Pathway Tools Walkthrough

This walkthrough will provide information on how to create a BioCyc export from genome sequences using the Pathway Tools software. For information regarding Pathway Tools: https://academic.oup.com/bib/advance-article-abstract/doi/10.1093/bib/bbz104/5669859

1. Installation

To download visit: http://bioinformatics.ai.sri.com/ptools/dist-473256968.html

There is a request for logins, we have our own license so request the detail from your supervisor.
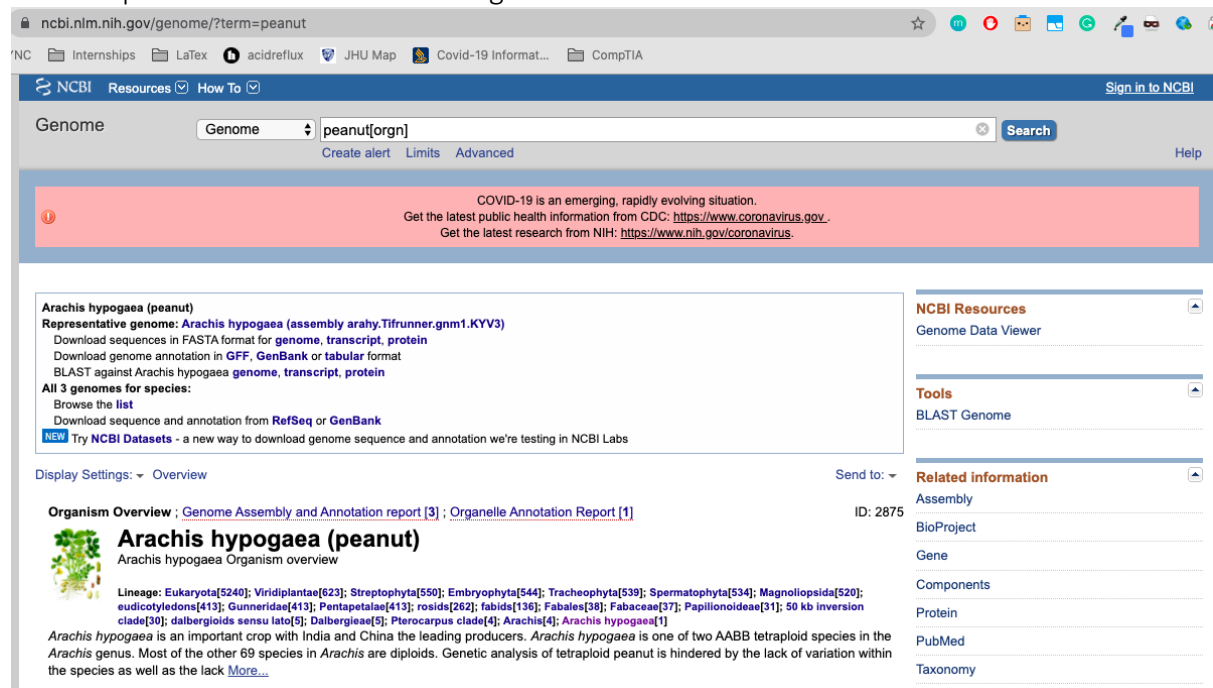
The easiest to start off with is the GUI version which is also the focus of this walkthrough.

2. Getting Genome Sequence

The next step is to download the genome sequence, visit: https://www.ncbi.nlm.nih.gov/genome/

