SAGE ABI Sequence File Processing
Updated March 15, 2006

This example uses the "*aureococcus*" account, but the protocol will work for any SAGE account.

Transfer file from ABI3730 to UNIX account (Mac OS X example):

You will receive an email from the sequencing staff to let you know when the sequence data will be ready to download. Also included in the email is the percent successful reads and average length of read. Click on the link in the email and enter in the username and password. Then click on the name of the sequencing plate you wish to download (the file will then be un-zipped to your Desktop).

Open a terminal and transfer the data to your SAGE account at MBL:

```
cd ~/Desktop
tar cvf name_of_file.tar name_of_file
gzip name_of_file.tar
scp name_of_file.tar.gz aureococcus@evol5.mbl.edu:/habitat/aureococcus/sagedata/.
```

Now use Secure Shell (ssh) to log into the aureococcus account on FOG.MBL.EDU:

```
ssh -l aureococcus evol5.mbl.edu
ssh -l aureococcus fog.mbl.edu
cd ~/sagedata
gunzip name_of_file.tar.gz
tar xvmf name_of_file.tar
rm name_of_file.tar
```

Process ABI files using UNIX account:

Move into the "sageanalysis" directory:

```
cd ~/sageanalysis
```

Run the "sagetags_linux" program, giving new data folder names as arguments (you can enter multiple folder names). For example:

```
sagetags_linux name_of_file
```

The "sagetags_linux" program will process all of the data to extract sage tags and their frequencies. You can also run "sagetags_linux" without arguments to process the existing data. You will get an email giving an overview of the results.

Viewing Details of the Results (Optional):

Once "sagetags_linux" is done working, you can move into the results folder to look at the detailed results:

```
cd ~/sageanalysis/results
ls
```

Here is an example of some of the output files and what they contain. For each library (A library in this example) you will find:

```
042804.A.fasta.screen.qual.adj              vector-screened, quality adjusted raw sequence data
042804.A.fasta.screen.qual.adj.ambigtags    tags with sequencing ambiguities
042804.A.fasta.screen.qual.adj.ditags       ditags and their frequencies
042804.A.fasta.screen.qual.adj.nonditags    NlaIII fragments outside of the ditag size range
042804.A.fasta.screen.qual.adj.rot          return of effort curve data
042804.A.fasta.screen.qual.adj.sagelog      text log of SAGE processing
042804.A.fasta.screen.qual.adj.tags         tags and their frequencies
```

There are also files for all libraries combined:

```
042804.alllib.summarystats                  summary stats table for all libraries (tab-delim)
042804.alllib.tags                          export file of tags and all their frequencies
```

Obtain Summary Results File to Upload to GMOD Database:

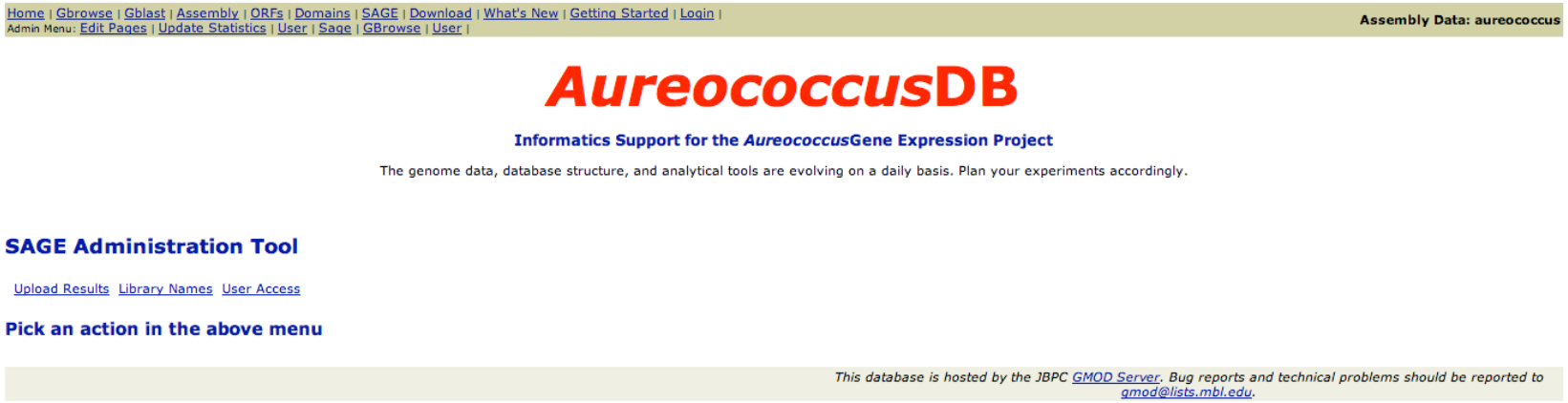
You only need one of the results files to upload to your GMOD database (*date*.alllib.tags). Log out of the MBL unix accounts using the "exit" command until you are back on your own computer. Use the following command to copy the *date*.alllib.tags to your desktop (using March 15, 2006 data as an example):

```
scp aureococcus@evol5.mbl.edu:/habitat/aureococcus/sageanalysis/results/031506.alllib.tags ~/Desktop/.
```

