Importing and exporting groups data

Video Script

Associated with "04_BioCyc-Groups_Import-Export_071813.mov"

Groups supports the ability to import or export groups, which allows you to use outside sources of data or to use the results of a group outside of the BioCyc website. This functionality can be a very powerful tool for analyzing data not present in BioCyc or to use your own software to do further examination and manipulation of a group's data.

Import a file [*Note: partially reused from basics...]*

To import a file, we can use the Import from Uploaded File option. We'll take a look at the text file first and this is simply a list of E. coli genes, one per line. To create a new group we go up to the groups menu which lists all the groups commands and we'll select "**New Group from Uploaded File**." That prompts a dialog which lets us specify the file and a few parameters. The "Try to make objects" option will attempt to match entries in the file to objects in BioCyc. [click on try to make objects] Once this option is selected we can then specify that we want our data to be recognized as genes by selecting the "gene" choice. The matching of genes against BioCyc makes use of the common-name, synonyms, and unique identifiers stored in the BioCyc gene objects, plus the identifiers used in any database links stored in the BioCyc genes.

If we look again at our import file you can see that the first line is named, "Column 1". We'll select "First line is column headers", so BioCyc will name the column in our group "Column 1" and not attempt to match the line with objects in the database. If our file did not have such a header, we could leave this option unselected. We'll click "**Upload and create group**." The new group is now created.

Let's do the same for a set of compounds, but this time our file will also be more complex. [find compounds file, named compounds.txt] This file contains tab-separated text. Each line will still represent a row in the group that is created, but in addition to that each tab that is encountered will represent a divider that tells the upload tool where a new column should begin. [use arrow keys to show that it's a tab since it looks like whitespace?] The first column will be a set of compounds and the second column will contain other text representing each compound's chemical formula. This time when we select the "**Try to make objects of type"** option we will select the "compound" choice. [[[*WARNING: first column only bug needs to be fixed. The following line can be removed for now if necessary - Tim:* Since we don't want the second column to be correlated to BioCyc objects we will select the "**Try to make objects for the first column only"** option.]]] Let's click **"Upload and create**

group", and now our new group of compounds is made. If we look at the group, you can see that one of the compound names, mannosyl-D-glycerate, doesn't look the same. This means that no compound could be found in our currently selected organism that directly matches this entry. Let's delete that row by using the checkbox next to the row and using the **"Delete Checked Rows"** menu.

Import a group from a replicon coordinates file

In addition to uploading a file using the **"New Group From Uploaded File"** menu, a new group of replicon regions can also be created from replicon coordinates in a file using the **"group from file of replicon coordinates"** menu. The menu describes how the replicon coordinates file should be created. To find replicon names for the currently selected organism, view the organism summary page. [show organism summary page] In the case of EcoCyc there's only one replicon, which is just named, "Chromosome."

We have a file containing replicon coordinates here. [show file regions.txt] The file is a tab-delimited, containing replicon names and start positions, as well as optional fields for end positions and name labels for each region. If end positions are omitted, the region is assumed to be one nucleotide long. Note that start positions have a larger value than end positions to represent a region that has a reverse direction. Let's go back to our groups page and upload the file by selecting, "Groups > New > Group from File of Replicon Coordinates." Click "Upload and generate columns"… and now we have a group of replicon regions. In the first column we can see each region contains the replicon it's part of as well as the start and end positions.

Export to spreadsheet file

One method of exporting a group is to export it as a spreadsheet. Once we have a group's page open we can export it to a spreadsheet through the groups menu by selecting, **"Groups > Export > to Spreadsheet File"** [operate menus] Leave the "Use frame IDs instead of common name" box checked to make it easier to re-import later. The file then downloads to your computer…. [locate local file] There it is. You can open it in Excel or other spreadsheet software. [open newly created file in Excel].

Export to FASTA file

Returning to our group, let's get their sequences using the DNA Sequences property. If a group contains a column of sequences that are compatible with the FASTA format, such as nucleic acid or amino acid sequences, the sequences can be exported in this format using the Export >To Fasta File… menu. The first column in our group will be the name labels in our exported file and the currently selected column will be the sequences. Let's select the sequences column and then perform the FASTA

export via the To Fasta File… menu. [navigate and peform export] The result is downloaded to your computer. [open file]

Send to PortEco

If you have a group with a set of E. coli genes, you can also export those genes to the PortEco Cluster My Genes website for further analysis. To do this, select a column in your group with the E. coli genes you'd like to use with PortEco. [show E. coli genes column] Then, under the Groups menu, go to Export, and then select Export Genes to PortEco Cluster My Genes. [select menu item] You will then be forwarded to the PortEco's Cluster My Genes website to use with your selected genes. [porteco website, maybe show results of the search]