Search up the desired name of the organism in the search bar:
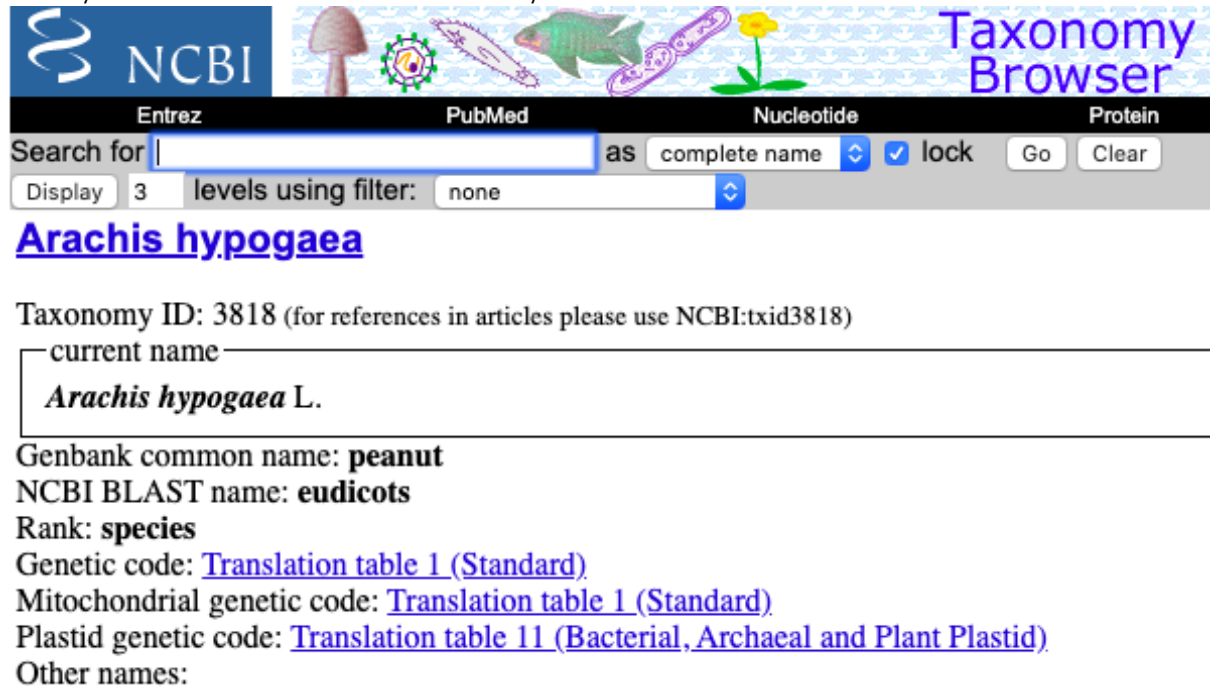

Figure 1. Result of searching peanut in NCBI

There are several formats as can be seen in the box above the organism overview section. The needed file is the genome annotation file which in Figure 1 is available in GFF, GenBank, or tabular format. The easiest to use is the GenBank file so download it and unzip the downloaded file.

While on NCBI, note the taxonomy ID and organism specie as it will be needed later. To get the taxonomy ID, go to the Taxonomy link in the Related Information section in the right-side pane.

Get the specie taxonomy by clicking on the link with the description: "species" underneath then you will see a clade for subspecies on the next page but click on the organism specie to get the taxonomy.

Then you will land on the NCBI Taxonomy Browser:

Figure 2. Peanut taxonomy browser

Note down the taxonomy ID. It is 3818 in Figure 2.

3. Using Pathway Tools

To launch the executable, navigate to the installed pathway tools directory. It is usually located within the home directory in the subdirectory of pathway-tools.