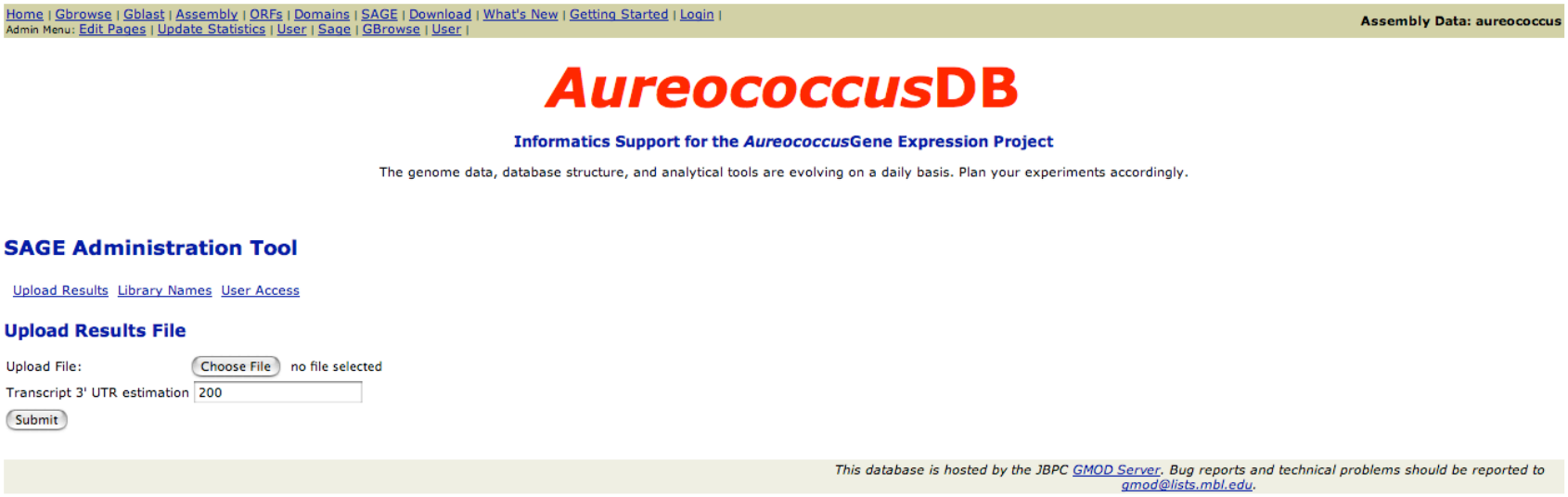
Update GMOD:

To upload the tags file to GMOD, you first must log into the GMOD database as an administrator (your account may already be set-up for this). Select the "Sage" button on the second menu bar:

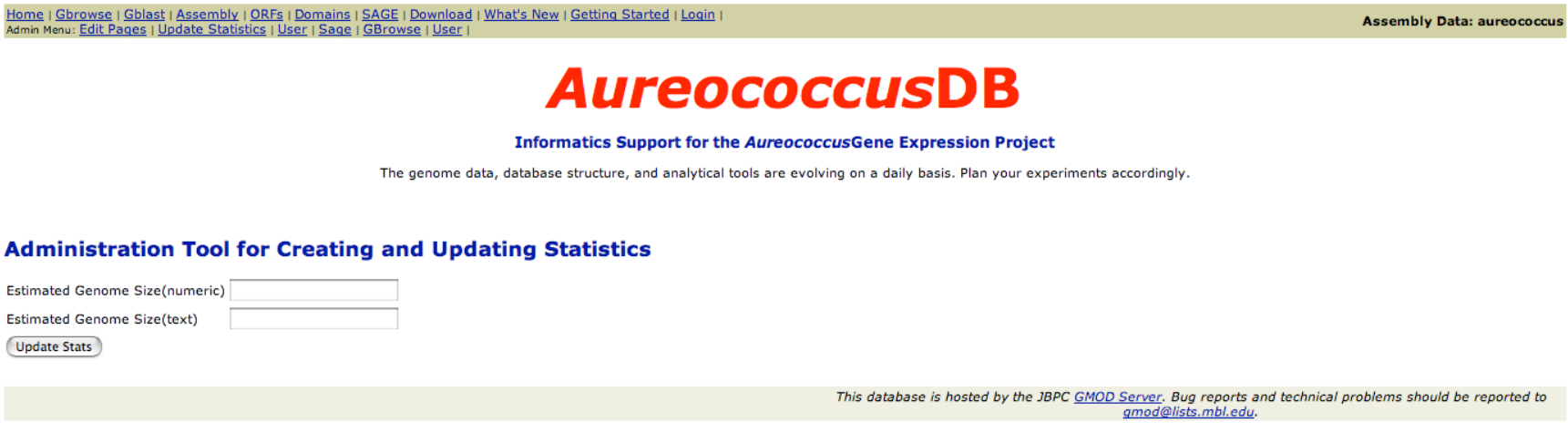
Once on the SAGE Admin page, select “Upload Results”:



Once on the Upload page, select your tags file and then UTR length. Once you hit submit, it can take several hours to run the analysis if there are lots of reference contigs (running overnight is an easy way to ensure it is complete):



Once the SAGE tags have been uploaded, they won't appear automatically on the website. The last step is to update the website using the "Update Statistics" tool:



You may additionally wish to blast all of the tags against GenBank to grab extra annotation information. Currently a GMOD administrator is needed to perform this – contact gmod@lists.mbl.edu