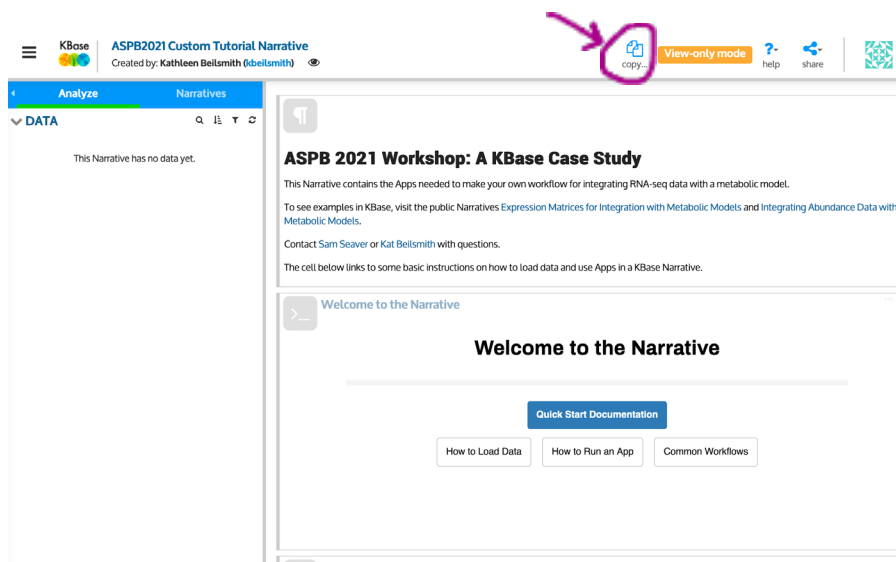


Uploading custom data for A KBase Case Study

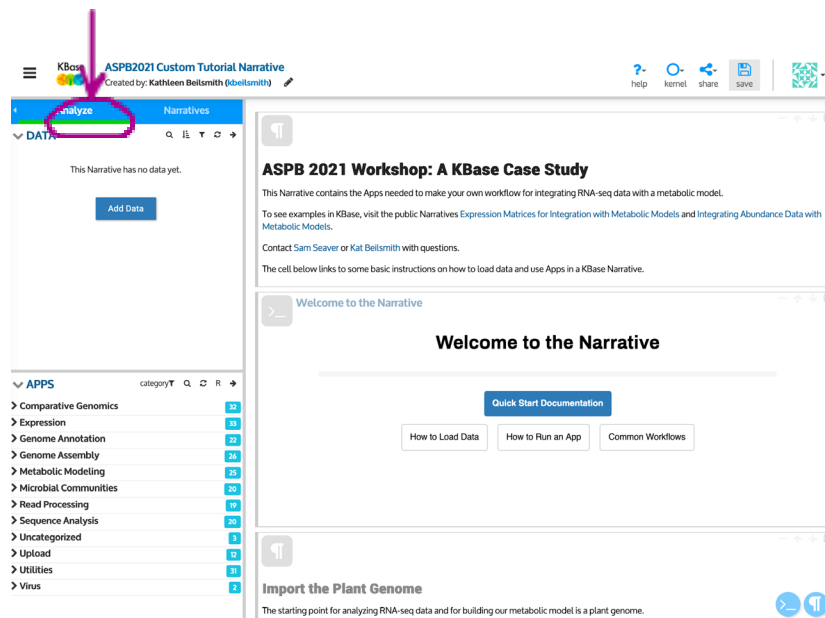
In this workshop, we will showcase a method for integrating transcript abundance data with a genome-scale metabolic model. Users may wish to work with the genome of a plant species other than *A. thaliana* to create a model. In addition, users may wish to import a custom gene expression dataset for integration with the model. Instructions below describe how to create KBase data objects from private or public files in a [customizable tutorial](#) for the workshop.

General instructions

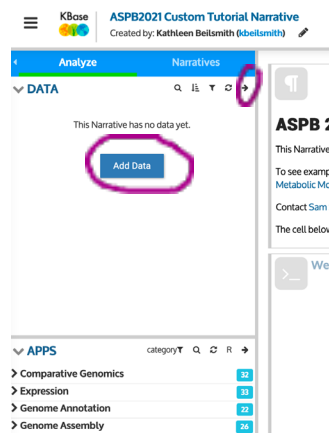
Log into KBase and click the link here to open a [customizable tutorial](#) Narrative for the workshop. Click the blue “Copy” button at the top of the page to make your own private copy of the Narrative that will be saved to your account. You can rename this Narrative and edit the workflow within. This Narrative copy will be visible only to you until you share it.



