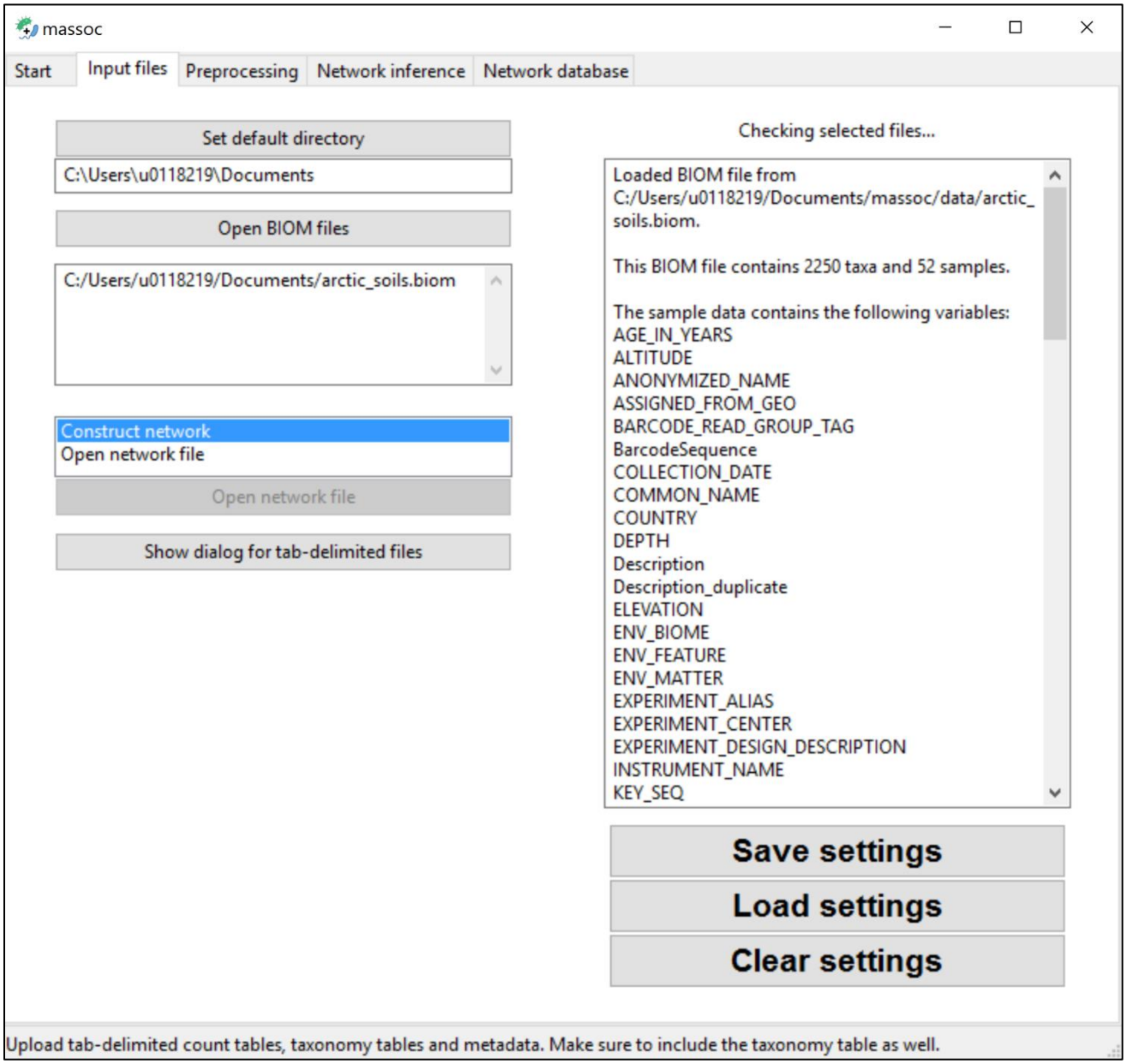
Case study: corralimorphs from the Palmyra atoll

In this case study, we are going to perform an analysis of data collected from arctic soils. This data can be found here: https://qiita.ucsd.edu/study/description/104

Chu H, Fierer N, Lauber CL, *et al.*: Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ Microbiol.* 2010; **12**(11): 2998–3006.

In the original BIOM file, the mapping file is not included. The BIOM file in the Github repository has already had the mapping file added:
https://github.com/ramellose/massoc/raw/master/data/arctic_soils.biom

First, set the default directory and load the data. The BIOM file will be checked automatically, and if it can be imported, the pane on the right will list properties of the BIOM file.

