

**iPlant Tools and Services Workshop**

**2013 Maize Genetics Conference, March 14th - St. Charles, IL**

**Agenda**

|  |  |
| --- | --- |
| Time | Presentation |
| 10:00 AM | **Arrive/Sign – in** |
| 10:15 AM | **Welcome – Presenter/Participant Introductions** |
| 10:30 AM | **Overview of the iPlant Collaborative** |
| 10:45 AM | **Overview of the iPlant Discovery Environment (DE)** |
| 11:30 AM | **iPlant Data Store – Managing “Big Data”** |
| 12:00 PM | **Lunch Break** |
| 01:00 PM | **RNA-Seq in the Discovery Environment and Atmosphere** |
| 02:20 PM | **The DNA Subway as an Annotation Platform** |
| 02:30 PM | **Introduction to GBS and GWAS in the Discovery Environment** |

**Useful Information**

**Follow Along At:** **http://www.iplantc.org/tsw\_maize13**

**Account issues/information:** user.iplantcollaborative.org

**iPlant forums:** ask.iplantcollaborative.org

**Support issues:**  support@iplantcollaborative.org

**General questions:**  411@iplantcollaborative.org

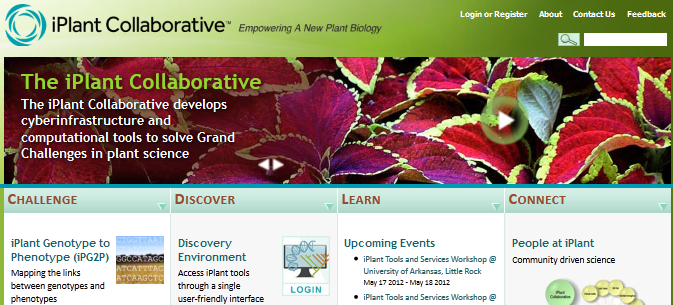
**iPlant staff @ this workshop**

Sheldon McKay mckays@cshl.edu

Josh Stein steinj@cshl.edu

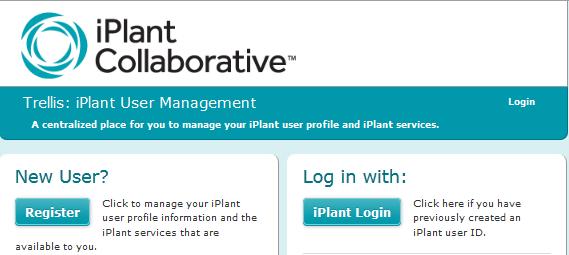
**iPlant Post-workshop survey**

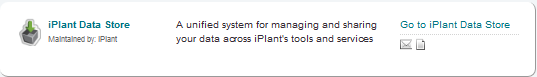
**https://www.surveymonkey.com/s/ToolsServices\_post**

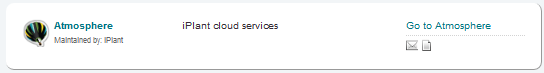
**Places you should know**

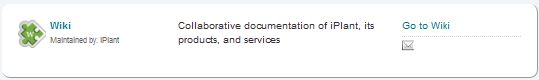
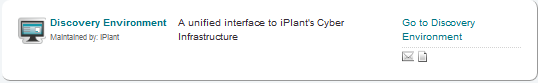
**iPlant Homepage:** Access to DE, Atmosphere, DNA Subway, and more.

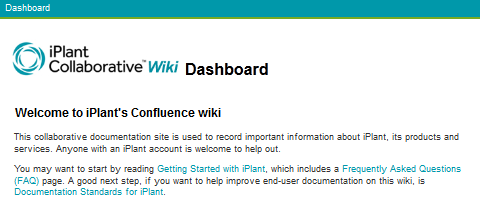
**www.iPlantCollaborative.org**



**iPlant User Management:** Account creation and management, access requests. Please login and check to see the following under **My Services** (Atmosphere, Discovery Environment, iPlant Data Store, Wiki). **user.iPlantCollaborative.org**







**iPlant Wiki:** Documentation, organization, everything iPlant

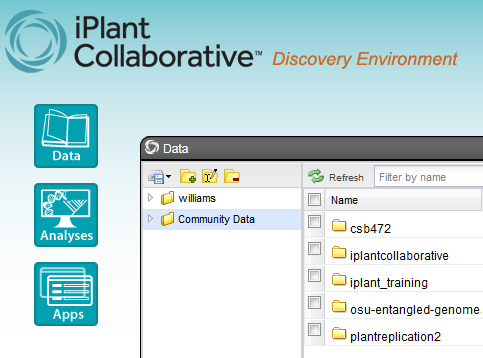
**pods.iPlantCollaborative.org/wiki**

**Tools and Services Workshop Guides/Tutorials**

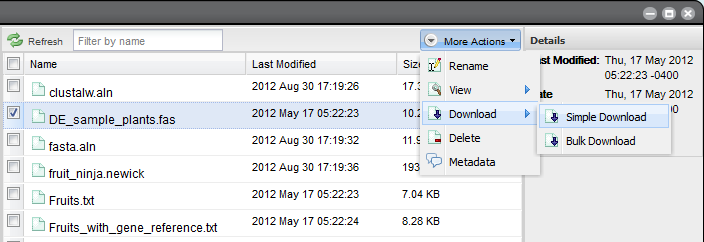
**Overview of the Discovery Environment (1.4.4)**

**Detailed Notes on the Wiki @: www.iplantcollaborative.org/de1**

**Goal:** Learn how to use the DE by creating a multiple sequence alignment



**1**

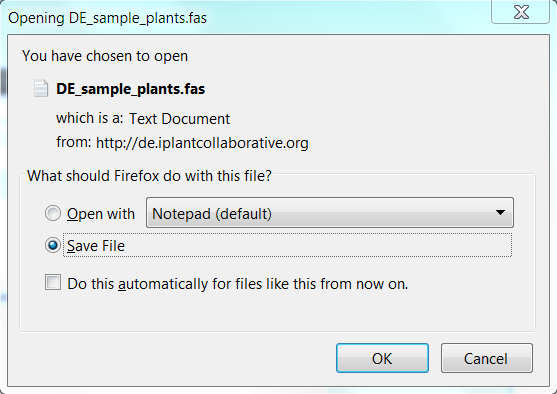


**4**

**3**

**Task 1:** Download a file from the community data folder.

1. Click **Data** from the DE workspace and locate the **Community Data** Folder.
2. Find the file **DE\_sample\_plants.fas** under the directory: *Community Data > iplant\_training > de\_walkthrough.*
3. Select the file **DE\_sample\_plants.fas** file by clicking on the box to the left.

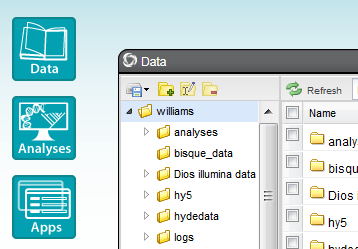
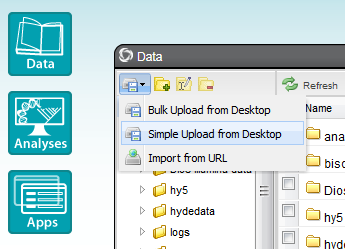


**5**

1. From the dropdown menu **More Actions** select **Download** > **Simple Download**.
2. This will start a download procedure to your local computer. Save the file wherever you would like.

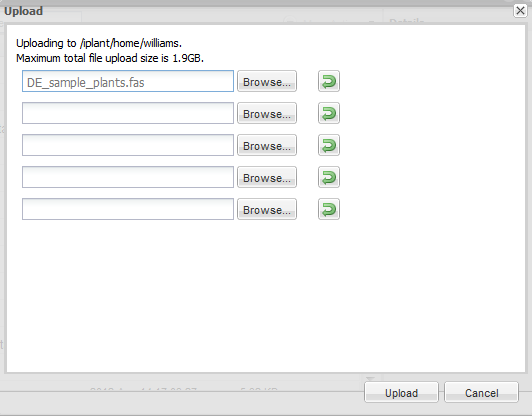
**Note:** For very large files you can also use **Bulk Download**. This runs via the Java application iDrop lite.

