**Genome Pipeline Assembler User Manual**

For assembly of genomic data

Date Updated: (05/04/16)

Table of Contents

[**1. INTRODUCTION** 2](#_Toc450218010)

[**2. STEP 1 - REGISTRATION** 3](#_Toc450218011)

[**3. STEP 2 – LOGIN** 3](#_Toc450218012)

[**4. STEP 3 – VERIFY ACCOUNT** 4](#_Toc450218013)

[**5. STEP 4 – CREATE A JOB** 5](#_Toc450218014)

[**5.1 Step 1** 5](#_Toc450218015)

[**5.2 Step 2** 6](#_Toc450218016)

[**5.3 Step 3** 7](#_Toc450218017)

[**5.4 Step 4** 7](#_Toc450218018)

[**5.5 Step 5** 8](#_Toc450218019)

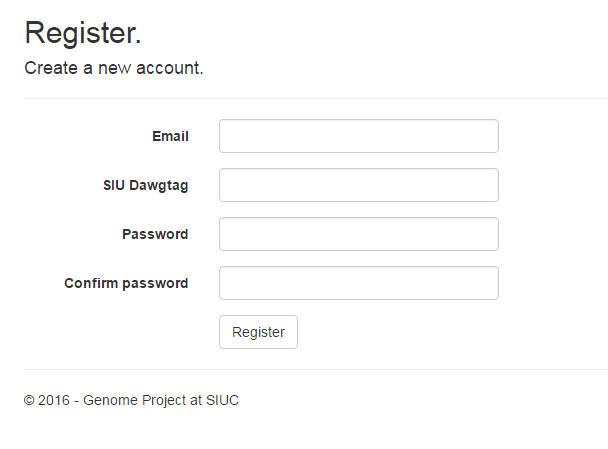
[**6. STEP 5 – VIEW A JOB** 8](#_Toc450218020)

# **1. INTRODUCTION**

This platform is used for the assembly and analysis of genomic data. The idea is to run data through as many assemblers at once to improve the data accuracy and efficiency of analysis for researchers.

# **2. STEP 1 - REGISTRATION**

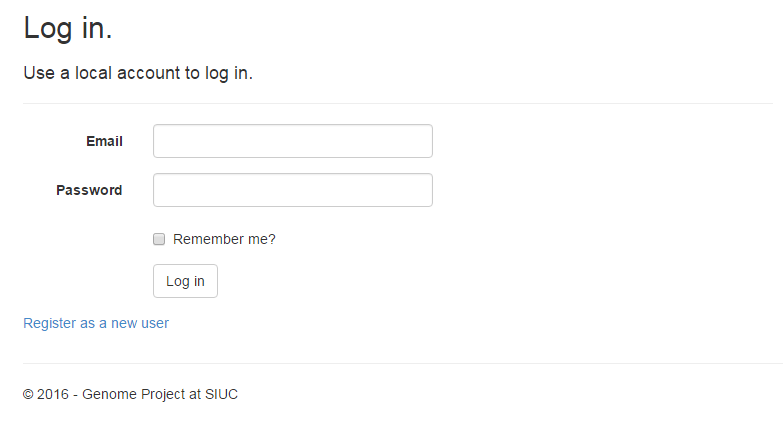
In order to access the system, you must first create an account. To access the registration page, simply go to the website or click registration. The following will be displayed to you.

* First enter in a valid email address such as [something@siu.edu](mailto:something@siu.edu).
* Then enter your dawg tag without the SIU.
* Finally, enter in a password. It must contain an uppercase letter, lowercase letter, symbol, and at least one number.

Once the registration has completed, you must now wait until your account has been activated by an admin. This may take from a day to a week to occur.

