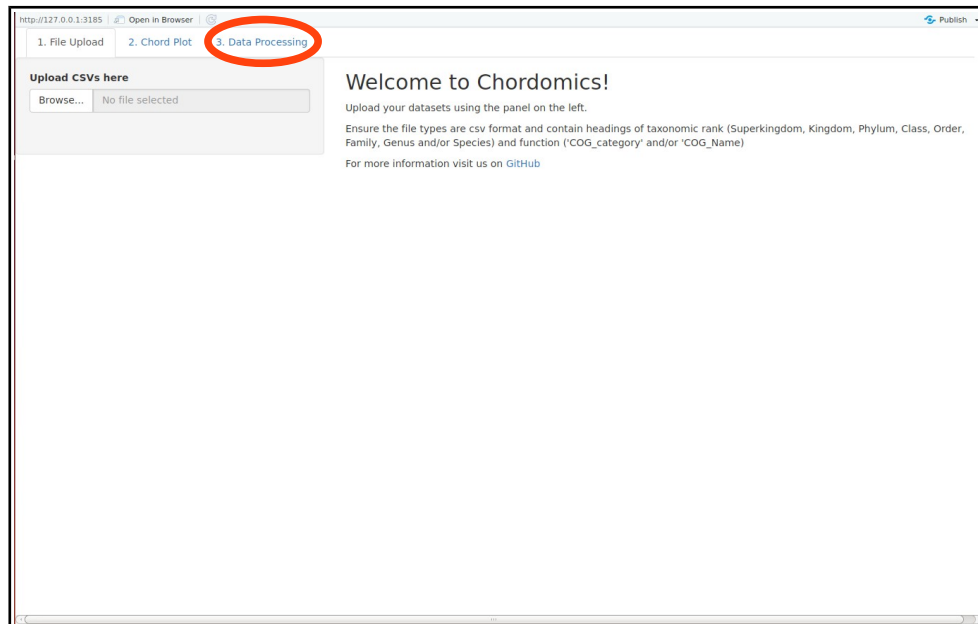
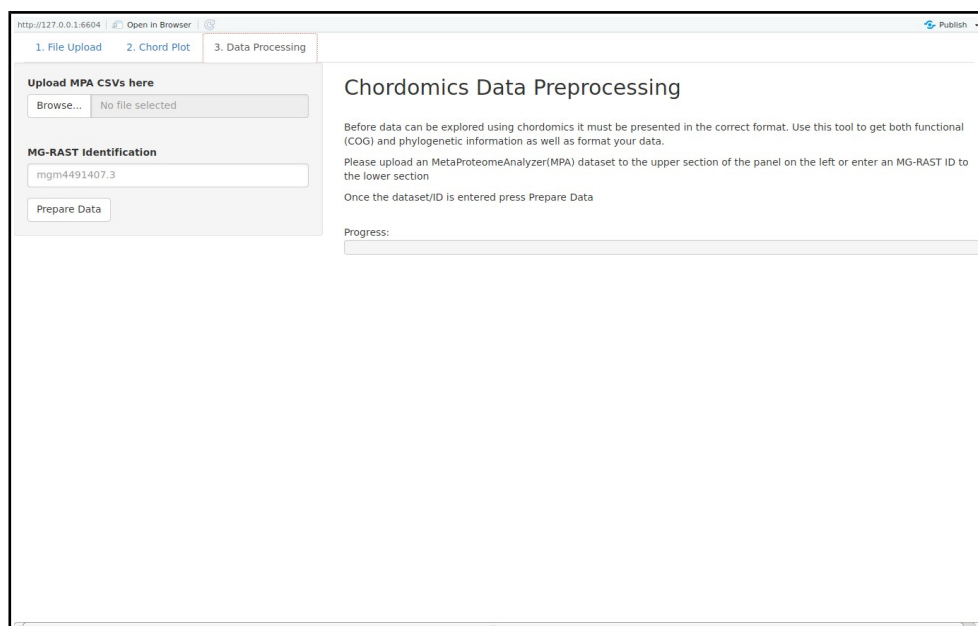


Process Data

To process your data go to the “Data Processing” tab.

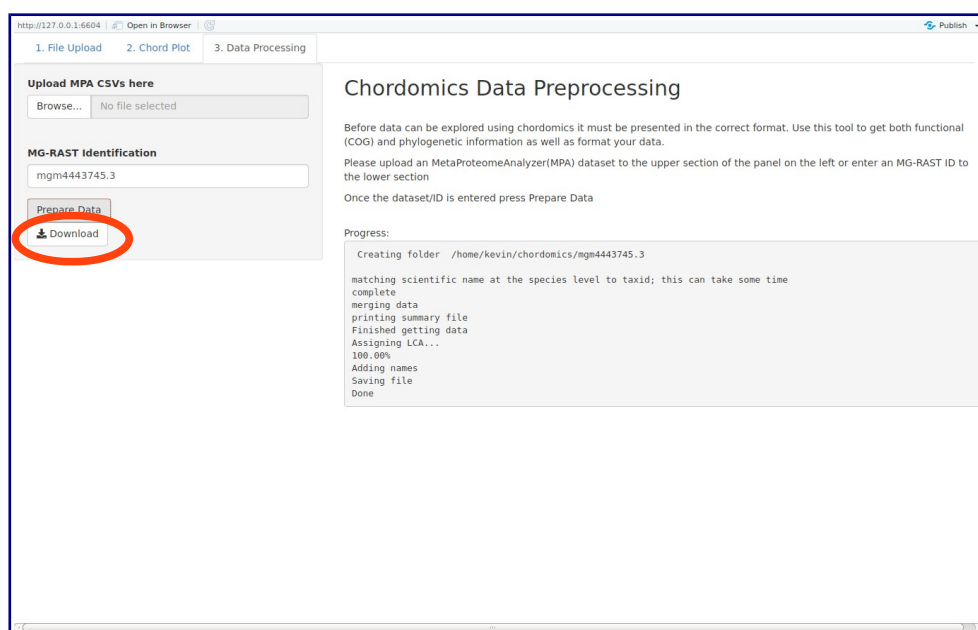


The software provides two options for data input. Given an MG-RAST ID, e.g. mgm4491407.3, relating to your dataset the app will automatically download the RefSeq taxonomy file and the COG functional file. These are then combined by read ID and assigned taxonomy down to the lowest common ancestor of the possible species. Finally the COG IDs are labelled with their category and name.



If an MPA output csv file is loaded in, the app provides a different processing pipeline. The “Proteins” column provides UniProt accession numbers for the possible proteins for each entry. These are used to find all possible COG IDs associated with each entry. These are largely the same as the possible proteins are very similar to one another. In cases where they are not an entry is split into however many possible COGs there is.

Once either a MPA dataset is loaded in or an MG-RAST ID is entered, press “Prepare Data” to run the appropriate processing pipeline.



The Progress window will keep you up to date on what is happening. For large datasets this may take quite some time.

When its finished a “Download” button will appear. If run from R Studio the software will not rename your file. This is a problem R Studio has with downloadHandlers. Specifying a name and adding “.csv” will successfully save your data.

Once your files are processed they can be uploaded into Chordomics using the “File Upload” tab for viewing.