Instructor Notes for
CyVerse

Once you have received your DNA sequences from the sequencing center, the next step is to analyze your sequences. You will be performing sequence analysis with applications available in CyVerse, a web-based platform that allows you to access computational tools for microbiome analysis.

First, you must create a CyVerse account and upload your DNA sequence files to the CyVerse file management system known as the Data Store.

<u>Goal:</u>

In this tutorial, you will learn how to:

- Create a CyVerse account
- Upload your sequence files into the CyVerse Data Store
- Share your sequence files with students

Create a CyVerse Account for you or your group

1. Go to https://user.cyverse.org/register.
   a. *Note: To avoid glitches, the Chrome browser is recommended.*
2. Fill in the requested information and use an email address to which you can receive an automated message. Select a username. Make sure to use an email address to which you can receive an automated message.
3. An automated email will be sent to the email address you used to create your account. This email will contain a link to confirm that the email address used was valid. **Check your email and follow the steps contained in the email to confirm your address.**
4. Once you confirm your email address, you will be directed to create a password for your new account.

Upload files to the CyVerse Data Store

Before we go over the process of uploading files, first make sure you have all the appropriate data files in the proper format.
- You should have received a folder containing two (2) fastq.gz files for each of your samples. One contains the forward reads and the other contains the reverse

reads for all the sequences in a sample. For each sample the two files should look something like this:

- ■ `SampleID_S#_L001_R1_001.fastq.gz`
- ■ `SampleID_S#_L001_R2_001.fastq.gz`
  - *Note: Each sample will be numbered uniquely with S#. A sample number was assigned to each sample during the sequencing process.*
  - *Verify that you have two (2) fastq.gz files for EVERY sample*

1. Open a web browser and go to https://de.cyverse.org/de/
2. Click on log in with your Cyverse ID
3. Log in using your CyVerse ID
4. Click launch for the CyVerse Discovery Environment. This is an area that can store files and documents for you to use within CyVerse. To upload your sequence files, click on the Data icon (stack of disks in a cloud), which can be found in the navigation bar on the left.
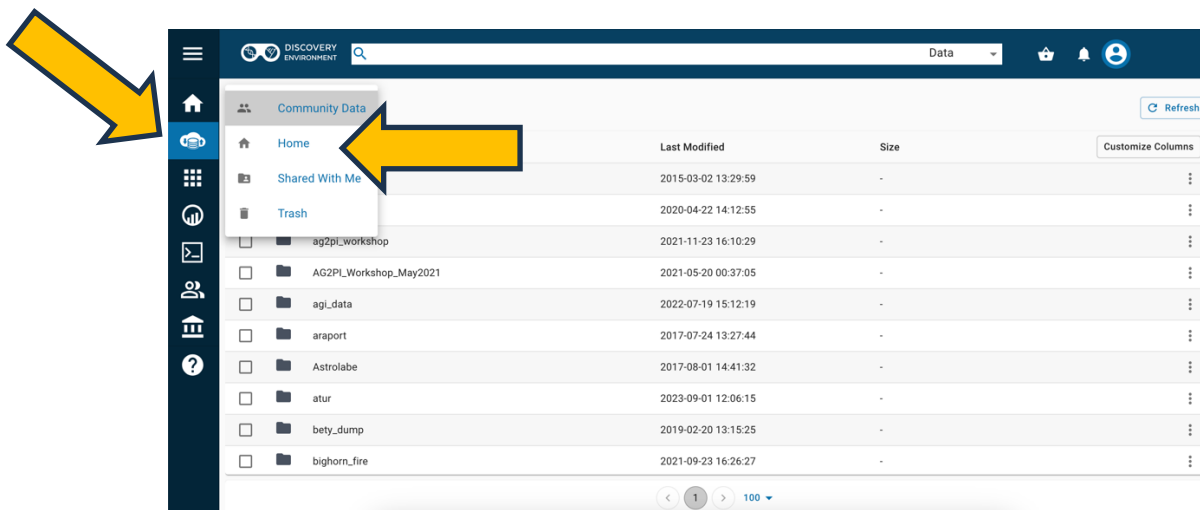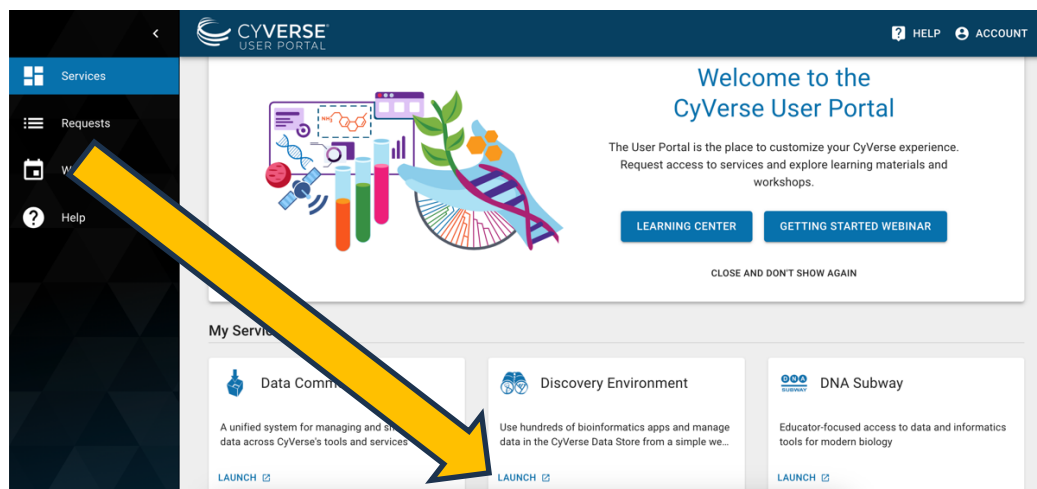
5. Clicking this icon will take you to the community data directory containing data files available to anyone on CyVerse. Click on the pull-down menu on the upper left and select "Home" to access your personal data store (screen shot above). Your page will initially be empty as you do not have any data yet.