# **3. STEP 2 – LOGIN**

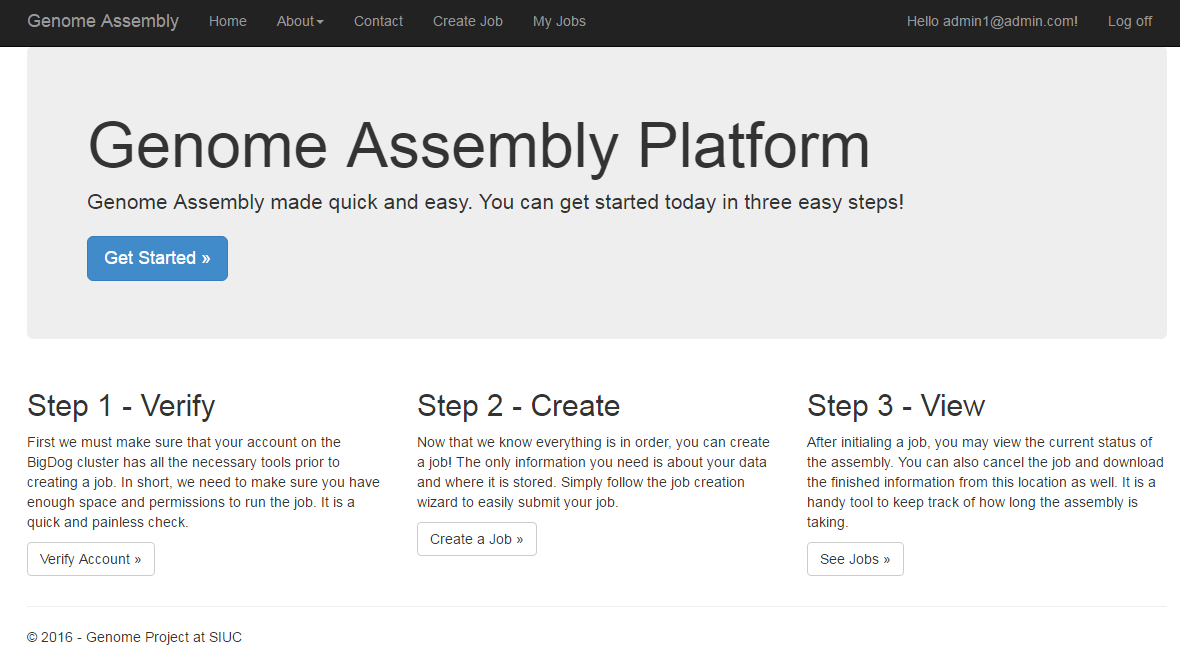
Once your account has been activated by an admin, you may now access the website. First, you must login. To login just visit the login page by typing in the website URL.

* To login, first type in the email you created in Step 1.
* Then type in the password you created in Step 2.
* If you would like your credentials to be remembered next time you visit the website, then select “Remember me”.
* Finally, click the login button.

If an error occurred, it will be displayed on the page. Otherwise, you will be logged in.

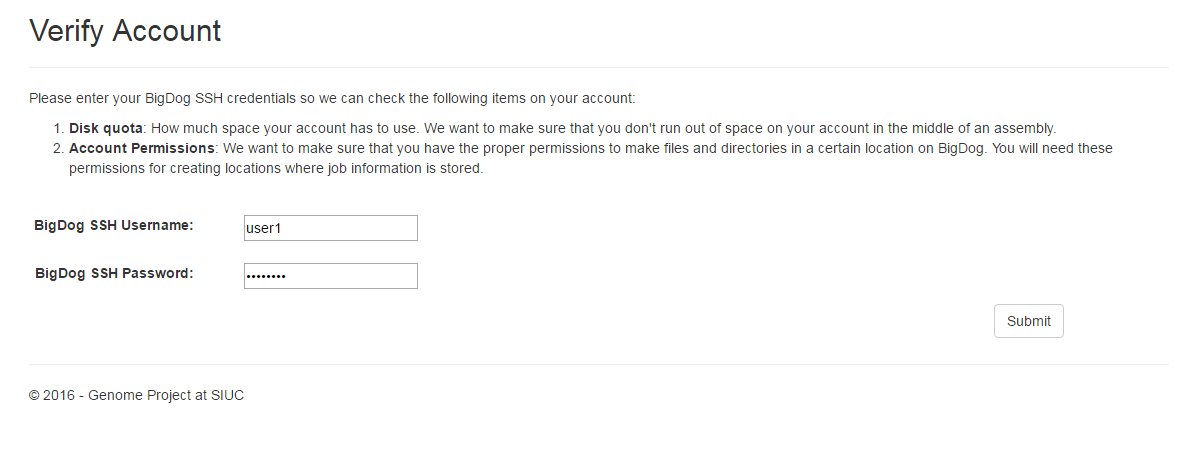
# **4. STEP 3 – VERIFY ACCOUNT**

After successfully logging in at Step 2, you will be greeted with the following page.



At this point, it is possible to do many things. But before being allowed to actually start a job, you must first verify your account. To do this, click the “Verify Account” button seen at the bottom left of the image above.

Once you have clicked the button, you will be taken to another page which will have the following on it:



At this point, we need to check that you have sufficient permission and enough space to do a genome assembly job. So you need to enter in your BigDog SSH username and password. If you do not have these things, then you need to contact the BigDog team.

