

Pre-Workshop Packet

Training materials to move you from Data to Discovery



Get Science Done



Reproducibly



Productively



Introduction to iPlant	3
What is Cyberinfrastructure?.....	3
What to do with this packet..	3
How long will this take? Do I have to complete every exercise?	3
What if I get stuck or need help?	3
1. Setup – Sign up for your iPlant account and prepare your computer	4
2. Data upload - Import your data using iDrop	5
3. Sample Analysis - Use DE to examine illumina sequence with FastQC	6
4. Launch a virtual machine – Connecting to Atmosphere.....	7
5. Additional Exercises and notes for Livestream attendees.....	9
Resources you should know about.....	10
How to Acknowledge iPlant	11
Funding.....	11

iPlant's Vision - The iPlant Collaborative enables life science researchers and educators to use and extend iPlant's foundational cyberinfrastructure to understand and ultimately predict the complexity of biological systems and their dynamic nature under various environmental conditions, thus stimulating the transformation of how the life sciences community does research.



Introduction to iPlant

The iPlant Collaborative was funded in 2008 by the National Science Foundation to develop cyberinfrastructure (software, high-performance computing, data management, and people) needed to support data-intensive biology. Since then, we have developed a number of resources for phylogenetics and genotype-to-phenotype science – resources you can use to get more science done and ask bigger questions while saving you time, effort, and funding.

What is Cyberinfrastructure?

Cyberinfrastructure (CI) is data storage, software, high-performance computing, and people – organized into systems that solve problems of size and scope that would not otherwise be solvable. Some of these components you likely have already – your personal CI. However, iPlant CI can help fill in gaps where you don't have access to resources you need, and expand your ability to address larger questions that involve bigger compute and a greater need to share data with collaborators and the community.

What to do with this packet...

These exercises are designed to get you to “touch” some of the important pieces of the cyberinfrastructure. This way, you can spend the in-person (or online) training time more effectively, troubleshoot, and ask better questions.

This packet **will not** address much of the rationale or science relating to the topics covered. You will however learn how to upload data so you can more quickly move to working with your own research.

How long will this take? Do I have to complete every exercise?

We think this will take 30-60 minutes to complete. You don't have to complete every exercise, but at a minimum, you should get your iPlant account. We also highly recommend you complete the iDrop exercise so that you can start working with your own data as soon as possible.

What if I get stuck or need help?

You can 1) post or search for an answer to your question/problem to ask.iplantcollaborative.org, 2) email support@iplantcollaborative.org, or if scheduled 3) attend any announced virtual office hour that has been scheduled prior to the workshop.

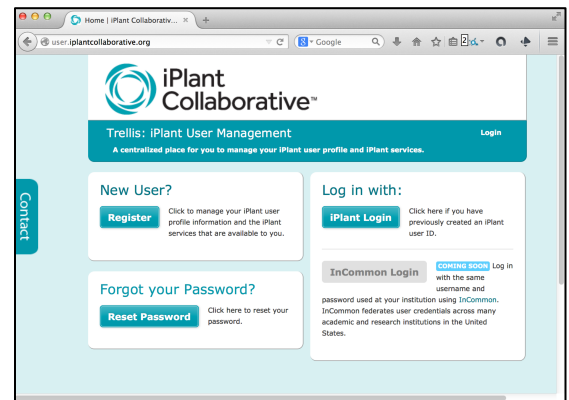
✂ 1. Setup – Sign up for your iPlant account and prepare your computer

iPlant accounts are free and anyone (students, researchers, industry professionals, non-U.S. users, etc.) can sign up for an account. It is recommended you use an institutional email (e.g. .edu, .org, etc.) to sign up for your account; this is a **requirement for access to the Atmosphere cloud computing services**.

Sign up for iPlant account and request access to Atmosphere

If you already have an account – skip to step 3 to register for Atmosphere

1. Visit the iPlant user portal at <http://user.iplantcollaborative.org/>
Click 'Register' to begin creating your account.
2. You will automatically receive a confirmation email, and you should follow the instructions to confirm your account as soon as possible.
3. Once you have confirmed your account, return to the user dashboard at <https://user.iplantcollaborative.org/dashboard/> check to see if **Atmosphere** is listed under the heading 'My Services.' If Atmosphere is listed, you don't need to do anything further. If Atmosphere is listed under 'Available Services', click the '**Request Access**' link. For justification, please enter "Attending a workshop."