a. Creating Database

Once launched, you will be faced with a landing screen of all the available PGDBs that you have. If this is the first time, then you will have 2 available i.e. the registry PGDBs: E. coli and MetaCyc. Below is my welcome screen with a few other PGDBs:
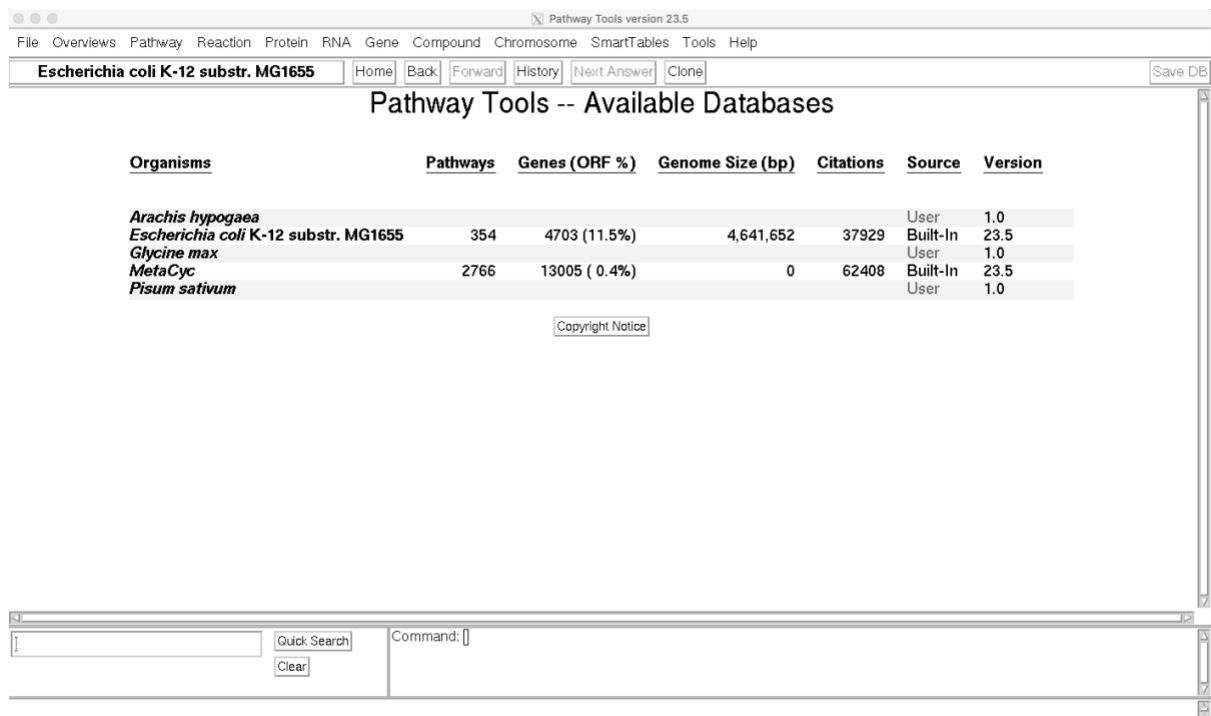
Figure 3. Pathway Tools software Welcome Screen