**Task 2:** Upload the file to your personal data store.



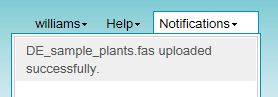
**6**

**7**

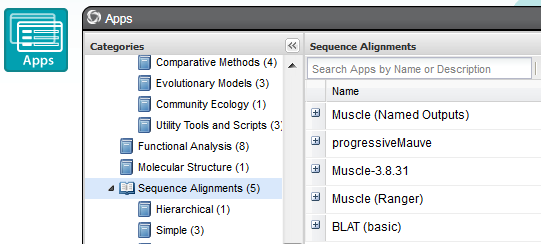
1. In the **Data** console, navigate to your home folder (make sure it is highlighted/selected).
2. Click on the **Import** icon (top left corner); a drop-down menu will appear from which you can select **Simple Upload from Desktop**.
3. Click **Browse** to navigate to the file you downloaded in step 5 (“DE\_sample\_plants.fas”). Click **Upload** to upload this file to your iPlant Data Store. You will receive a notification when your file is successfully uploaded

**8a**

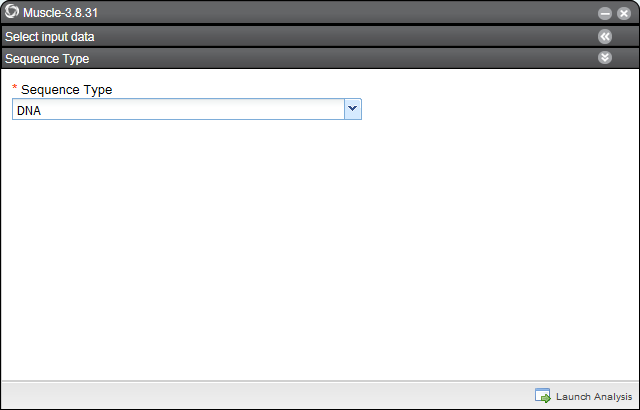
**8b**



**Task 3:** Find and select the multiple sequence aligner **MUSCLE** and run the app.

1. Click **Apps** from the DE workspace and select the aligner **MUSCLE-3.8.31** (Location: *Public Applications> Sequence Alignments>Simple)*. Click on the actual app name.

**9**

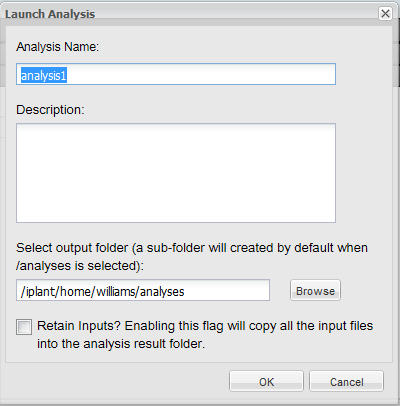
1. Click **Browse** and search your data store home folder. Select the **DE\_sample\_plants.fas** file. Then click **OK**.
2. Under ‘Sequence Type,’ select **DNA**, and then click **Launch Analysis.**
3. On the Launch Analysis console, name your job and select **OK.**

**11**



**10**

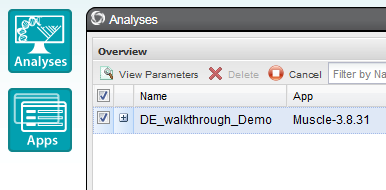
**11**



**12**

**Task 4:** Monitor the progress of your analysis and save parameters.

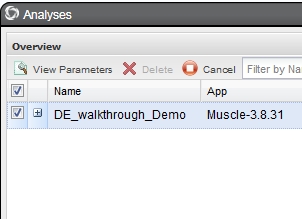
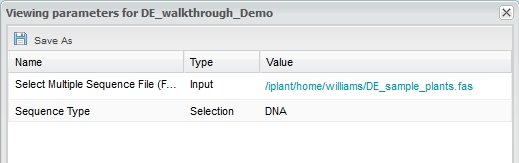
1. Click on **Analyses** from the DE workspace and monitor the status of your submitted job.



**13b**

**13a**

1. **Select** your analysis and click on the **View Parameters** icon. Save the parameters of your analysis to your home folder.

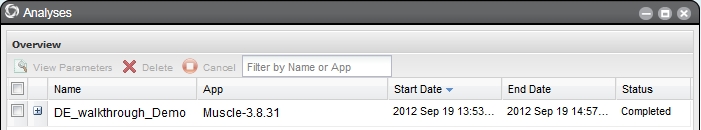


**14a**

**14b**

**Note:** When the status of your job is completed, results will appear in your **analyses** folder (or wherever you specified when you launched the job).

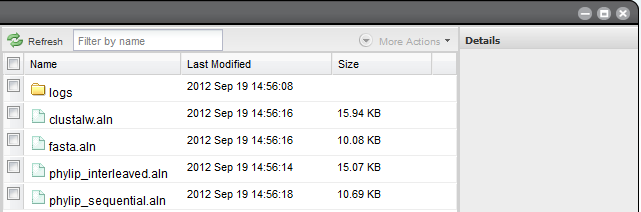
**Task 5**: View your results.



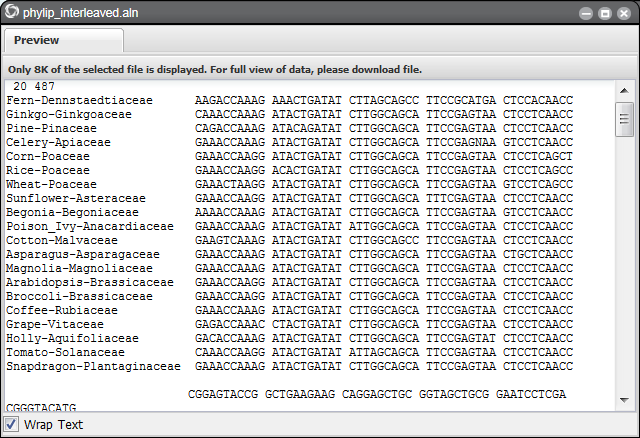
**15**

1. In the **Analysis** console, once your status appears as ‘Completed,’ click on the name of your analysis. (You could also navigate to your expected output folder from the Data console).
2. You should have a folder (named according to your job title) with the following outputs:

‘logs’ - (a folder of log files), ‘clustalw.aln,’ ‘fasta.aln,’ ‘phylip\_interleaved.aln,’ and ‘phylip\_sequential.aln.’



**16**

1. Click the **phylip\_interleaved.aln** file to view your aligned sequences. You can download these outputs or use them in further analyses.

**17**

**Overview of the iPlant Data Store**

**Detailed Notes on the Wiki @: www.iplantcollaborative.org/ds1**

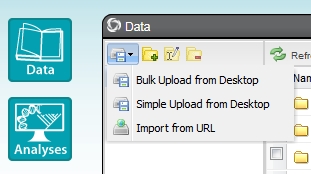
**Goal:** Import files into the data store, annotate them with metadata and share them with a colleague.

**Task 1**: Import a file into the DE from a URL

1. Follow your instructors’ direction to choose an ftp link for import. You can select any link you like - here’s one from Ensembl:

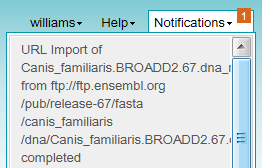
**ftp://ftp.ensembl.org/pub/release-67/fasta/canis\_familiaris/dna/**

**Canis\_familiaris.BROADD2.67.dna\_rm.chromosome.MT.fa.gz**



**4**

1. Click **Data** from the DE workspace.
2. Select your home directory from the listing.
3. Click on the **Import from URL** icon from the **Import** drop-down menu.
4. Paste the URL in the space provided. (*delete any unnecessary spaces before, after, or within the URL)* Hit **Enter** or click **Upload**.



1. Click on **Notifications** in the DE workspaceto monitor your notifications for the message that the upload is completed.