6. To create a new folder where you will store your sequences, click on the `Folder` button on the upper right hand corner.

7. A new box will appear. Type in the name of what you would like your folder to be called. Here, I am naming my new file `Beetle_Diet_Microbiome.` Then click `Create.`
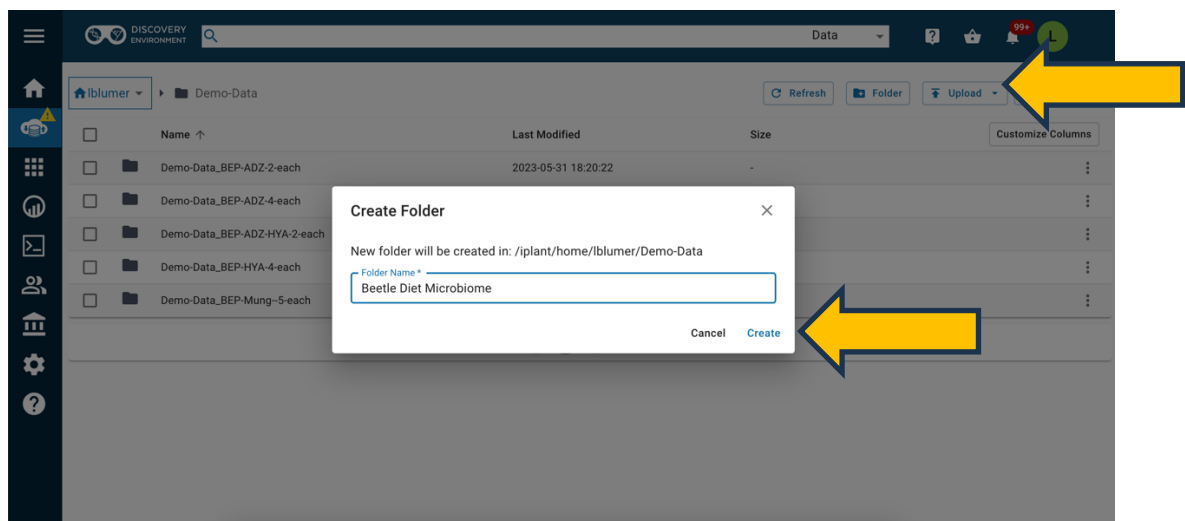
8. After clicking `Create`, you should see your new folder appear on the screen. Click on the name of the folder to enter the folder. It will be empty at this point.