Computer requirements

Please bring your own Wi-Fi enabled laptop to the workshop. Make sure your laptop has the following:

- **VNC Viewer:** Download the DMG (for MAC) or exe (for PC): <http://www.realvnc.com/download/viewer/>
- **Java:** Please have the latest version of JAVA (www.java.com) installed and enabled
- **Browser:** Please have an up-to-date web browser (Recommended Firefox or Safari)

Also, it is recommended that you have administrative rights to your computer so that you may install or software or adjust settings as needed.



2. Data upload - Import your data using iDrop

iDrop – The easiest way to get your data uploaded to iPlant

There are several methods to get data into the iPlant data store. In this packet, we will only cover iDrop – a user-friendly standalone application that can serve the majority of use cases for the majority of users. Feel free to upload any data you wish to analyze, especially in the context of the workshop. If you don't have sample data, just upload any document as a test.

Note: If you strictly prefer to use command line interfaces to the data store, see the iCommands documentation on the iPlant Learning Center or Wiki. Try installing and configuring iCommands and doing some simple data transfers. More details will be covered at the workshop.

Transferring Data with iDrop

1. Download the iDrop version for your operating system (Windows, Mac, Linux) – see the links at <http://www.iplantc.org/id1>
2. Enter your iPlant Username and Password.
3. To upload – select the place on your local computer (left column) where the files are located:
 - **Windows** – select your hard drive (e.g. C:\) and then navigate your directory
 - **Mac/Linux** – select the files or folder you wish to move (most items will be in the /home directory)

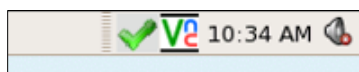
Then drag the file(s) or folder(s) you wish to transfer into your iPlant Data Store (right column).

4. To download – select the file(s) or folder(s) in your iPlant Data Store and drag them to a location on your local computer (left column)

Tip: In the right column of iDrop (your iPlant Data Store) your home folder is the same name as your iPlant username. You will see other folders of data at this level - these are directories that either are setup by other iPlant services or directories of data shared with you. If you get lost navigating, use the **Tree** menu at the top of the application to navigate back to your home folder.

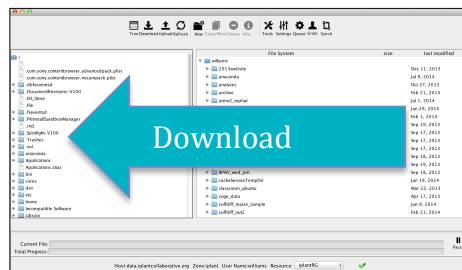
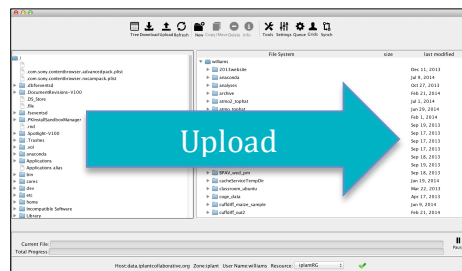


Windows task bar



Mac/Linux

Tip: When iDrop starts, it may appear in your taskbar as a **green check** (usually lower right for Windows computers, and upper left for Mac/Linux). Click on the green check and select **iDrop** from the menu to start the interface.



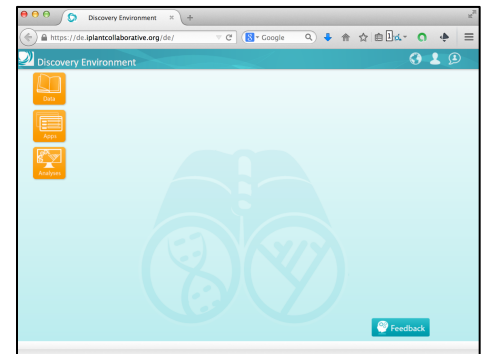
3. Sample Analysis - Use DE to examine illumina sequence with FastQC

How the Discovery Environment relates to our cyberinfrastructure

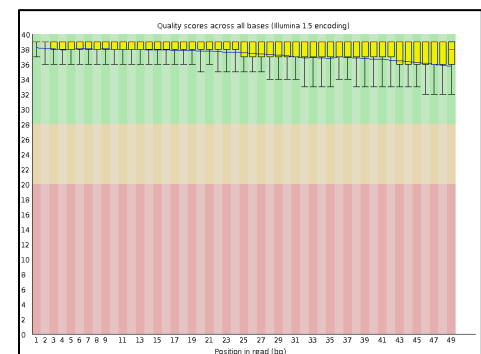
The Discovery Environment (DE) was designed to be an all-purpose bioinformatics workbench – Most data management and popular bioinformatics analyses (e.g. genome assembly, RNA-Seq, phylogeny methods, GWAS, etc.) can be done from start to finish within the DE.