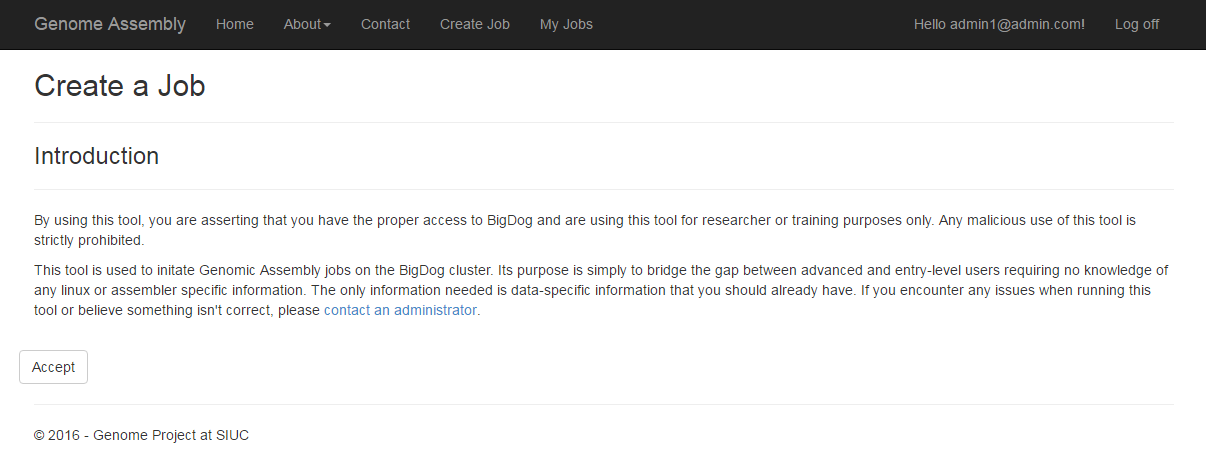
Once you have entered in the credentials, press the submit button. At this point, if your account is good to go, you will be told as such at the top of the page. Otherwise, you will be notified what needs to change and who to contact.

Once your account as successfully been verified, you will be able to finally create a job.

# **5. STEP 4 – CREATE A JOB**

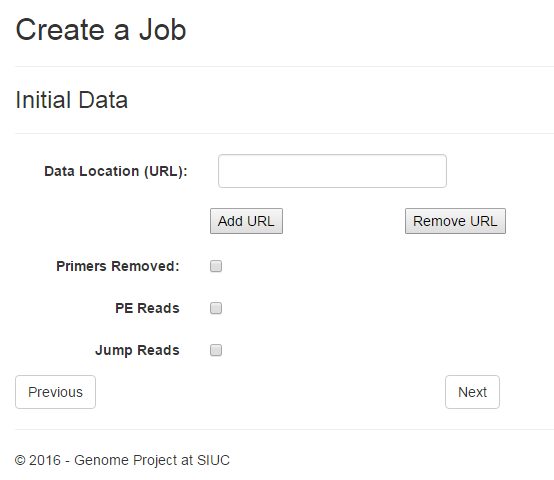
The following sections outline what is to be expected when creating a new job.

## **5.1 Step 1**



At this step, you simply need to agree to the conditions of using this product. Once you accept, you will be able to move to the next step.