9. To upload data from your local computer, you can click on the `Upload` button on the upper right. A drop down menu will provide options for where you would like to upload data from. To upload from your local computer, click on `Browse Local.`
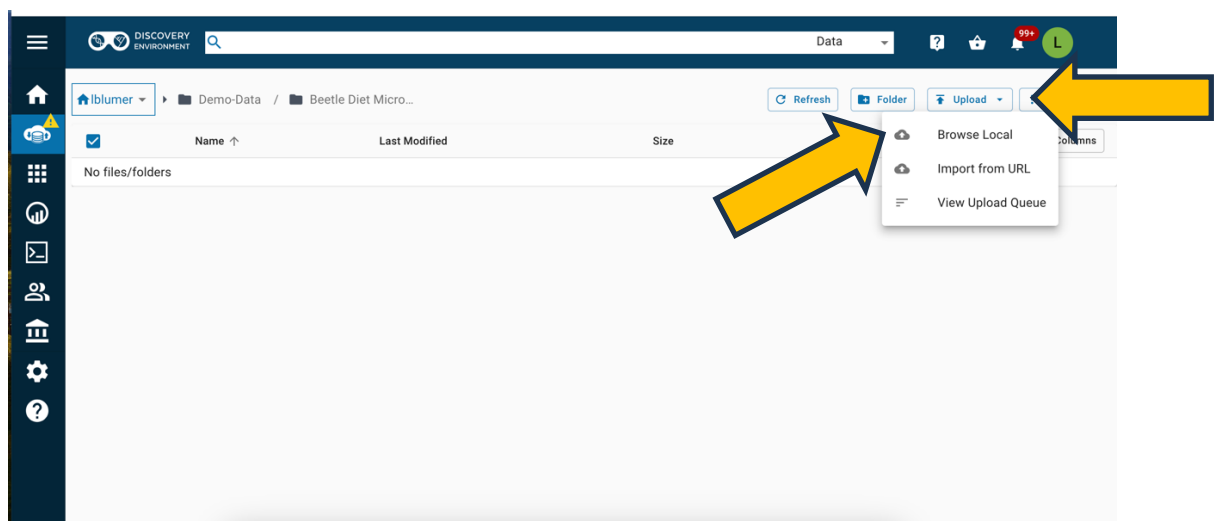
10. A new window will appear that will permit you to locate directories and files on your local computer and directories on your local computer for cloud storage. Navigate to the files you would like to upload, select, and upload the files. Remember to upload both the forward and reverse read files (two fastq.gz files) for each sample.

11. You will now see your files in your folder. If you do not see them, refresh your page by clicking on the `Refresh` button.

12. If you have more files to upload, repeat the upload process until all your files have been uploaded. The "Basic" (free) tier subscription to CyVerse permits you to upload as much as 5GB (5000MB) of data. We have found this to be more than sufficient for as many as 10 experiments with bean beetles containing 12 samples and a negative control (13 samples). However, if you will be continuing to conduct bean beetle microbiome experiments in the future (after completing your work in the Bean Beetle Microbiome Project) or working with larger sample sizes, and using CyVerse to store your sequence data, you may find it necessary to purchase a higher tier subscription (https://cyverse.org/subscribe).

13. Once you have successfully uploaded all your files. You may share those files with students so they may process the sequence data using the DNA Subway Purple Line application in CyVerse.