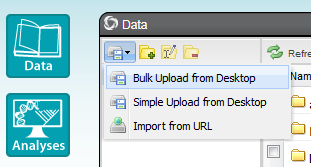
**6**

1. Click on **Data** from the DE workspaceand check your home folder to confirm that you see the imported file.

**Task 2:** Import a “large” file using iDrop lite in the DE

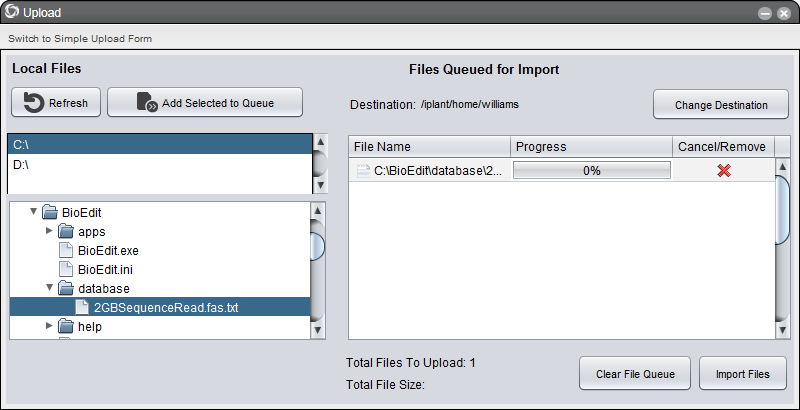
1. From the **Data** console, ensure you are somewhere in your home directory (where you have permission to write)

**9**

1. Choose **Bulk Upload from Desktop** from the **Import** icon.

**Note:** If prompted, give permission for Java (iDrop lite) to run. Selecting “Always trust content from this publisher” is recommended.

1. Under the **‘Local Files’** column select a file to upload from your local file system. (*For this demo, a small file (10-20MB) is best. The process for large (3-4GB) files is exactly the same. You can use the file we downloaded in the DE demo if you choose.*)
2. Click **Add Selected to Queue** icon to add the file to your import list (*you can upload multiple files if you wish*).
3. Click **Import Files** to start the upload. When the import is complete you will get a pop-up notification. You may then close iDrop lite.

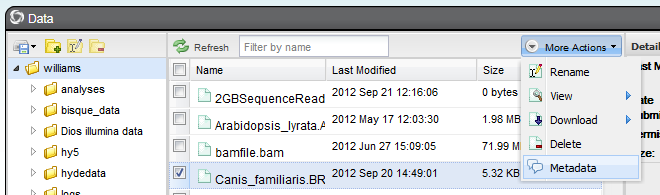


**12**

**11**

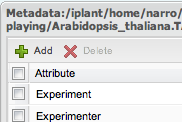
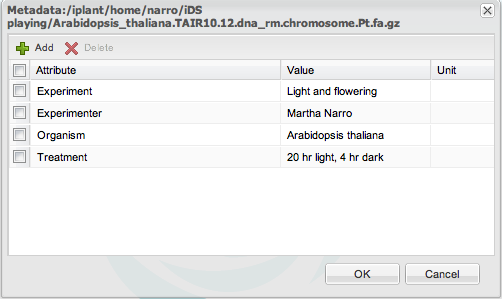
**Task 3:** Markup your files with metadata.

1. In the **Data** console, locate the file you imported from a URL in task 1.
2. Select the file and then click ‘More Actions’ to open a drop-down menu; click **Metadata**.



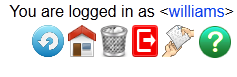
**14**

1. Click on the **Add** button and then enter information for the Attribute, Value, and Unit categories, then click **OK**. You can repeat steps 14-15 for file from task 2 as well.



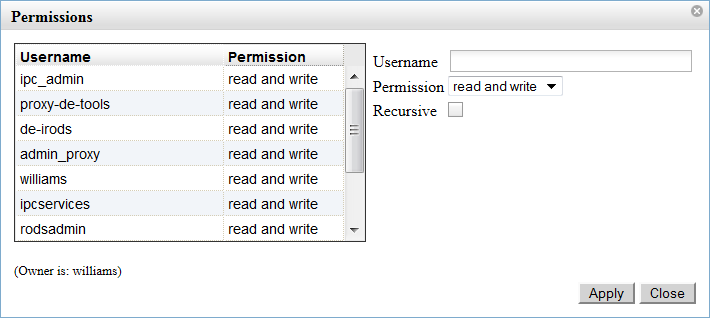
**15**

**Task 4:** Share your data with a colleague/ other user (DAVIS Web method)

1. Open a new web browser (or tab) and go to **data.iplantcollaborative.org** to access the DAVIS web interface to the data store.
2. Enter your iPlant credentials and click **Login**.
3. Click on the **Home** icon to navigate to your home folder.

**18**

1. Highlight one of the files you imported/uploaded to your home folder (*don’t actually click on the file name which will start a download*)
2. Click on to **Access Control** (right-side menu) and view the file permissions. You can only share files for which you have sufficient permissions. Close this window (You can also give permission for another iPlant user to access your file within the data store by adding their username here and configuring their access – see online documentation.) Click **Close** when done.



**20**

1. Click on **Share/Unshare** to share this file, and then with **Share** items selected, click **OK.** You now can copy the link you have generated, and share it with anyone (iPlant user or not). Copy the URL by left-clicking (apple: option – click) to copy URL.



**22**

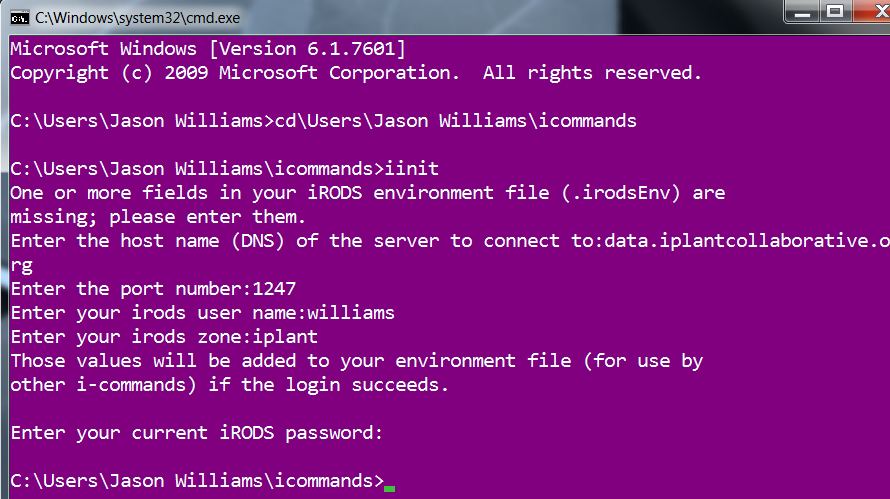
**Getting Started with iCommands**

**Detailed Notes on the Wiki @: www.iplantcollaborative.org/uic1**

iCommands is a command line utility provided by the iRODS project (www.irods.org); iRODS is the system upon which the iPlant Data Store is built. This quick guide will help beginners through the first (and often most difficult) step – transferring large datasets.

**Task 1**: Download and configure iCommands

1. Navigate to the link above (www.iplantcollaborative.org/uic1) and download the appropriate iCommands for your operating system (Windows, MacOX, Linux,)



1. Unzip (decompress) the files, making sure to note the file path (i.e. where you unzipped the files to) e.g. /usr/local/bin in a Linux system or C:\Users\user\_name> in a Windows system.
2. Open a command line terminal (terminal in Mac or Linux, enter “cmd” in the Windows start menu search bar.
3. Navigate to the extracted (decompressed) icommands folder
4. Run the **iinit** program (in Linux you may need to run this as: “./iinit”)
5. Configure your iCommands setup with the following parameters:

Host name (DNS): **data.iplantcollaborative.org**

Port: **1247**

iRods user name: **your\_iplant\_username**

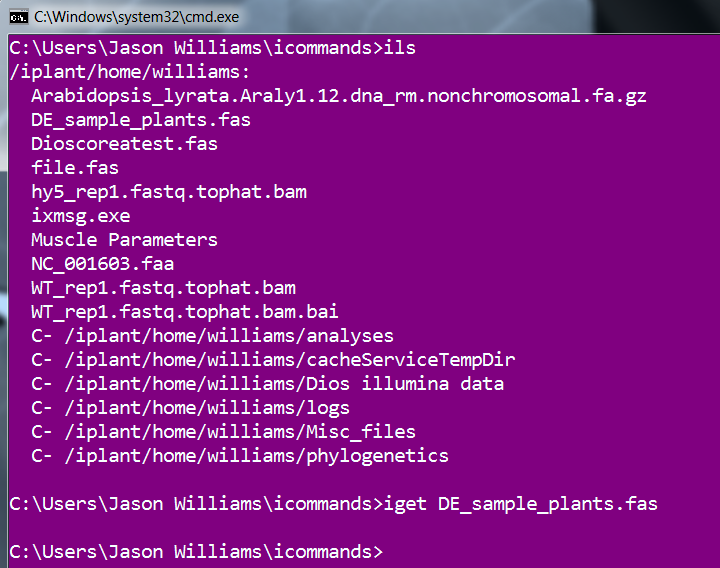
iRods zone: **iplant**

iRods password: **your\_iplant\_password**

1. If you are successful you will be returned to the shell prompt and will be able to log in in task 2.

**Task 2**: Use **iget** to download data from your iPlant Data Store

1. Enter the command “**ils**” to view the contents of your iPlant home directory. *Note: If you were not already logged in, it would be necessary to enter* ***iinit*** *to log in.*
2. Choose an appropriate file to download (depending on your local connection speed). *Note: you could use* ***icd*** *to navigate to a different directory within your iPlant Data Store (e.g. icd /iplant/home/your\_iplant\_username/your\_destination\_folder)*
3. Download the file by entering the following **iget** command: **iget your\_file’s\_name**



*Note: This will download the file into your icommands directory. In order to choose the source or destination directory the actual example command is:* iget [options] source path destination path, for example in **Windows** this could be:

**iget iplant/home/your\_username/your\_directory/your\_file C:\your\_local\folder**

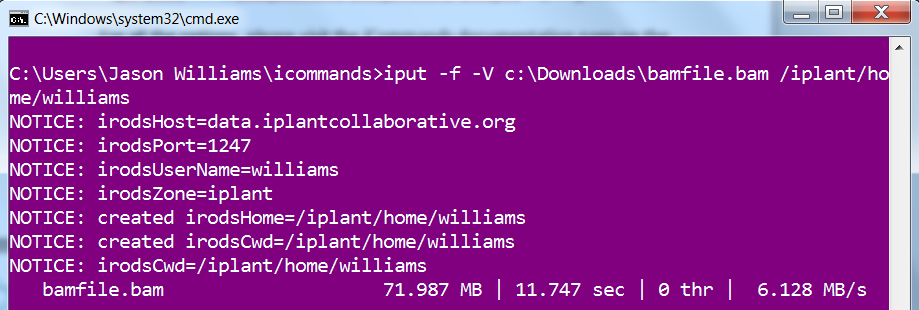
For all the options, please visit the iCommands documentation page on the iPlant wiki and at:

**https://www.irods.org/index.php/iget**

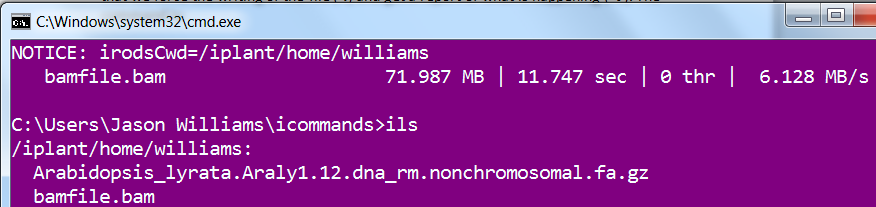
**Task 3**: Use **iput** to upload data from your local source to the iPlant data store

11. After determining the local file’s path, use **iput** with the following options **–f –V** (so that we force the writing of the file (**-f**) and get a report of what is happening (**-V**). The command is:

**iput –f –V your\_source\_directory/your\_source\_file /iplant/your\_username/your\_directory**



*Note:* *In this case, I moved a file called* ***bamfile.bam*** *from C:\Downloads to my iPlant home directory.*

12. Use **ils** to verify the upload of the file.

For all the options, please visit the iCommands documentation page on the iPlant wiki and at:

**https://www.irods.org/index.php/iput**

*Note: iCommands will automatically multithread large files, and is a robust way to transfer large data. Advanced users will be able to take advantage of iCommands abilities to design workflows for all their data management needs.*

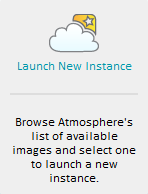
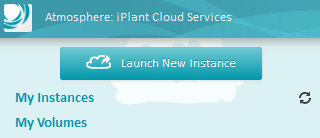
**iPlant Atmosphere Hands-on Lab**

**Detailed Notes on the Wiki @: www.iplantcollaborative.org/atm1**

**Goal:** Use Atmosphere to examine Illumina reads and share your desktop with another user.

**Task 1**: Launch, monitor, and connect to your Atmosphere instance

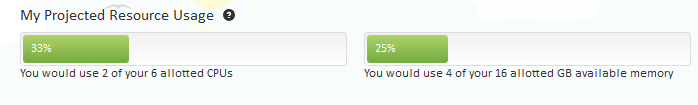
1. Navigate to Atmosphere at **https://atmo.iplantcollaborative.org/login/**
2. Login using your iPlant credentials.
3. Click **Launch New Instance** either on the navigation panel (left) **or** on the home screen.



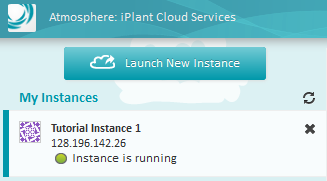
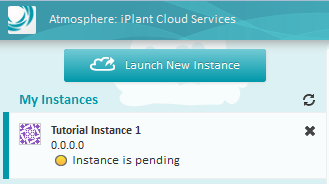
**1**

1. Under ‘Create a New Instance’ search for **NGS Viewers v3 8/20/2012**.
2. Give your instance a name (*Suggestion: ‘Tutorial Instance 1’).*
3. Select ‘Instance Size’ to be **m1.small (2 CPUs, 4 GB Memory)**; this is the default.
4. Check ‘My Projected Resource Usage’ to determine how many CPUs and how much memory you have. (*If you need additional resources, you must either terminate 1 or more running instances, or request additional resources*).

**7**



1. Click **Launch Instance**
2. Check the ‘My Instances’ heading in the navigation panel to view the status of your instance, e.g. pending, running, etc. (*If you no longer want an instance that is pending, click the ‘x’ icon to terminate that instance).*
3. In 10-15 minutes (depending on system load) your instance should display the status **running**. You may also receive an email notification when your instance is ready.

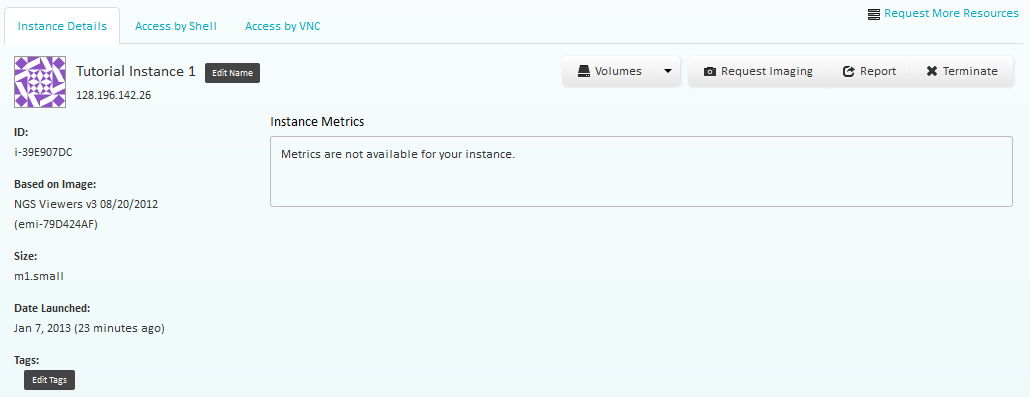
**10**

**9**

1. Under ‘My Instances’ click on the instance you created in step 8 to view details on your created instance.

**12**

**11**

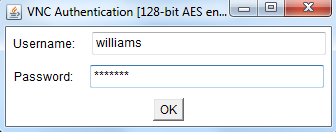
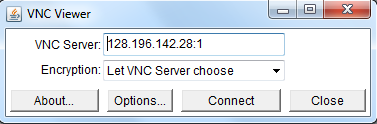


1. Click on the **VNC** tab to connect to this instance using VNC Viewer (*This requires Java is installed and running; alternatively you may download VNC viewer from the link)*.