## **5.2 Step 2**

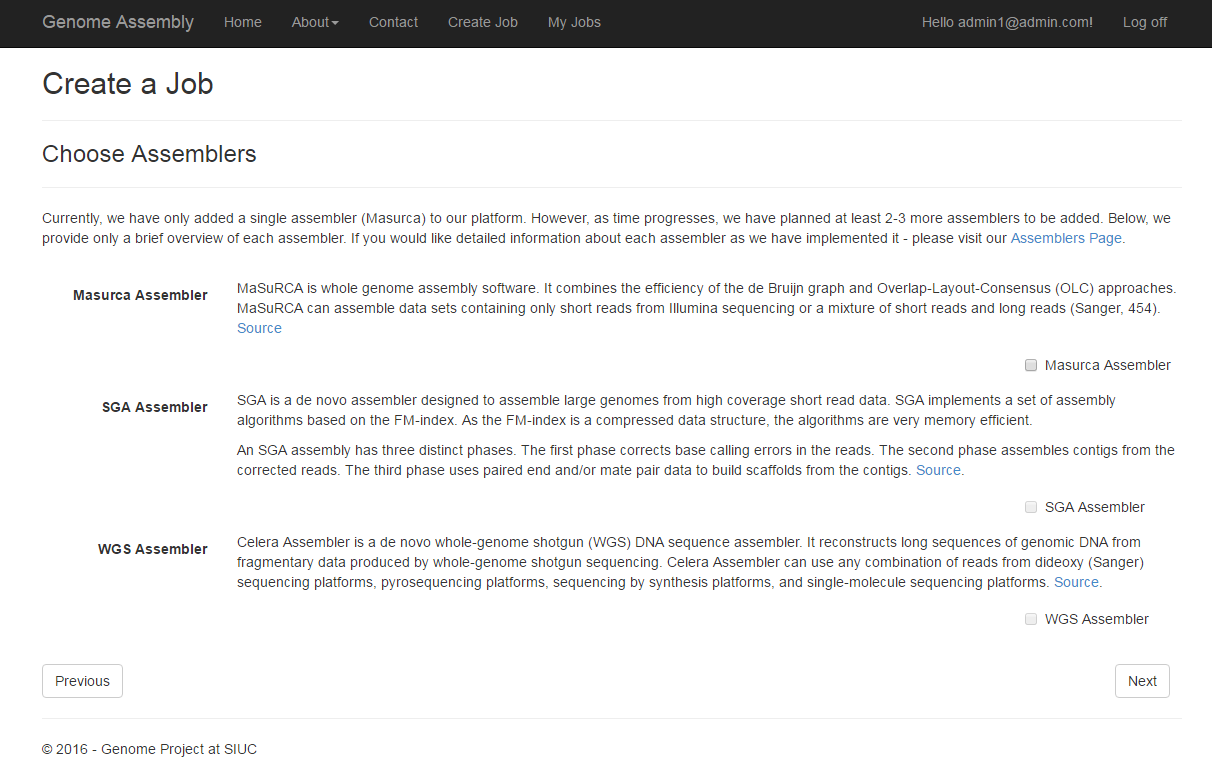


At this point, you will need to input data-specific parameters.

1. Data Location: Enter the URL to your data. If there are multiple URLs, then click the “add url” button and more boxes will appear. (Example: [www.siu.edu/mydata.fastq](http://www.siu.edu/mydata.fastq)).
2. Primers Removed: Check this box if the primers in your data have been removed. If they haven’t been removed, then they will be removed by the program.
3. PE Reads: If your data is paired-end data, then check this box and you will have to enter the PE Length.
4. PE Length: Length of the paired-end reads.
5. Jump Reads: If your data is jump read data, then check this box and you will have to enter the Jump Length.
6. Jump Length: Length of the jump reads.

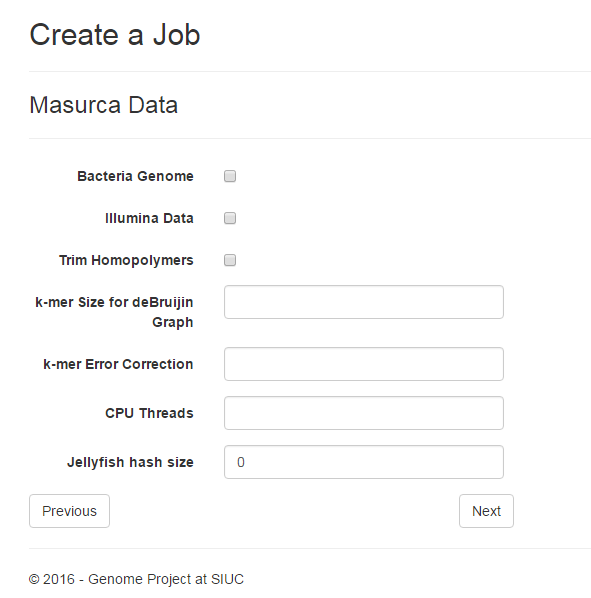
Once you are satisfied with the data entered, press next. If there any errors, they will be reported to you now and you will be unable to move to the next step if there are any.

## **5.3 Step 3**



At this point, you must select the assemblers that you wish your data to be run through. Currently there is only the Masurca assembler so that is the only option available. Select that and press Next.