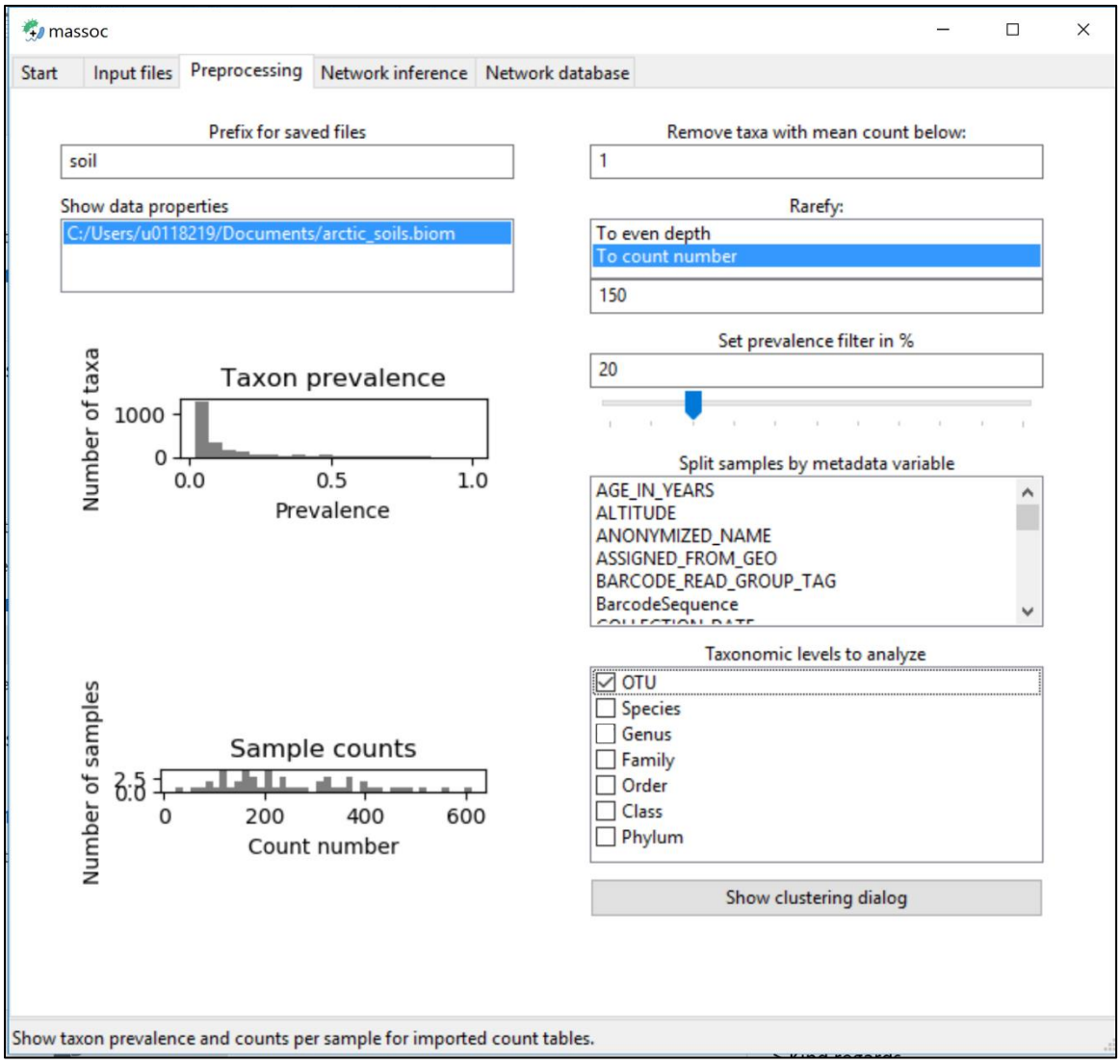


Next, select some preprocessing steps. The prefix is set to ‘soil, so any intermediate files generated by *massoc* can be identified with this prefix.

