

## **DNA sequence base calling, contig assembly and editing using *phred*, *phrap*, and *consed* via Secure Shell (SSH) connection**

The unix/linux phred software reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base. Phrap is a program for assembling shotgun DNA sequence data. Consed/Autofinish is a tool for viewing, editing, and finishing sequence assemblies created with phrap. Phred and phrap are commonly used by DNA sequencing facilities and large scale sequencing projects. Consed is a great tool for the end user to evaluate the sequencing data, correct basecalling errors, detect low-quality regions that will require further sequencing, etc. These tools are free for academic users, and can be obtained from <http://www.phrap.org/phredphrapconsed.html>.

These are unix programs. It is convenient to run the software on a networked linux/unix server accessible to the entire research group and use the software remotely via telnet or SSH connection from individual researcher's PC (Win or Mac) computers. The final output (a FASTA sequence file) can then be downloaded to the PC.

The linux server needs the phred/phrap/consed software properly installed. Additionally, it needs to allow ssh connections from the user's computers, and have X-windows installed. It is also useful to have the "Midnight Commander" utility installed for easier navigation and file management. The Linux server can be a fairly low-end computer. I have used a Pentium 200MHz system with 64 MB memory for assembly and analysis of small configs (up to about 50 sequencing reads, i.e., sequencing of individual genes) without any problem.

On the client PCs, the following two software packages must be installed:

- SSH (Secure Shell) client to connect securely to the Linux server.
- X-server to display the graphical output from the linux server on your PC screen. Free x-servers are scarce, but several commercial ones cost less than \$30. Also check free X-server from the cygwin project (<http://www.cygwin.org>).

The following steps describe how to process sequencing data in the remote session.

1) Start SSH Secure Shell Client and connect to the Linux machine

If login was successful, you are now logged on the linux server and should see a message like this:

```
Last login: Fri Apr 26 13:27:33 2002 from someuser.tamu.edu
```

2) Unless it is done automatically by the ".profile" or ".bashrc" file during login, you must change the DISPLAY environment variable to tell the linux server where to send the graphical output of the consed program. You need to know the IP address or host name of your client computer. To use the numerical IP address, find it out in the command prompt (DOS) window on your computer by running **ipconfig**. Once you

know the hostname or IP address (e.g.123.456.78.9), issue the following command on the linux machine, using the obtained IP address:

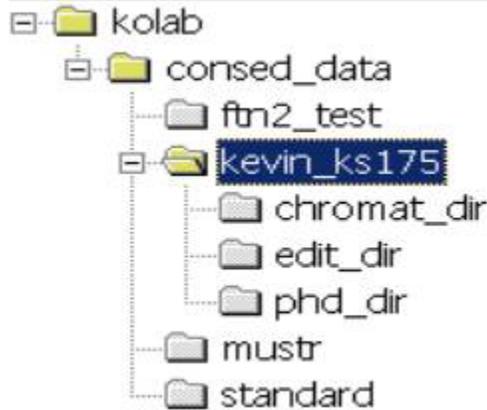
```
$ DISPLAY=123.456.78.9:0.0 <ENTER>
```

3) For easier navigation on the linux machine, start the Midnight Commander (“mc”):

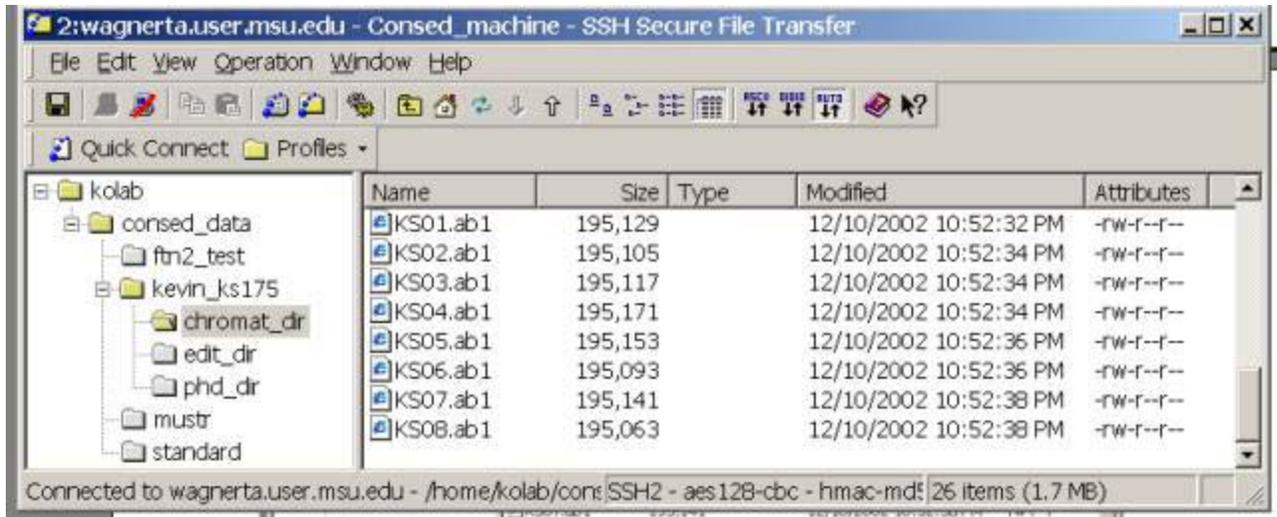
```
$ mc <ENTER>
```

Midnight Commander is used in the same way as the “Norton Commander” used to be in MS DOS.

4) If not already present, create a new folder, named “consed\_data”. In this folder, create a new folder for your sequencing project, with three sub-folders: chromat\_dir, edit\_dir, and phd\_dir (see example below).



5) Upload sequencing chromatogram files (\*.abi or \*.ab1 files) from your PC to the “chromat\_dir” folder (most SSH clients have the ability to do FTP transfers). When done, you will have several \*.ab1 files in the "chromat\_dir" folder. **It is important that your file names do NOT contain any spaces. If necessary, rename the files.** Go back to the SSH terminal window, navigate into the chromat\_dir where you should see the files you just uploaded.



6) To perform base calling contig assembly, run the phredPhrap “script”

```
$ phredPhrap <ENTER>
```

Things will start happening, you should see text messages scrolling on the screen. If all went well, the last message should say: you may now run consed .... Your sequencing traces have been processed, base calling performed and the contig(s) assembled. The assembly file (\*.ace.1) may now be used to display and edit the config.

7) Start consed:

```
$ consed <ENTER>
```

A new window will show-up in the X-server window. Select the desired “\*.ace” file (usually the one with the highest number) and click “display” button. The \*.ace file will be loaded and the contig(s) and individual reads listed. Select the contig of interest and click “Show contig”. Read the consed manual to learn how to use it. A screenshot of the consed program environment is shown below.