To create a PGDB go to the navigation bar, click on Tools -> PathoLogic or click Ctrl + T:
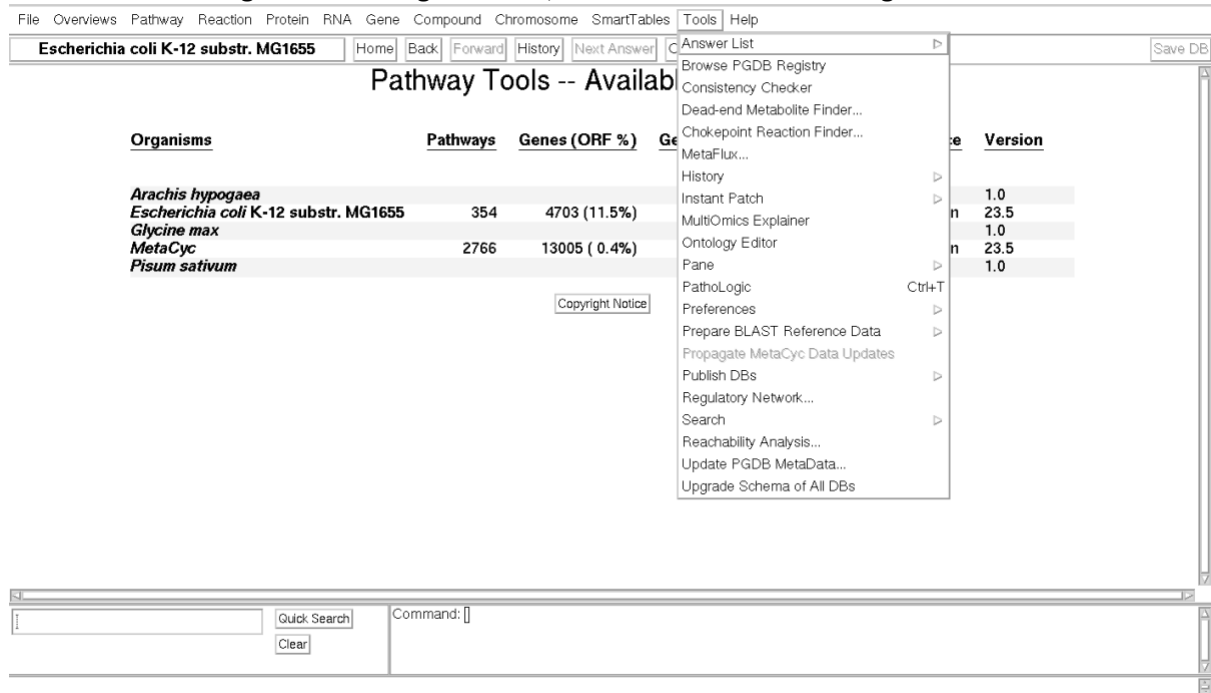
Figure 4. Launching PathoLogic

The PathoLogic tool allows for the creation of a PGDB using the taxonomy ID gathered from NCBI. Once the interface is launched navigate to Database -> Create New.

Figure 5. PathoLogic

Then you will be prompted to enter the project information:

Figure 6. Project Information

Here, give your organism an ID preferably in caps, to continue with the peanut example, I fill:
Organism Project ID: PEANUT

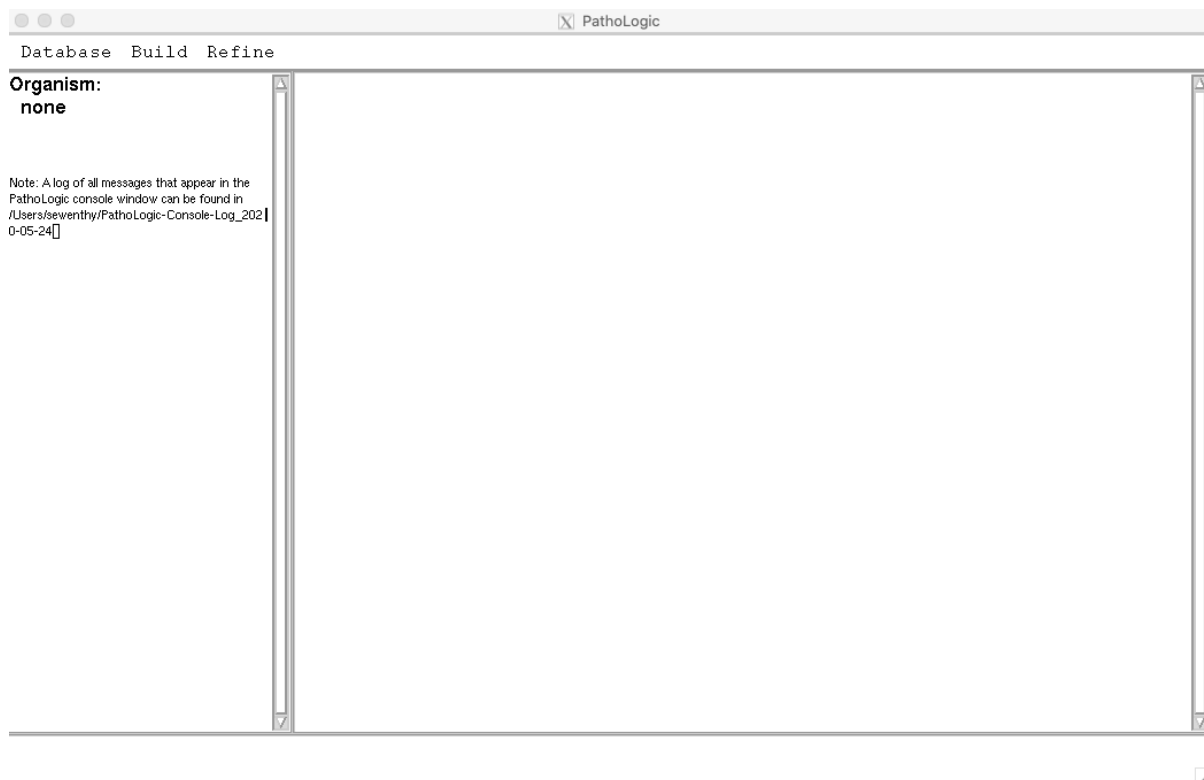Next only the Organism taxonomic class is needed to be filled and that is what is gotten in Step 2. In my case, I fill: 3818 [RET] and the rest is auto filled. Just click OK and you are done.