14. You may share a folder containing all the files for a given experiment as follows:

    - Select the folder of sequence files you wish to share by checking the box to the left of the folder name. Once selected, the upper menu buttons will change.
    - Then select Share on the upper right of the directory window.
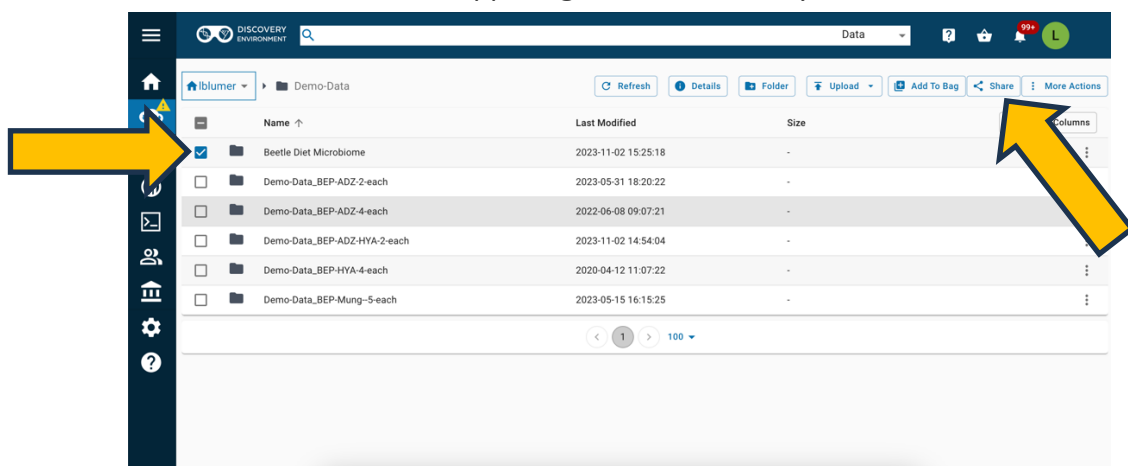
    - A Share window will open. Type the username or email of the user with whom you wish to share the selected folder (see below). CyVerse will search for the user and will display the institutional information about the user. To the right of the user information is a pulldown menu. Select "Read" from that menu. You

want to permit users to read the sequence files but not write new data or own the data in your CyVerse Data Store. If CyVerse does not find a user whose name you entered, they either have not requested a CyVerse account, or they have not confirmed their CyVerse account by responding to the automated email CyVerse sent to the email they entered when requesting an account. Ask them to check their spam folder or have them request a new CyVerse account using a personal email (such as gmail).
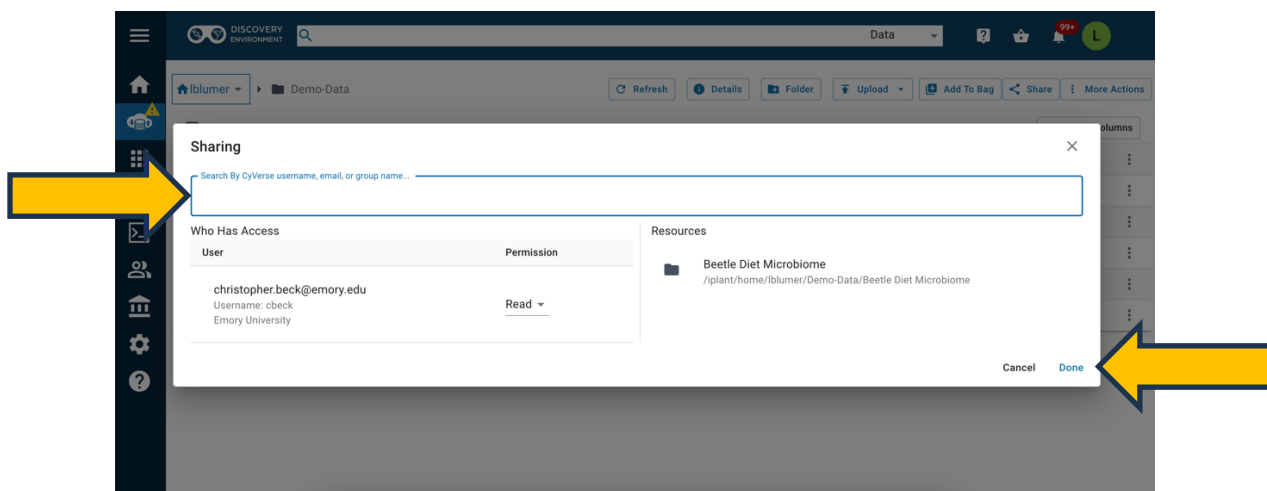
- Keep entering the username or email of each student (or teaching assistant or faculty member) with whom you wish to share files and for each select "Read" access. Note that individuals who have "Read" access cannot re-share data files with other CyVerse users. A user must "Own" the data to share those data with other users.
- Click "Done" on the lower right to complete the sharing request.

15. Now you may log out of CyVerse, or simply close your browser.

Proceed to the **DNA Subway Purple Line Tutorial.**