Access quality of short sequence reads with FastQC

1. Login to the iPlant Discovery Environment (click *Launch* from the ‘Discovery Environment’ icon on the iPlant homepage:
<http://www.iplantcollaborative.org>)
2. Click **Apps** from the DE workspace and select the aligner **FastQC 0.10.1** (Location: *Public Applications> NGS>QC and Processing*). Click on the actual app name to run the App.
3. Under “Analysis Name” leave the defaults or make any desired notes.
4. Under “Select Input data” for ‘Input file, click **Browse**, then navigate to and select the **SRR1028781.fastq** file. (Location: *Community Data > iplant_training> ars_workshop>fastqc*). Then click **OK**.
5. Click **Launch Analysis**. You will receive a notification and may close the Apps window.
6. Click on **Analyses** from the DE workspace and monitor the status of your submitted job (You may have to click refresh to view updated status).
7. In the **Analysis** console, once your status appears as ‘Completed,’ click on the name of your analysis to navigate you to the results. Download **SRR1028781_fastqc.zip** using the **simple download**, unzip the files and open the results in a web browser.



iPlant Discovery Environment (DE)



Per base quality graph from FastQC

Expected Outputs

- “logs” - log files,
- “SRR1028781_fastqc” – folder of figures and results in HTML,
- “SRR1028781_fastqc.zip” – results in zip file

4. Launch a virtual machine – Connecting to Atmosphere

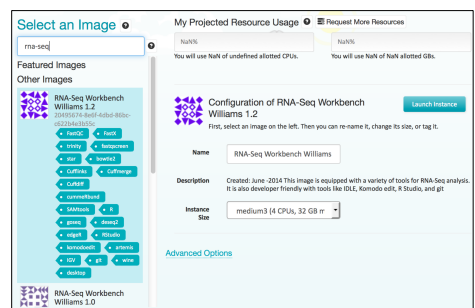
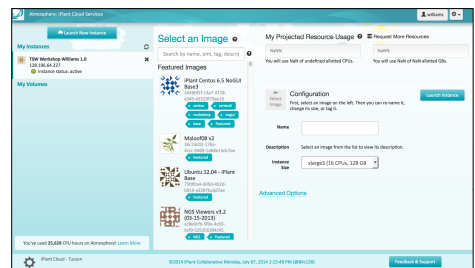
On demand cloud computing

Atmosphere is one of the most versatile components of the iPlant CI. Anything that you would normally be able to do with your local laptop/desktop, you can do on a virtual machine in the Atmosphere cloud. The advantage of using Atmosphere is that you can get access to greater resources (currently up to 16 CPU, 128GB RAM machines). Additionally, those resources are co-localized with the iPlant Data Store so that moving to and from your instance is very easy to do.

Tip: To use Atmosphere, you must have an email address from an academic/governmental institution and request access to Atmosphere through the user portal. To request access, login to user.iplantcollaborative.org and check to see if Atmosphere is listed under 'My Services.' If it is not, scroll down and click the "Request Access" button next to Atmosphere to complete a request form.

Launch an Atmosphere instance

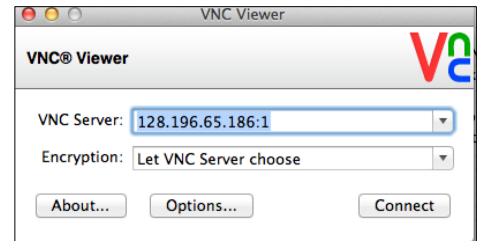
1. Login to Atmosphere (click the *Launch* link from the iPlant homepage at www.iplantcollaborative.org)
2. Click **Launch New Instance** either on the navigation panel (left) or on the home screen.
3. Under 'Select an Image' select **TSW Workshop Williams 1.0**
4. If desired, give your instance a name. Leave the instance size at the default (tiny1) size.
5. Click **Launch Instance**. Your instance should be ready in 10-20 minutes.