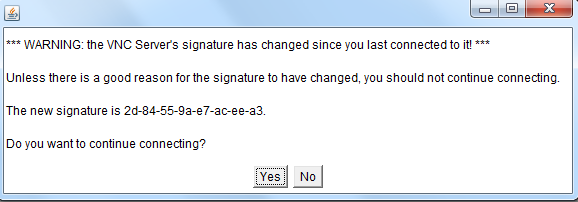
**13**

1. Give Java permission needed to run in your browser.
2. When the VNC Viewer window appears, click **Connect**.
3. The first time you connect, you may receive a warning that the server’s signature has changed. Click **Yes** to continue connecting.



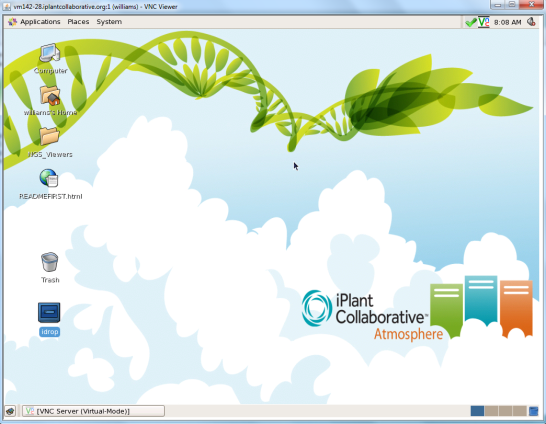
**14**

1. Enter your iPlant credentials and click **OK**.
2. Your instance may be locked when connected, so re-enter your iPlant password if prompted and click **Unlock**.



**15**

1. You are now viewing the desktop of your atmosphere instance.



**18**

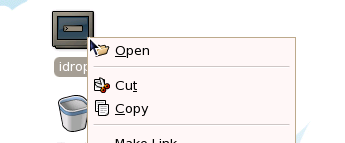
**17**

**16**

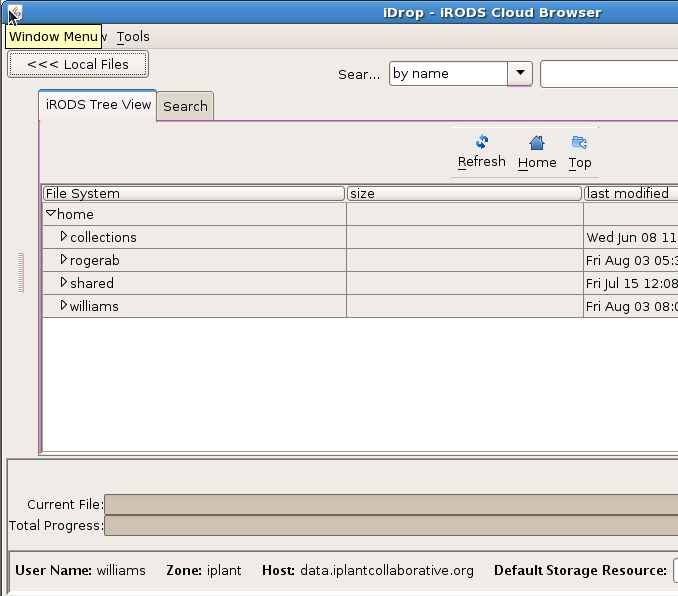
**18**

**17**

**Tip:** You may wish to adjust the size of your desktop while utilizing the VNC connection. To do so select **system** in the Linux toolbar and under **Preferences** select **screen resolution**. Select the resolution you wish to use, and the select **apply**.

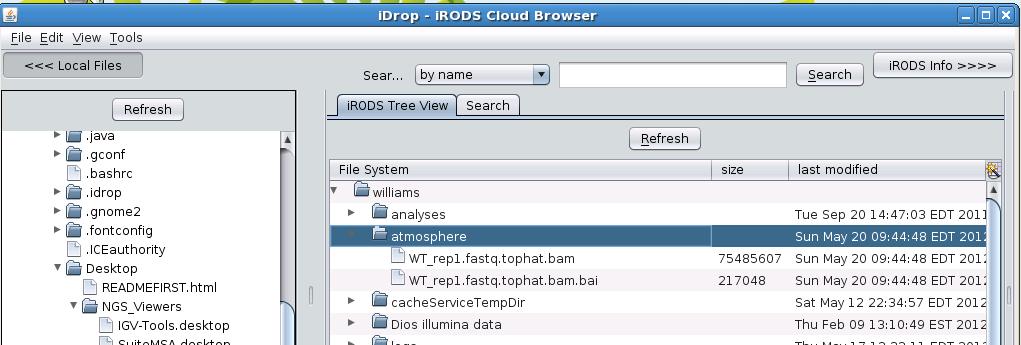
**Task 2**: Import data into your instance

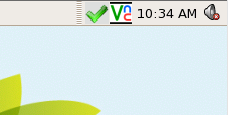
**19**

1. On the Desktop, right-click on **idrop** and select **Open.** Enter your iPlant username and password when prompted to ‘Please log in to your iDrop data grid.’
2. Click **<<<Local Files** so that you can see both your local and data store files in the iDrop console.
3. In ‘Local Files’ navigate to your ‘NGS Viewers’ folder on the desktop: *home > iplantusername > Desktop > NGS Viewers*

**20**

1. In ‘iRODS Tree View’ Open *shared > iplant\_training > atmosphere and* and drag the **WT\_rep1.fastq.tophat.bam** and **WT\_rep1.fastq.tophat.bam.bai** into your local ‘NGS Viewers’ folder. Select **Yes** as you move each file. Close iRODS



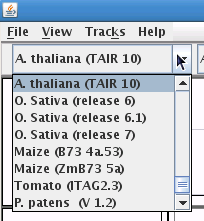


**21**

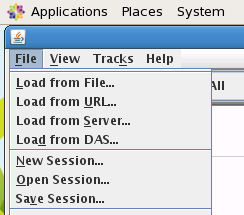
**22**

***Note:*** *When iRODS is running, you will see a green checkmark in the panel. Right click this icon to use iRods. Clicking the iRods icon on the desktop again will give an error message because iRODS is already running.*

**25**

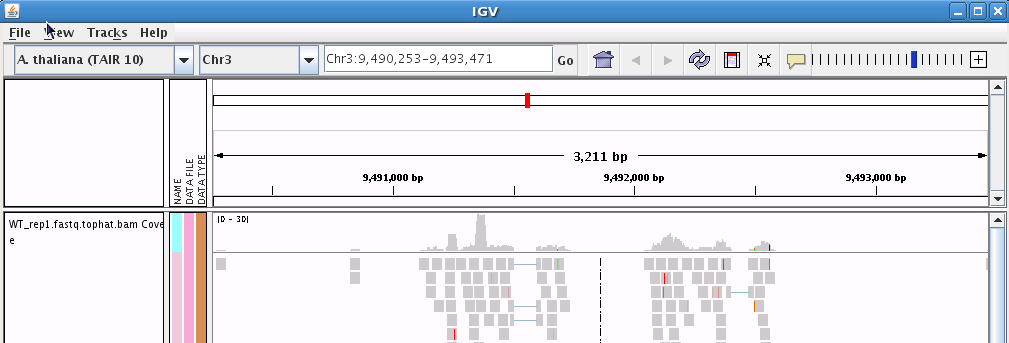
**Task 3**: Examine data with IGV (Integrated Genome Viewer)

1. Click on **NGS\_viewers** folder on the desktop.
2. Click on the **IGV 2.1.21** icon to start the program.
3. In the list of genomes drop-down box, select **A.thaliana (TAIR 10)**.

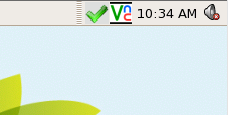


**26**

1. In IGV, under the ‘File’ menu click **Load from File**. Navigate to the ‘NGS\_Viewers’ folder (*Desktop > NGS\_Viewers*); load the **WT\_rep1.fastq.tophat.bam** file.
2. You can zoom in to see mapped reads. A good example can be found at: Chr3:9,490,253-9,493,471.