To start all data imports, go to the Data panel on the top left. (This panel can be expanded or collapsed using the small arrow at the far left of the blue header). The “Analyze” tab should be selected and underlined in green as shown below.



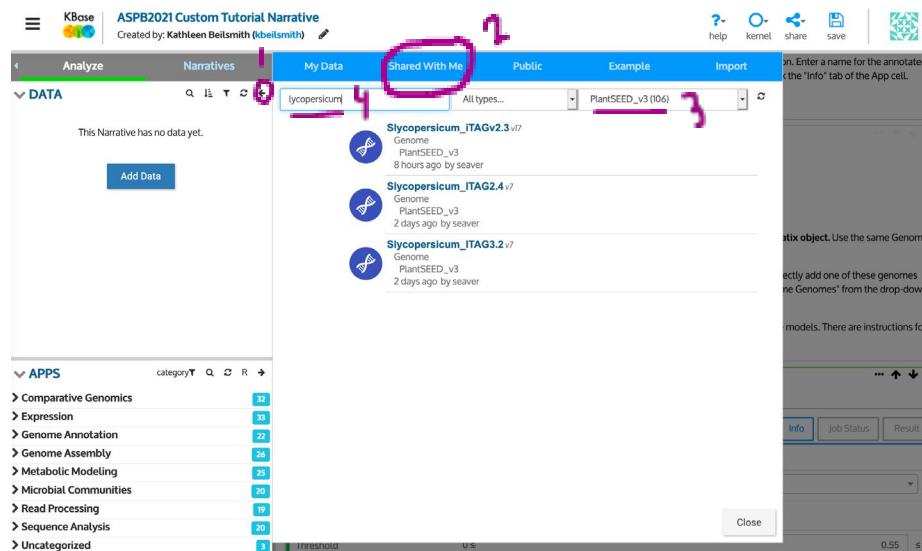
Click the small arrow at the top right of the Data panel. (The “Add Data” button, which sometimes appears as a blue “+” in the lower right corner of this panel, will get you to the same place.) This will open up a data browser. More details on the browser tabs can be found [here](#).



Import a pre-annotated genome from Phytozome

The App “Annotate Plant Enzymes with OrthoFinder” can take several hours to run. For this reason, a Narrative with pre-annotated genomes for all plant species in the JGI Phytozome collection will be available for the workshop. **If you would like to build a metabolic model for a species other than *A. thaliana* during the workshop, please send the species name and your KBase user name to kbeilsmith@anl.gov and we will share the Narrative containing the pre-annotated genomes with you.** A genome can be accessed from this Narrative by selecting the “Shared With Me” tab of the data browser. From the “All Narratives...” drop menu in the top right, select the Narrative titled “PlantSEED_v3.” Then, using the search bar at the top left, type a species name and press enter. The pre-annotated genomes available for that species will

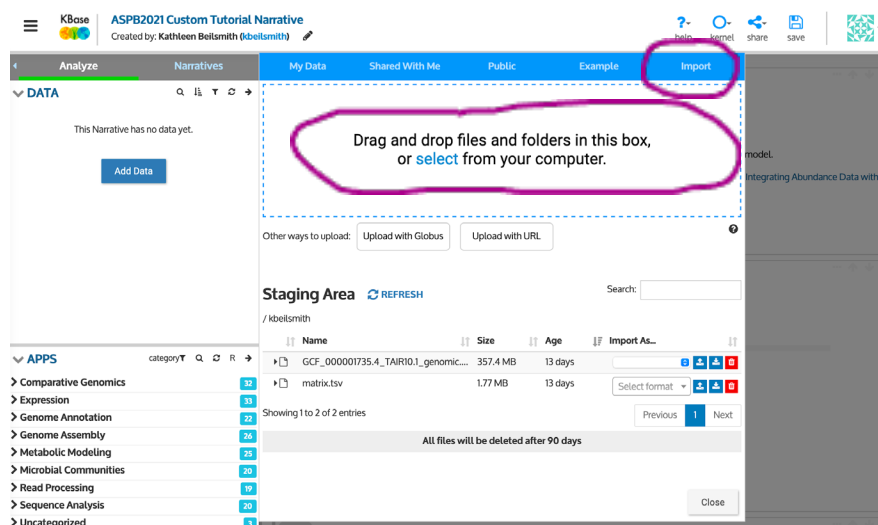
appear in a list in the data browser. Hover over an entry in the list and a blue “Add” button will appear to the left of the entry. Click this button to add the pre-annotated genome to the Narrative.