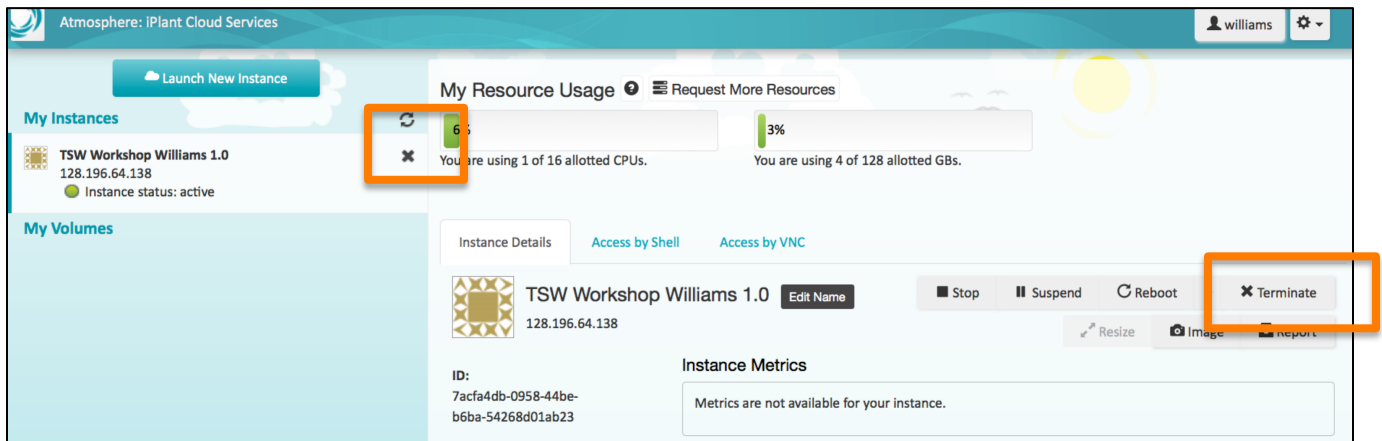
Connect via VNC

6. Download VNC viewer (<http://www.realvnc.com/download/viewer/>)
7. Locate your Instance IP address (beneath your instance name)
8. Enter your IP address + “:1” in the ‘VNC Server’ field (e.g. 161.803.39.887:1) and click connect.
9. When connecting for the first time to an instance, you will be prompted to save a signature. Select **yes** and continue.

Tip: Once you connect to an Atmosphere instance, use iDrop (as described earlier in this guide) to transfer data to and from your new instance. See the iPlant online Learning Center



Once you are connected to the Atmosphere desktop, you can continue on to the challenge exercises. Otherwise, please go back to the atmosphere homepage select your image and click the black (X) next to the instance name or the ‘**Terminate**’ button to terminate the instance. If you need more assistance, visit the iPlant Learning Center Atmosphere page tour (<http://www.iplantcollaborative.org/learning-center/atmosphere>)



5. Additional Exercises and notes for Livestream attendees

Get more familiar with Discovery Environment

- Use the App “**NCBI SRA Import 1.2**” to import the sequence data at the following URL:
`ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR536/SRR536773/SRR536773.sra`
- Use the App “**NCBI SRA Toolkit fastq-dump 2.3.4**” to generate fastq files from the SRA file that was imported.

Get more familiar with Atmosphere and Command line

You can use Atmosphere to practice some of the command line skills we will use for some elements of the workshop. If you have never used the command line, you can find some excellent tutorials at the Software Carpentry (www.software-carpentry.org) and the lesson on the UNIX shell (terminal/command line) are on that site here:

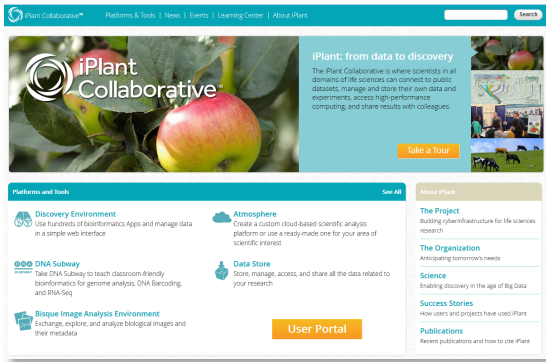
<http://software-carpentry.org/v5/novice/shell/index.html>

Attending a workshop from Livestream

If you are attending the workshop from Livestream, here are a few pointers:

- All of the latest information on the workshop will be posted on the workshop wiki page. There may be last-minute changes or technical glitches that affect our plans for the broadcast so if you have trouble attending, see the wiki page or email support@iplantcollaborative.org
- You will need to connect to the presentation from a computer that supports audio. You probably don't need to do anything else than connect with an up-to-date browser such as Firefox or Safari.
- If you are following along with the Atmosphere demo, we recommend you launch your instance the day before. Follow the instructions on the wiki so that you launch the correct image.
- There will be an etherpad chat for the workshop (virtual and in-person attendees alike); please use this to post questions instead of the Livestream chat which we may not be able to attend to as quickly.