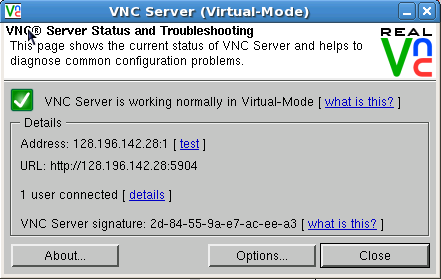


**27**

**Task 4**: Share your instance (desktop) with another user.

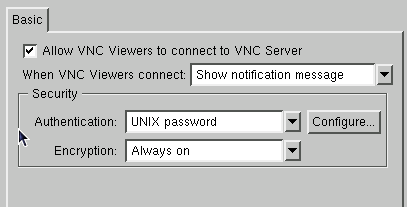
1. Click on the **VNC** icon on the desktop panel. This will open a VNC panel. **Copy down the URL (*e.g. http://128.XXX...)* to share later**.

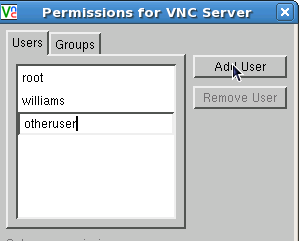
**28**



1. Click on **Options**. Under the security panel, click on **Configure**.
2. Under ‘Permissions for VNC Server,’ click **Add User**. Enter the iPlant usernames for anyone with whom you wish to share and then click **OK.**

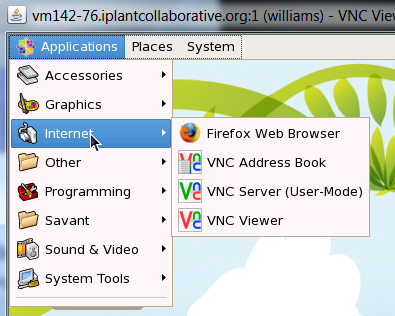
**Note:** *Only give access to your instance to someone who is authorized to use it. Never give anyone else your iPlant credentials.*

1. On the ‘Options’ panel click **Apply**.



**30**

**29**

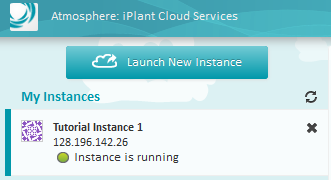
1. Give the URL copied in step 29 to your added users. They can now join your instance.

**Note**

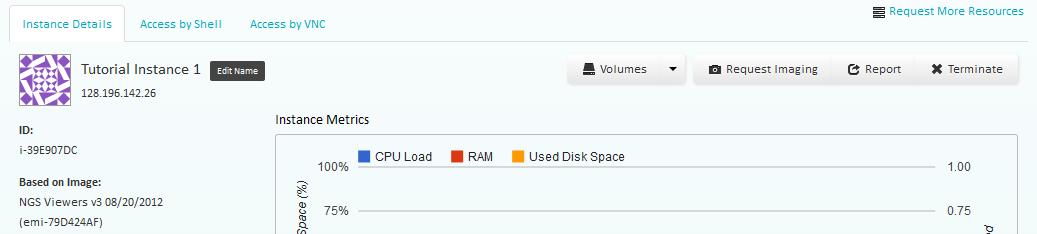
N**ote:** *If the VNC Icon disappears you can restart it under the Applications menu* (*Applications > Internet > VNC Server (User-Mode)*)

**Task 5**: Terminate Your Instance

1. Back on the Navigation panel; click the “**x**” button to terminate the instance. Alternatively you can click the **Terminate Instance** button on your instance’s status screen.



**334**



1. If you are sure you wish to terminate your instance click **OK**. (*Your instance status will change to ‘Instance is shutting-down’. Once the instance is terminated, your resource quota will be updated.)*

**Note:** Once you terminate your instance all data and modifications to that instance are permanently lost. You may transfer data back into your iPlant data store. Additionally, see the documentation for information on how to save data to an EBS volume or Imaging your instance.

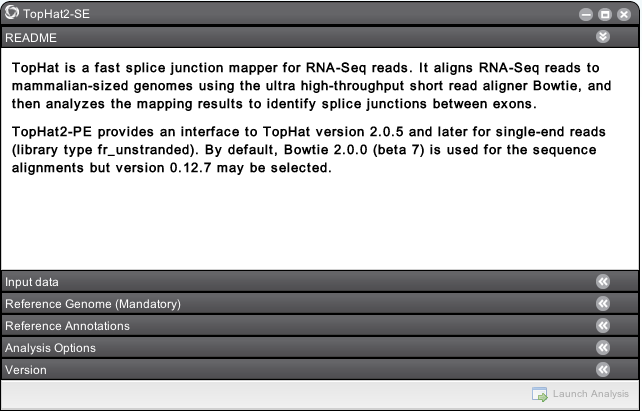
**Using the DE to Examine Differential Expression**

**Within an RNA-Seq Dataset**

**Detailed Notes on the Wiki @: www.iplantcollaborative.org/rs1**

**Goal:** Use RNA-Seq to compare expression levels for genes between WT and *hy5\-* samples and to identify new transcripts or isoforms.

**Task 1**: Align read data to the Arabidopsis genome using **TopHat**

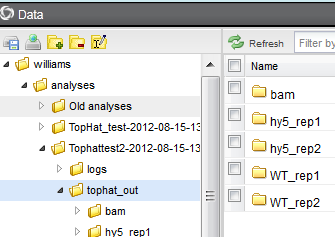
1.  Click on **Apps** from the DE workspace and select **TopHat2 – SE** under *Public Applications > NGS > Transcriptome Profiling*.

**2**

1. Under **Input data** use the “Add” button to browse and select each of four FASTQ files located under *Community Data > iplant\_training > intro\_rna-seq and select the files in the folder*  *01\_ input\_data* Alternatively you can click and drag these files from the data folder into the Input data window
2. Under **Ref. Genome (Mandatory)** for ‘Select a reference genome from the list’ select **Arabidopsis thaliana (Ensembl 14)** (*Note: This is equivalent to the TAIR 10 release)*

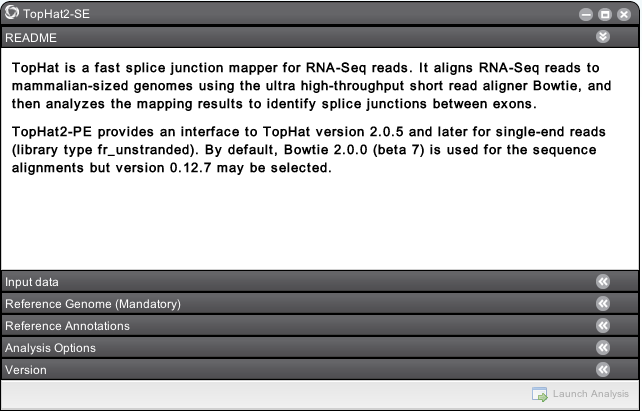
**3**

1. Under **Ref. Annotations** select **Arabidopsis thaliana (Ensembl 14)**; click **Launch Analysis.**
2. Name your job (e.g. TopHat), add a description if desired, and click **OK**.
3. Click on **Analyses** from the DE workspace to monitor the status of your job. You will also receive notifications.
4. When your job is completed click the job name in the **Analysis** console or navigate to the output in our **Data** directory. In the **tophat\_out** folder created you should verify you have created 5 folders; a ‘bam’ folder and one folder for each of the wild type/hy5 reads*.*



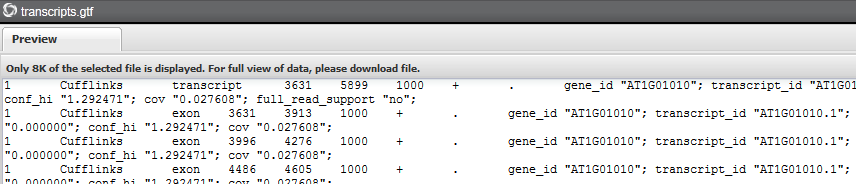
**7**

**Task 2**: Assemble transcripts using **Cufflinks**

1. Click on **Apps** from the DE workspace and select **Cufflinks2** under *Public Applications > NGS > Transcriptome Profiling.*
2. In the **Select Input Data** section, open a separate data window for the 'bam' folder from tophat\_out (above) and select and drag all four bam files into the input box. For convenience, a batch of TopHat bam files can be analyzed together but these files can also be processed concurrently in independent Cufflinks runs.