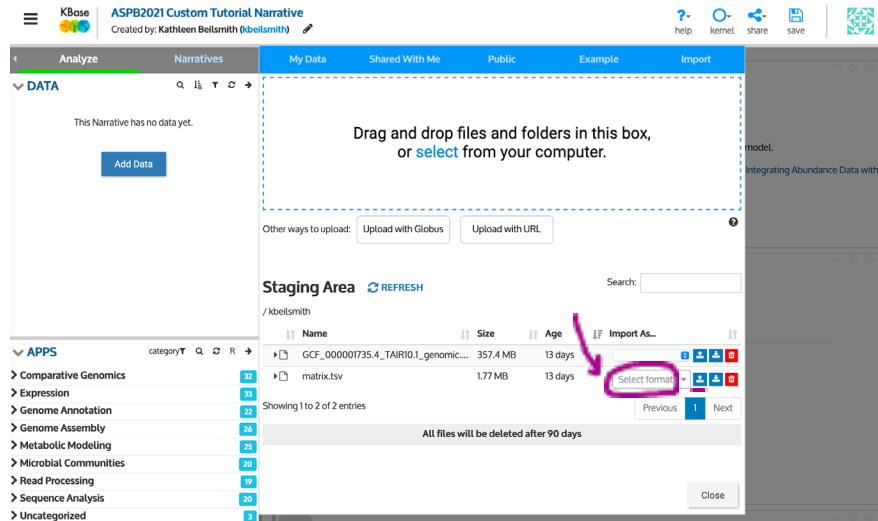
Import expression data from your computer

Expression data should be in a .tsv file with rows for each gene and columns for each sample. The abundances in each cell can be normalized or transformed; the import App will have a place to indicate whether the data is in log2, log10, etc.

Select the tab “Import” to upload data. A box will allow you to drag and drop files from your computer for upload to the “Staging Area.” Once files are in the Staging Area, they can be imported to the Narrative.



For each file in the staging area, use the drop menu to the right to select the type of data for import. In this case, select “Expression Matrix.” Then, the blue button to the immediate right of the drop menu can be used to begin the import process.



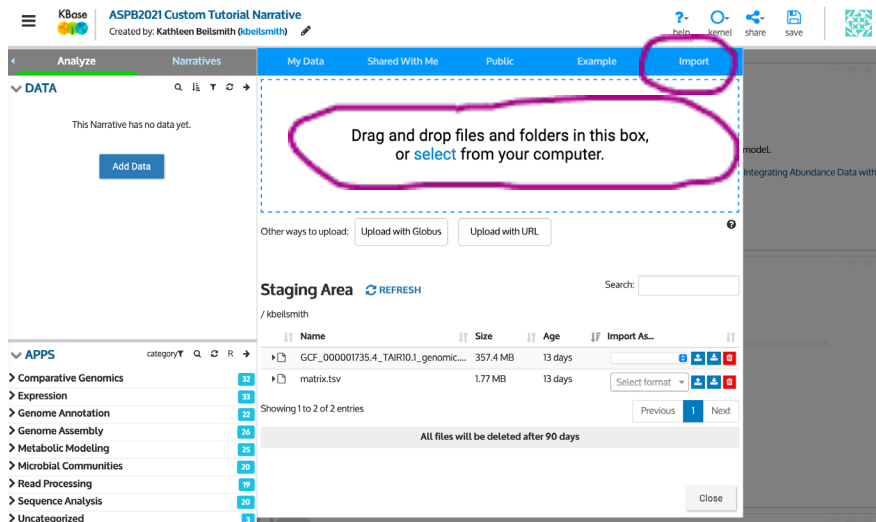
This will open an import App in the Narrative, where you can specify some file properties and a name for the resulting data object. When the App is finished running, the file will appear as an object in the Data panel. More detailed instructions can be found [here](#).