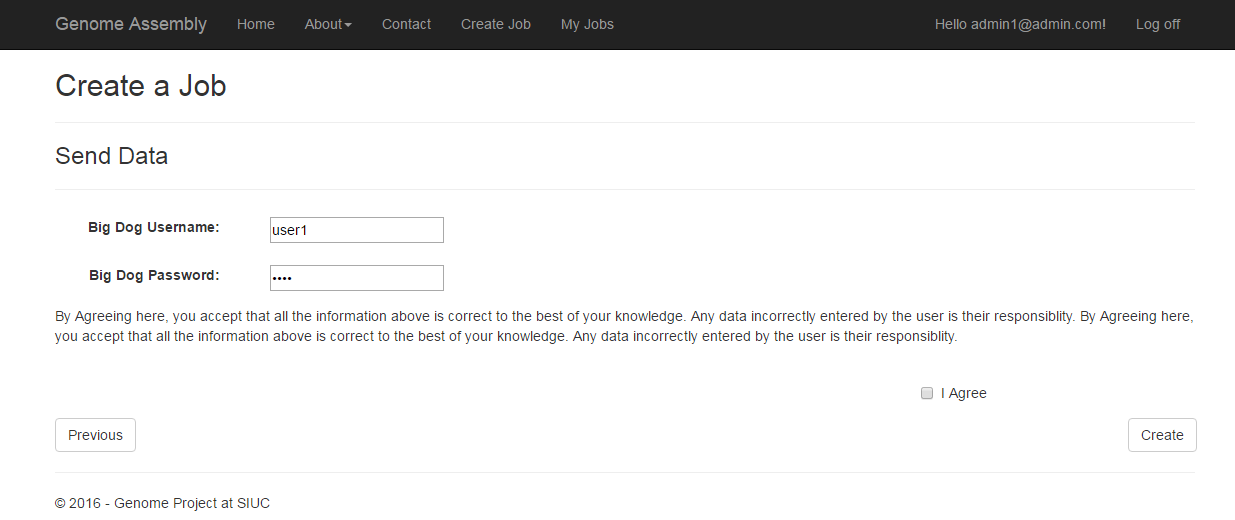
## **5.4 Step 4**

Now you must enter in the Masurca-specific parameters.

1. Bacteria Genome: If your data is a bacterial genome, select this option.
2. Illumina Data: If your data is Illumina data, select this option.
3. Trim Homopolymers: If you would like us to trim the homopolymers, then select this option.
4. K-mer Size for deBrujin Graph: The size of the k-mer value used in the production of the deBrujin graph.
5. CPU Threads: The number of CPU threads you wish your assembly to use. We default this to max.
6. Jellyfish hash size: This is usually a very large number used for analysis.

Once you are satisfied with the data entered, press next. If there any errors, they will be reported to you now and you will be unable to move to the next step if there are any.

## **5.5 Step 5**

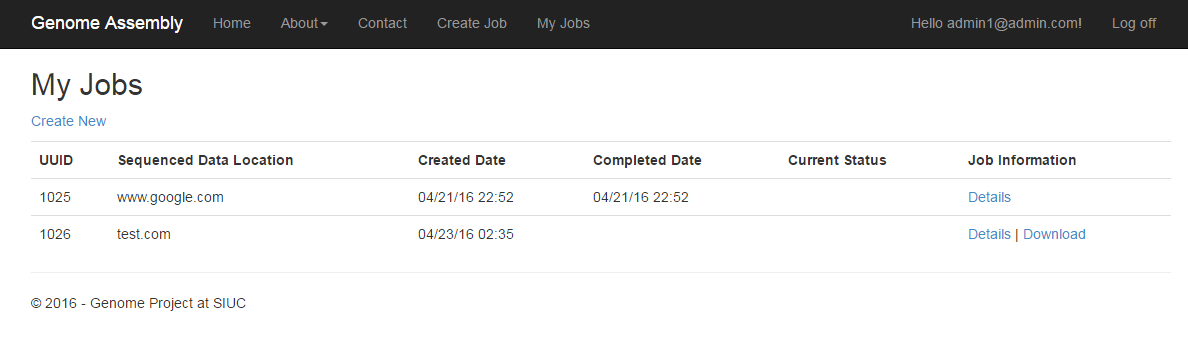


The final step requires that you enter in your BigDog SSH username and password to initiate the job. After doing that, press “I Agree” and hit create and the job will be created.