Once the process of creating the database is running, you will be prompted to enter a Replicon editor. Whether you choose to enter depends on whether you have downloaded and extracted the genome sequence in Step 2. If you are not ready, simply click No for now.

b. Populating Database

Once you have the genome annotation file, you are ready to populate the database. If you were ready in the previous step (3a), you would have already entered the Replicon specification screen but in case you were not, navigate to Build -> Specify Replicons in the PathoLogic navigation bar:
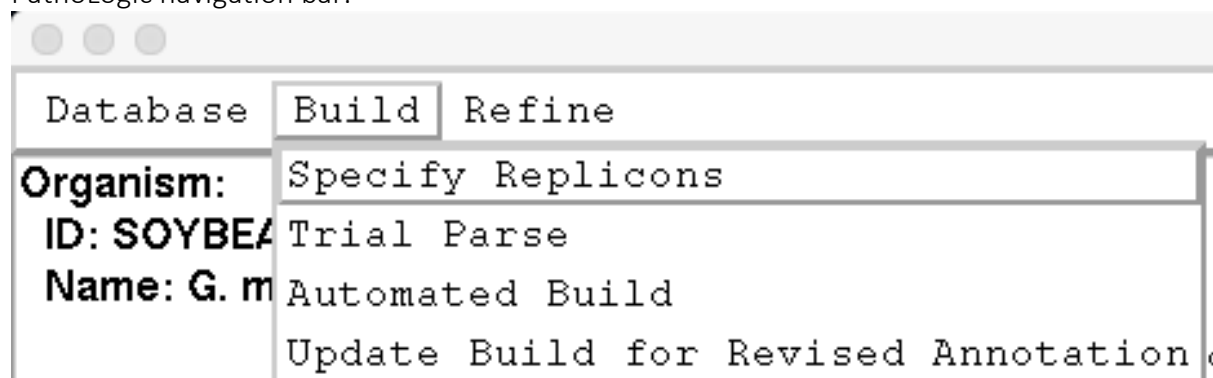

Figure 7. Specify Replicons option

This is where the GenBank file comes in handy. If you have other annotation formats, you can still use it but there might be more information that you have to fill in, read the instruction above the Replicon specification section.

Figure 8. Replicon specification

To import the annotation file, navigate to the Replicon section and click Browse on the far right-side. Navigate to where your extracted GenBank file is and click OK.

Now this is where there is quite a long wait. Note that the window will not show any activity and the file name shown "GFC_0000..._genomic.gbff" in Figure 8 is not there on your screen as of now. This is because Pathway Tools is reading the file and if you are on Mac, look over to the terminal window running the Pathway Tools instance and you can see messages like: "114936176 bytes have been tenured, next gc will be global.
See the documentation for variable EXCL:*GLOBAL-GC-BEHAVIOR* for more information."