Import raw RNA-seq reads

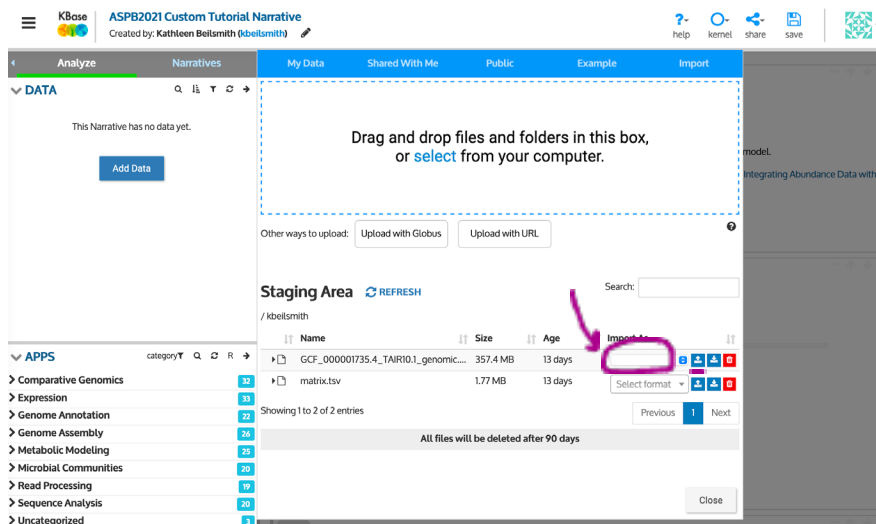
An Expression Matrix can also be generated from raw RNA-seq reads in KBase. Follow the instructions [here](#) to import reads from NCBI or a computer. Then, follow the instructions for RNA-seq analysis [here](#).

Import a genome from your computer

Genomes can be imported as .gbk files or with the combination of a .gff file and a .fasta file. Select the tab “Import” to upload data. A box will allow you to drag and drop files from your computer for upload to the “Staging Area.” Once files are in the Staging Area, they can be imported to the Narrative.



Use the drop menu to the right of the uploaded file in the Staging Area to select the type of data. In this case, select either “GenBank Genome” or “GFF Genome.” Then, the blue button to the immediate right of the drop menu can be used to begin the import process.



This will open an import App in the Narrative, where you can specify some file properties and a name for the resulting data object. When the App is finished running, the file will appear as an object in the Data pane.

In the case of a GFF Genome, you will need to provide the name of the matching .fasta file in the Staging Area to the import App. More detailed instructions are available [here](#). The genome will need to be annotated using the “Annotate Plant Enzymes with OrthoFinder” App in the Narrative.

Import a genome from Phytozome or RefSeq

KBase contains a number of publicly available genomes that have already been loaded from databases like Phytozome or RefSeq. To directly add one of these genomes into the narrative, click on the tiny arrow at the top-right of the Data pane. Then click on the "Public" tab and choose "Phytozome Genomes" or "RefSeq Genomes" from the drop menu at the top left of the data browser. Start entering an organism name in the search bar to the right and press enter to find genomes. When you hover on a genome in the list, a blue "Add" button will appear in the space to the left of the genome. Click this button to add the genome to the data pane for the Narrative.

The screenshot shows the KBase interface for a custom tutorial narrative. The top navigation bar includes 'Analyze', 'Narratives', 'My Data', 'Shared With Me', 'Public' (highlighted with a red circle and labeled '1'), 'Example', and 'Import'. Below the navigation bar, the 'Public' tab is active, displaying a search bar with 'Arabidopsis' and a dropdown menu set to 'Phytozome Genomes'. A list of four Arabidopsis genomes is shown, each with a blue 'Add' button (labeled '4') and a lineage tree. The genomes are: Arabidopsis lyrata v1.0, Arabidopsis lyrata v2.1, Arabidopsis thaliana TAIR10, and Arabidopsis thaliana Araport11. The left sidebar shows the 'DATA' pane with an 'Add Data' button and the 'APPS' pane with a list of applications. The right sidebar contains instructions for adding genomes and a section for the 'Annotate Plant Enzymes with OrthoFinder' app.

There are more instructions on adding public data [here](#). The genome will need to be annotated using the "Annotate Plant Enzymes with OrthoFinder" App in the Narrative.

Please contact the workshop organizers at the addresses below if you have any trouble getting set up for the workshop.

Sam Seaver
seaver@anl.gov
Computational Scientist
Argonne National Laboratory

Kathleen Beilsmith
kbeilsmith@anl.gov
Postdoctoral Appointee
Argonne National Laboratory