Suitable preprocessing steps can be selected with the prevalence and count plots. In this case, there appear to be many taxa that only occur in a few samples, and many samples have very low counts. To filter taxa, the minimum mean count is set to 1, and a prevalence filter of 20% is applied. T

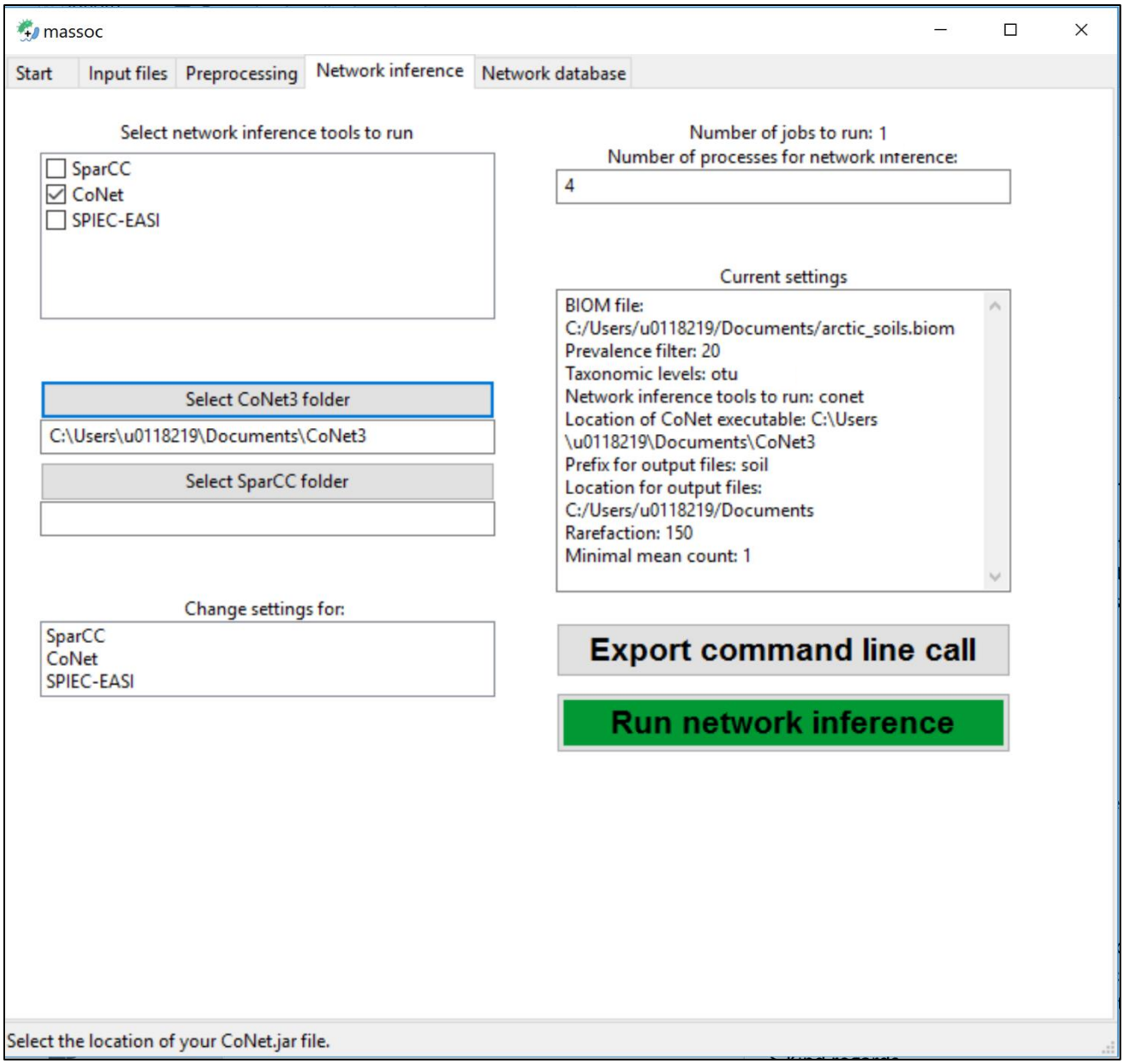
However, there are also some samples with low counts. Set the count number for rarefaction to 150; any samples below this number will not be included, and all remaining samples will be rarefied to 150 counts.

Finally, we can agglomerate taxa before network inference to check associations at different taxonomic levels. In this case, select OTU.



The next step is to infer networks. Select CoNet in the checkbox and make sure to specify the location of your CoNet3 folder. Here, we are just going to run CoNet with the settings specified in the CoNet.sh script. Currently, it is not possible to change these settings in the *massoc* executable, but an update should fix this in the near future.