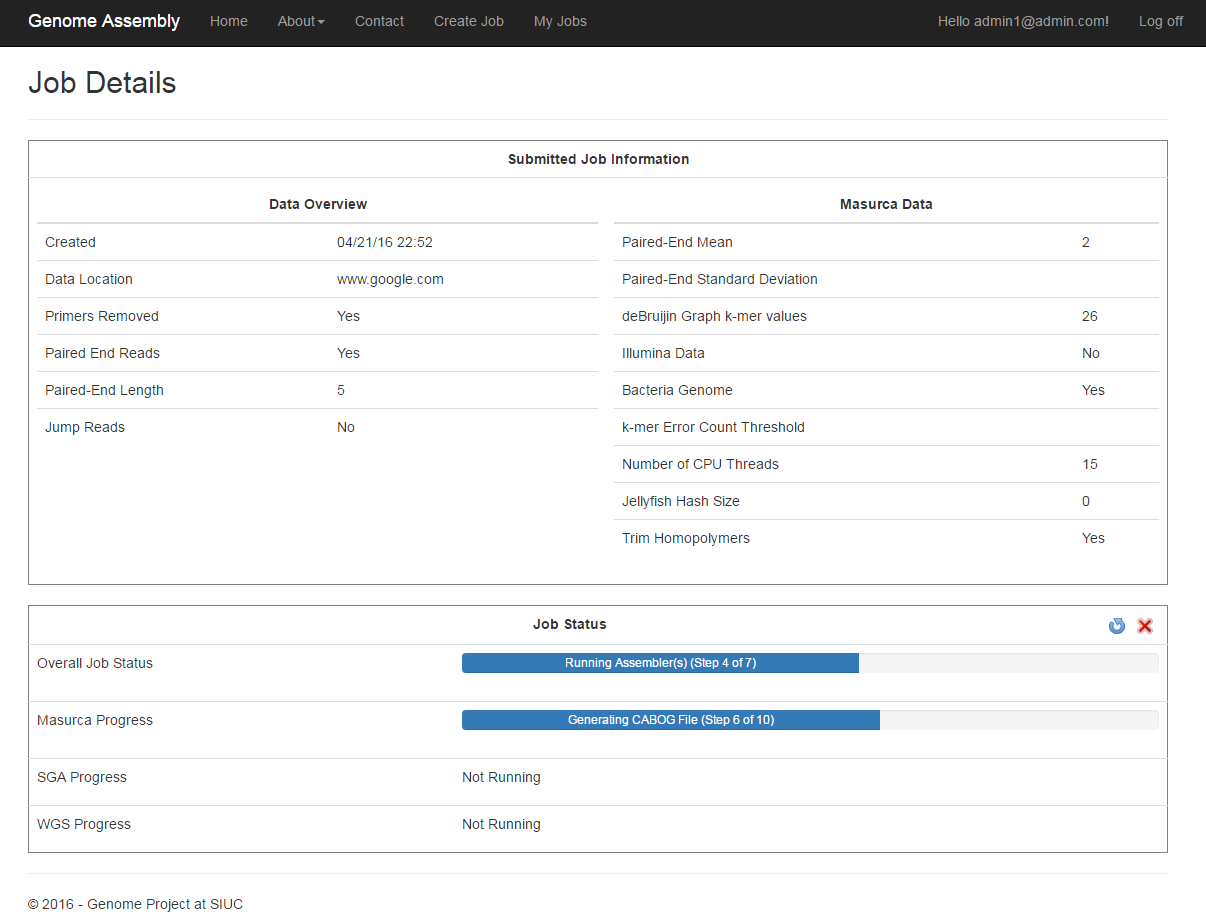
If there are any errors, they will be reported to you now and you will be unable to submit the job until they are fixed.

# **6. STEP 5 – VIEW A JOB**

In order to view a current job or cancel it, you need to view the details of the job. After submitting a job successfully, you will automatically be redirected to the details page of a job. But if you happen to lose that link, then you can simply view all of your jobs by clicking “My Jobs” in the tool bar at the top of the page. You will see the following page:



Except your jobs will be listed there if any. Then click the “Details” link under Job information. Once you have done that, you will see the following page:



At this point, you are able to see all of the information about the submitted job as well as where it is currently at in the analysis process. You’ll notice under the “Job Status” section at the bottom there is a blue refresh symbol and a red x off to the right. Those allow you to update the status of the job manually as well as cancel the job.

Simply clicking the refresh button will update the job automatically.

Clicking the red x button will require you to login with your SSH username and password to confirm that you really want to end the job.

Otherwise, all information will be reported back to you here as well as the download link.

Once the job is completed, the download link will appear both at the top of the page and the bottom of the page with any additional details about the job.

Since a finished job will only be available for a short time, it is important that the data is downloaded by clicking the link. It will then be downloaded to your computer immediately.