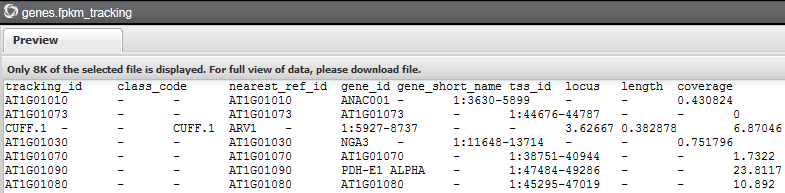
**9**

1. Under **Reference Sequence** under ‘Select Reference Genome Annotation’ we select the same genome build that we used in the TopHat assembly: **Arabidopsis thaliana (Ensembl 14)**
2. Click **Launch Analysis**. Name your job (e.g. Cufflinks), add a description if desired, and click **OK**.
3. Click on **Analyses** from the DE workspace to monitor the status of your job. You will also receive notifications. When your job is completed click the job name in the **Analysis** window or navigate to the output in our **Data** directory. In the cufflinksoutput folder that is created you should find folders for each replicate as well as a folder called **gtf**.
4. In the other folders created by Cufflinks (e.g. hy5\_rep1) you should find GTF and FPKM files. Click on **transcripts**.**gtf** to view annotated transcripts with their TAIR 10 release annotations.



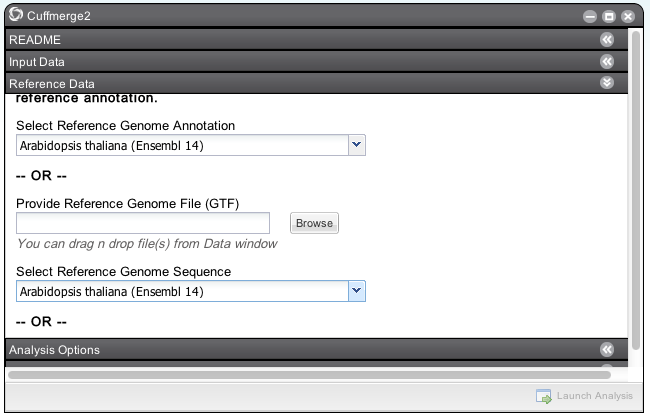
**13**

1. In the same folder (hy5\_rep1) click on the **genes.fpkm\_tracking** file to preview coverage expressed in fragments per kilosbase of exon per million mapped reads.



**14**

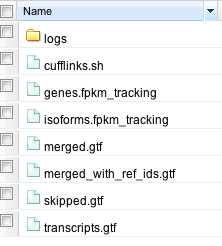
**Task 3**: Merge all assembled transcripts into a single transcriptome annotation file with **Cuffmerge**

1. Click on **Apps** from the DE workspace and select **Cuffmerge2** under *Public Applications > NGS > Transcriptome Profiling.*

**16**

1. For **Input Data (Mandatory)** under ‘Enter one or more input SAM/BAM files’, use the four files in the **gtf** folder created in Task 2 (step 12). If Task 2 has not completed, you can skip ahead by selecting pre-run data in the folder under: Community Data/iplant\_training/intro\_rna-seq/03\_cufflinks/gtf
2. Under **Reference Data** under ‘Select Reference Genome Annotation’ we select the same genome build that we used in the TopHat assembly: **Arabidopsis thaliana (Ensembl 14)**
3. Click **Launch Analysis**. Name your job (e.g. **Cuffmerge**), add a description if desired, and click **OK**.

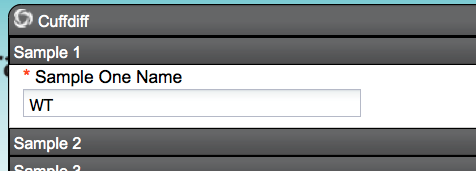
**19**



1. Click on **Analyses** from the DE workspace to monitor the status of your job. You will also receive notifications. When your job is completed click the job name in the **Analysis** console or navigate to the output in our **Data** directory. In the folder that is created you should find 7 files.

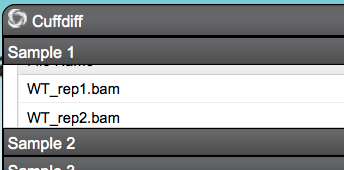
**Task 4**: Compare expression using **CuffDiff**

1. Click on **Apps** from the DE workspace and select **CuffDiff2** under *Public Applications > NGS > Transcriptome Profiling.*



**21**

**21**

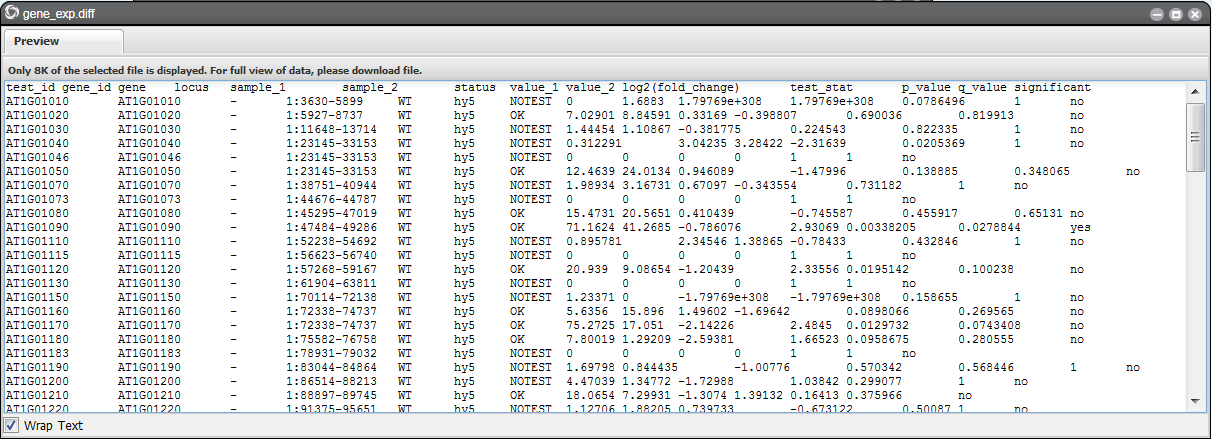
1. Under **Sample 1** enter **WT** for ‘Sample One Name.’ Then add (or drag) bam files from the two wild type replicates in **Tophat\_out** from the Tophat run (Task 1 Step 7).
2. Under **Sample 2** enter hy5 for ‘Sample two Name.’ Then add (or drag) bam files from the two HY5 replicates in **Tophat\_out** from the Tophat run (Task 1 Step 7).

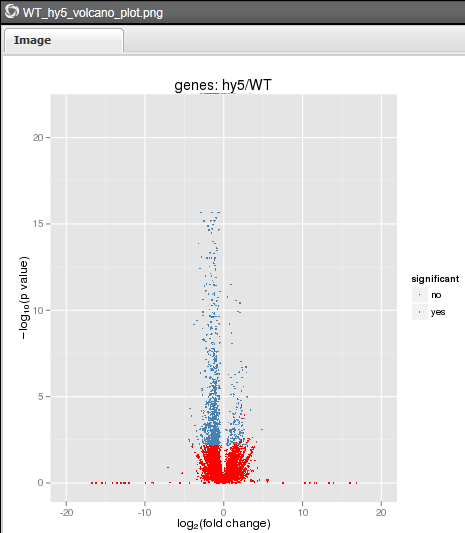
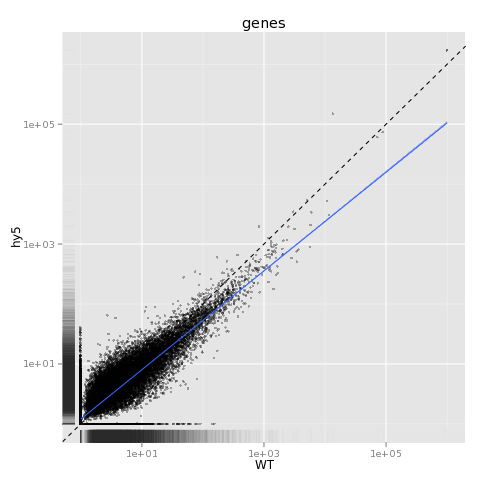
**22**

1. Under **Reference Data** under ‘Custom annotation File’ browse for the **merged\_with\_ref\_ids.gtf** created from Task 3 (step 19).
2. Click **Launch Analysis**. Name your job (e.g. **CuffDiff**), add a description if desired, and click **OK**.
3. Click on **Analyses** from the DE workspace to monitor the status of your job. You will also receive notifications. When the job is completed, click on the job name to navigate to the job output.