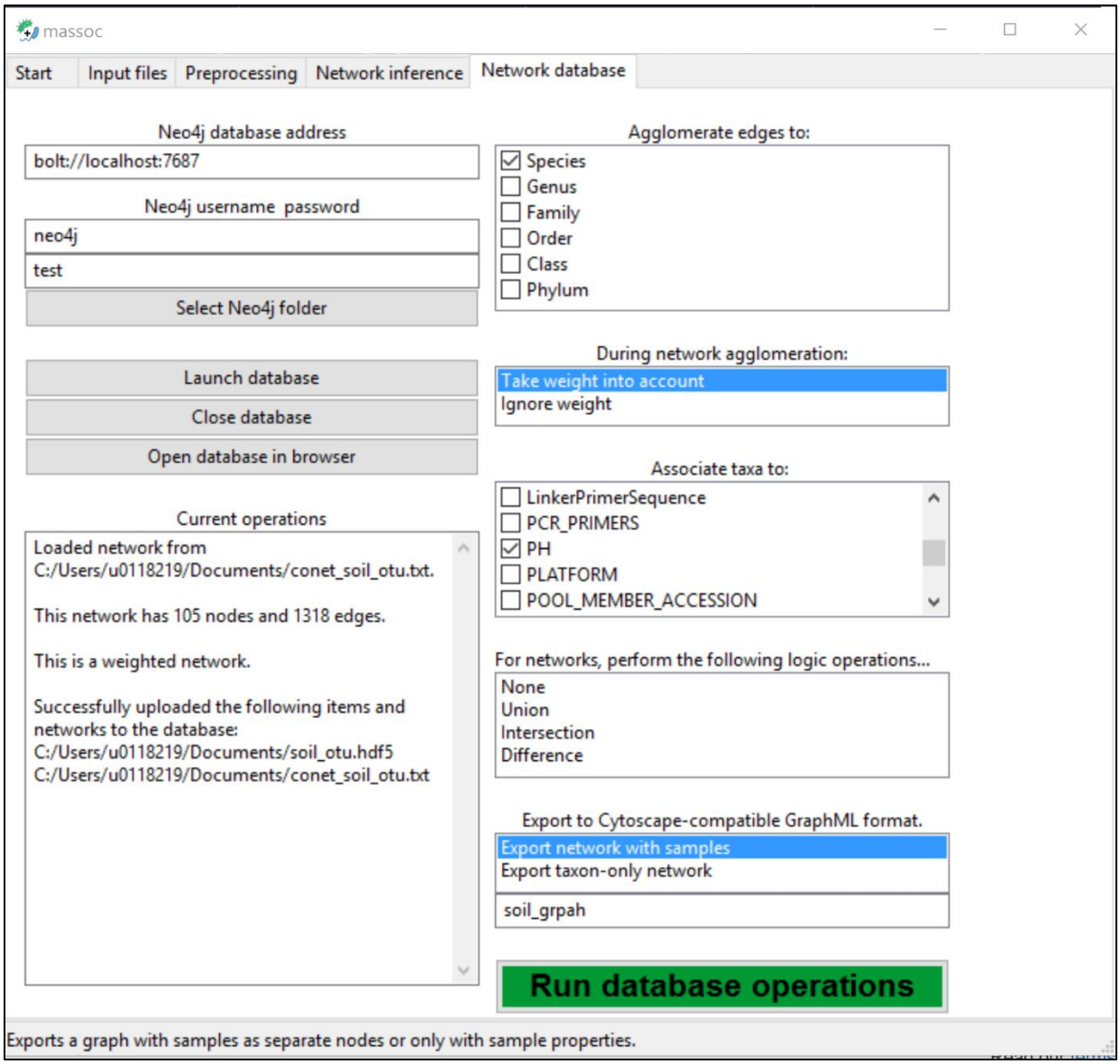
The review pane on the right should show some of the settings that have been selected so far. Run network inference by clicking the green button.



After you complete network inference, a text file of the edges will be written to disk. However, to obtain a rich GraphML file that is compatible with Cytoscape, it is necessary to first connect your network data to the original BIOM file. This demo assumes you already installed Neo4j Server on your machine. If not, please refer to https://neo4j.com/docs/operations-manual/current/installation/.

Specify your address, username and password for the Neo4j Server. Also select the folder that contains your Neo4j Server distribution. Afterwards, click *Launch database*. This will upload the BIOM files and network files to the database.

After the database launches successfully, you can carry out further operations. Set edge agglomeration to *Species* and take weight into account. Sample associations can also be interesting: select PH, DEPTH and ALTITUDE. To have sample properties as separate nodes in the GML export, select *Export network with samples.* Then click *Run database operations*.



A rich graph file named *soil_grpah.gml* will be written to disk, and you can import this into Cytoscape. Alternatively, open your network in the Neo4j console by clicking *Open database in browser.*