_dna  AAGGACACCGACTCTGAGGAGGAGCTCAAAGAGGCCITTCGGTGTGTTCCGACAAAGGACCAAGAACGGCTTCATCTCCGC
sample_protein K D T D S E E E L K E A F S W D K D Q N G F I T S A
L240 L250 L260 L270 L280 L290 L300
sample_dna  TGCTCAGCTTCGCCATGTCATCAACCAACCTAGGCCACAAAGCTCACGGACGAGGAGGTGCACAGATCATCCGTGAAG
sample_protein A E L R H V M T N L G E K L T D E E V D E N I R E A
L310 L320 L330 L340 L350 L360 L370 L380
sample_dna  CCGACGTGCACGGTGCATGCCAGATCAACTACGACGAGTTCGTCAAGGTCATCATGGCCAACTGA
sample_protein D V D G D G Q I N M D E F V K V W W A K *
L390 L400 L410 L420 L430 L440 L450 L460

```

pos:1 col:1 sample\_dna -->

**6. To save your changes:**

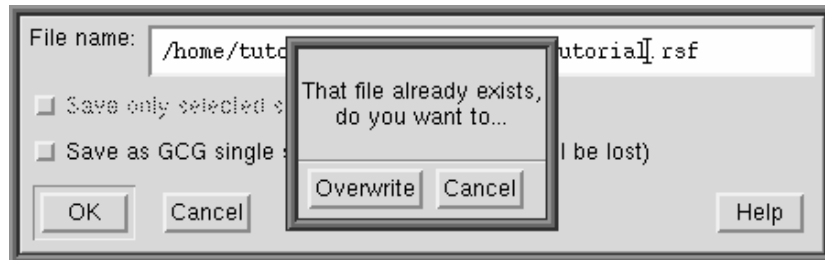


**a) From the File menu, choose Save As....**

The Save As dialog box appears.

**b) In the 'File name:' text box, type tutorial.rsf.**

**c) Click OK.**



SeqLab displays a prompt notifying you that a file by that name already exists.

The sample\_dna sequence you loaded in earlier in the lesson was saved in a file named tutorial.rsf. You now want to save a new version of tutorial.rsf with two sequences, sample\_dna and sample\_protein.



**d) Click Overwrite.**

Both sequences are saved in the file tutorial.rsf and are added to the list file currently loaded in the Main List.

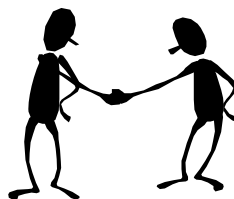
7. **Change to Main List Mode.**
8. **Click on 'File' and select 'Save List'.**
9. **Click on 'File' and select 'Exit'.**
10. **Close your BioDesk viewer, unless you are using one of our demo accounts, then you should use the EXIT button to kill the session.**

---

## *Summary*

### **In this tutorial you learned how to:**

- ❖ What GCG/SeqLab is and how it may work for you.
- ❖ How to get setup to use GCG/SeqLab.
- ❖ Where to get help.
- ❖ How to launch SeqLab.
- ❖ About List Files.
- ❖ How to add sequences into SeqLab.
  - ◆ From Personal Files
  - ◆ Manually
  - ◆ From a Database
  - ◆ From NCBI
- ❖ How to edit sequences.
- ❖ How to perform basic operations on sequences.
  - ◆ Simple Find
  - ◆ Reverse / Complement
  - ◆ Translation



# Introduction to GCG/SeqLab

## Part A

### Feedback

Date: \_\_\_\_\_

**How would you rate this Workshop in the following areas (1=poor, 10=excellent)?**

Quality of Content \_\_\_\_\_

Quantity of Content \_\_\_\_\_

Handout \_\_\_\_\_

Information Covered \_\_\_\_\_

Presentation \_\_\_\_\_

Overall Evaluation \_\_\_\_\_

Would you recommend this workshop to your students, colleagues, etc? (y/n) \_\_\_\_\_

What additional areas in bioinformatics/computing would you like training/workshops in?

Additional Comments/Suggestions?

Thank you for your Feedback!