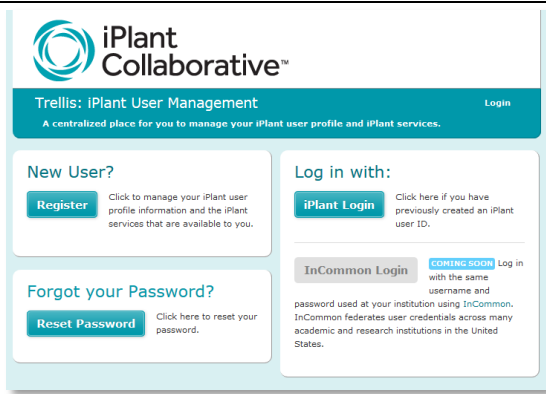
Resources you should know about



iPlant Collaborative Homepage

www.iplantcollaborative.org

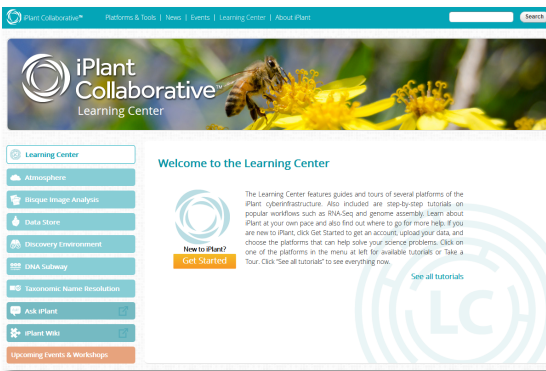
From the homepage you can learn about iPlant's mission, vision, and objectives (long-term or by quarter). You can also directly access our tools and platforms, learn about how to use iPlant platforms, and see how other persons, projects, and organizations make use of iPlant cyberinfrastructure.



iPlant User Portal

user.iplantcollaborative.org

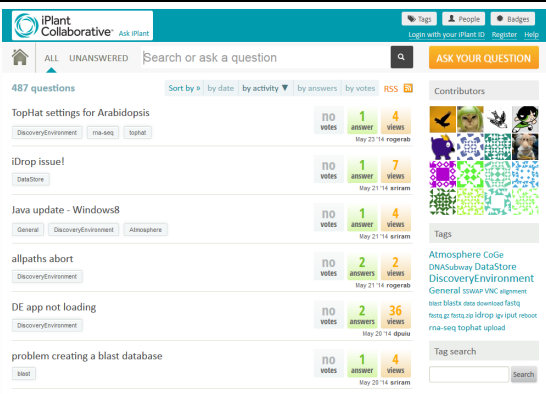
Create an iPlant account and also see what iPlant services you have access to. Once logged in, you can request access to services like **Atmosphere** from the user dashboard. From the user portal you can also reset your password.



iPlant Learning Center

www.iplantcollaborative.org/learning-center

The Learning Center features tours of the major iPlant platforms and tools as well as a growing list of science tutorials. If it is your first time using iPlant try the "Get Started" guide at www.iplantc.org/getstarted



ask.iPlant

ask.iplantcollaborative.org

This forum is a place for you to post any question about iPlant, from technical issues to questions and advice on a specific science objective. Any iPlant community member can help answer your questions, and the forum is monitored by iPlant support to ensure rapid responses.



How to Acknowledge iPlant

Please cite or acknowledge iPlant in any research that uses iPlant resources or extends the cyberinfrastructure. This may take the form of a citation, an acknowledgement, or both, as appropriate.

Acknowledging iPlant

The suggested format to acknowledge iPlant in a paper, a poster, or a presentation is:

This material is based upon work supported by the National Science Foundation under Grant Numbers #DBI-0735191 and #DBI-1265383. URL: www.iplantcollaborative.org

If you wish to additionally acknowledge an individual who assisted you from iPlant, the suggested format is:

We thank [consultant's name(s)] for [his/her/their] assistance with [describe the tasks accomplished], which was made possible through iPlant's Extended Collaborative Support program.

PIs should include a bibliography of articles or other manuscripts (published, accepted, submitted, or in preparation) that benefitted from use of iPlant resources as part of their annual Progress Report and Final Reports to their funding agencies.

Citing iPlant

If you would like to cite an iPlant publication, please cite this paper:

*Goff, Stephen A. et al., "The iPlant Collaborative: Cyberinfrastructure for Plant Biology," *Frontiers in Plant Science* 2 (2011), doi: 10.3389/fpls.2011.00034.*

For more information go to www.iplantcollaborative.org/citeus

Funding

The iPlant Collaborative is funded by the National Science Foundation under grant #DBI - 1265383.