*In the* ***cuffdiff\_out*** *folder you will see a number of outputs which are described in documentation (www.iplantcollaborative/rs1) including* ***gene\_exp.diff*** *which compares expression between the two samples (****WT*** *and* ***hy5)****. The* ***graphs*** *folder contains a few automatically generated plots using* ***cummeRbund****, part of the cufflinks RNA-Seq workflow.*

**Example results and plots**



******

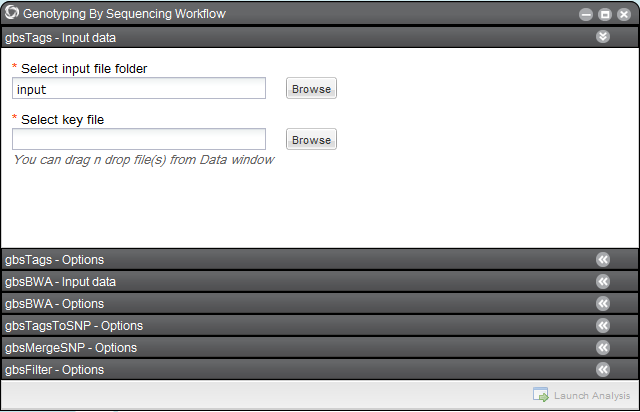
**Getting started with GBS and GWAS in the DE**

**Original TASSEL documentation: www.maizegenetics.net/tassel**

**Goal:** Explore the Discovery Environment implementation of the TASSEL package for GBS (Genotyping by Sequencing) and GWAS(Genome Wide Association Study)

**Task 1:** Start a GBS project using a supported genome

In this workflow, your input of gzipped Qseq files and a Barcode key file are used to an example rice dataset. An output tagCount file will be created for each Qseq input.

1. Click **Apps** from the DE workspace and locate the **Genotyping By Sequencing Workflow** app (Location: *Public Applications> QTL and GWAS>Genotyping By Sequencing Workflow)*.
2. Under ‘gbsTags – Input data’ for **Select input file folder** select the qseq folder located in community data directory: *Community Data > iplant\_training > gbs\_workflow > input > qseq*.

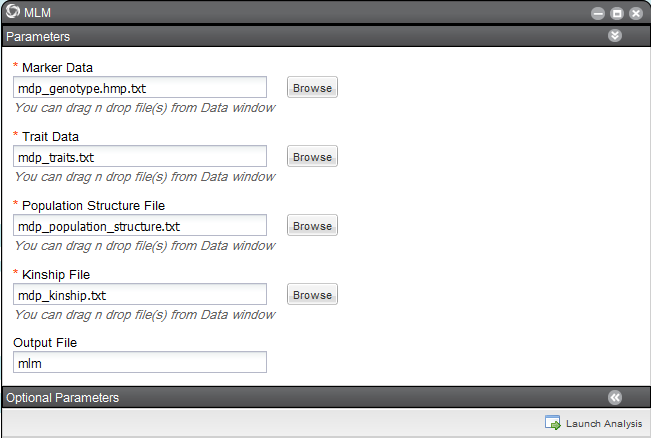
**2**

1. For **Select key file** select the **61VBPAAXX\_key.txt** file located in community data directory: *Community Data > iplant\_training > gbs\_workflow > input.*
2. Under ‘gbsTags – Options’ leave the **Start Chromosome** as **1**, change the **End Chromosome** to **2** (*This will save analysis time for the demo)*.
3. Under ‘gbs-BWA – Input data’ for **Select reference genome** select **Oryza\_sativa (japonica)**.
4. There are other options that can be adjusted, but in this instance we will leave all preset defaults. Click **Launch Analysis**, name your job, and select. OK.

You can use the **GBS Workflow with user genome** app if you want to supply your own genome for alignment, or use **UNEAK** app if your genome is not assembled.

**Task 2:** Start a GWAS project using genotype and phenotype data

With genotype data in Hapmap format and a table of trait values you can use TASSEL’s mixed linear model to look for association.

1. Click **Apps** from the DE workspace and locate the **MLM** app (Location: *Public Applications> QTL and GWAS>MLM)*.

**3**

1. Click on **Data** from the DE workspace, and navigate to *Community Data > iplant\_training>mlm\_gwas>input*
2. **Drag and drop** the following files for the following fields:
   1. ‘Marker Data’

**mdp\_genotype.hmp.txt**

* 1. ‘Trait Data’

**mdp\_traits.txt**

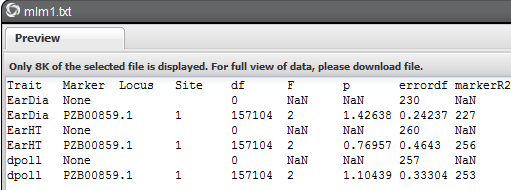
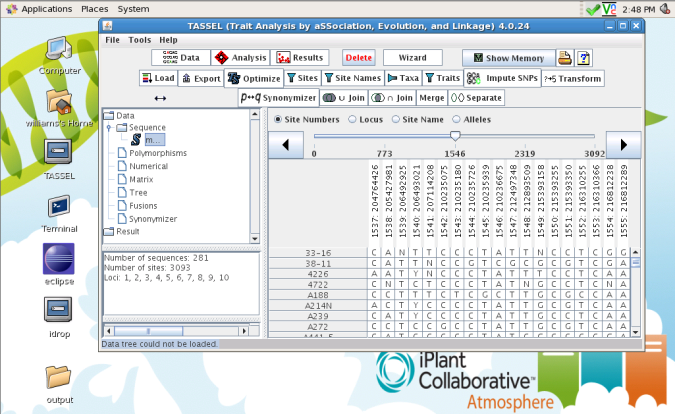
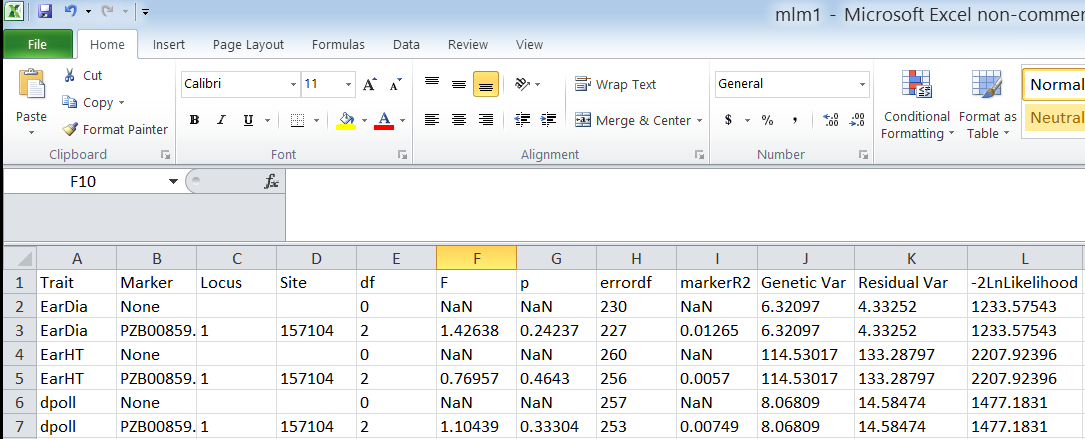
* 1. ‘Population Structure File’

**mdp\_population\_structure.txt**

* 1. ‘Kinship File’

**Mdp\_kinship.txt**

1. **Click** **Launch Analysis**, name your job and monitor its progress from the **Analysis** panel.

**Note:** You can download your results from the DE and further manipulate them (for example in excel). You can use the **MLM Workflow** app if you don’t have kinship and population structure data. You can also do further manipulations and visuzlazations with TASSEL using iPlant Amosphere (e.g. Population Genetics image – emi-33572B37)