You have to wait until the file name that you selected shows up like in Figure 8, then you can click OK. Note that this will take quite a long time on a slow computer and since Pathway Tools is very memory hungry, this process can take up at least 3GB of memory.

Once the Replicon is imported, navigate to Build -> Automated Build in Figure 7. Again this will be a very long process that takes up quite a bit of memory and all you can do is wait.

Once that is also done, navigate to Database -> Save DB and then quit PathoLogic. On the main window, you can navigate to File -> Available Databases and you should see your newly created PGDB. If you click on it from the list, you should see all the imported replicons, reactions, and pathways. Here, just to be safe click on Save DB at the top right corner.

c. Exporting Database

Now to export the PGDB into BioCyc format, you simply navigate to the database you want to export, navigate to File -> Export -> Entire DB to attribute-value and BioPAX files.
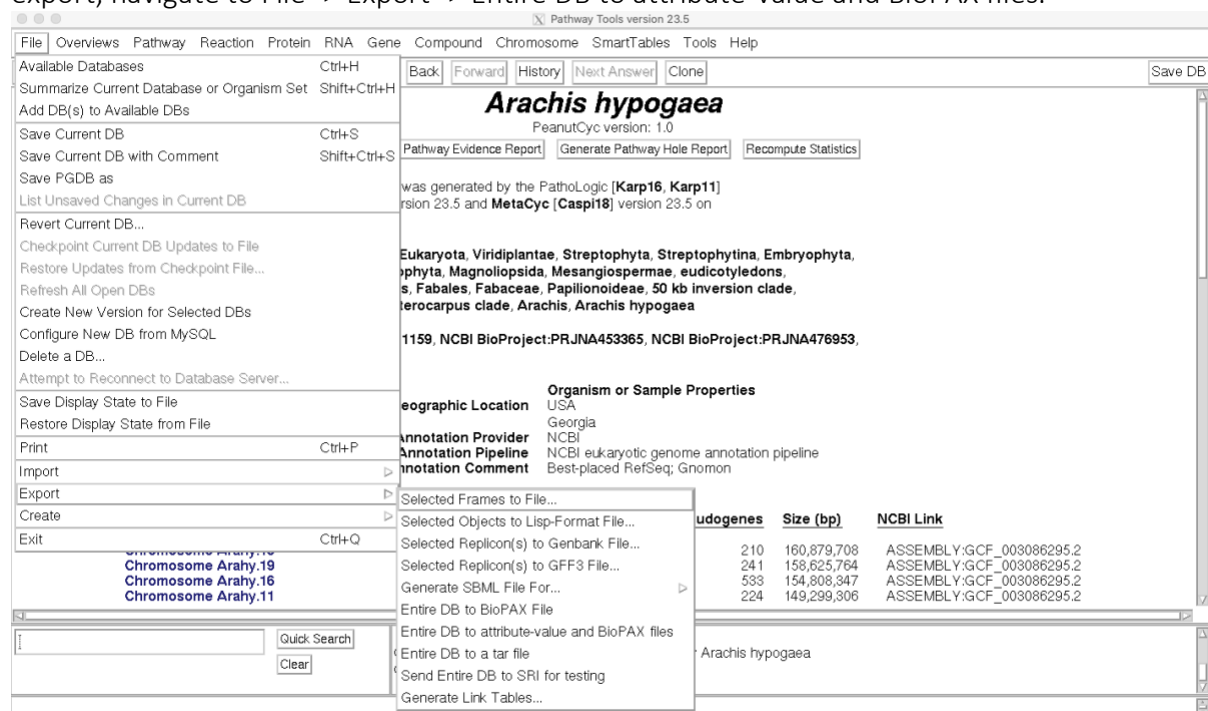

Figure 9. Export screen

Again, this will be a very long wait. After you are done, you should be able to find your export in your ptools-local folder.

AUTHOR:

Sewen Thy
s.thy@u.yale-nus.edu.sg