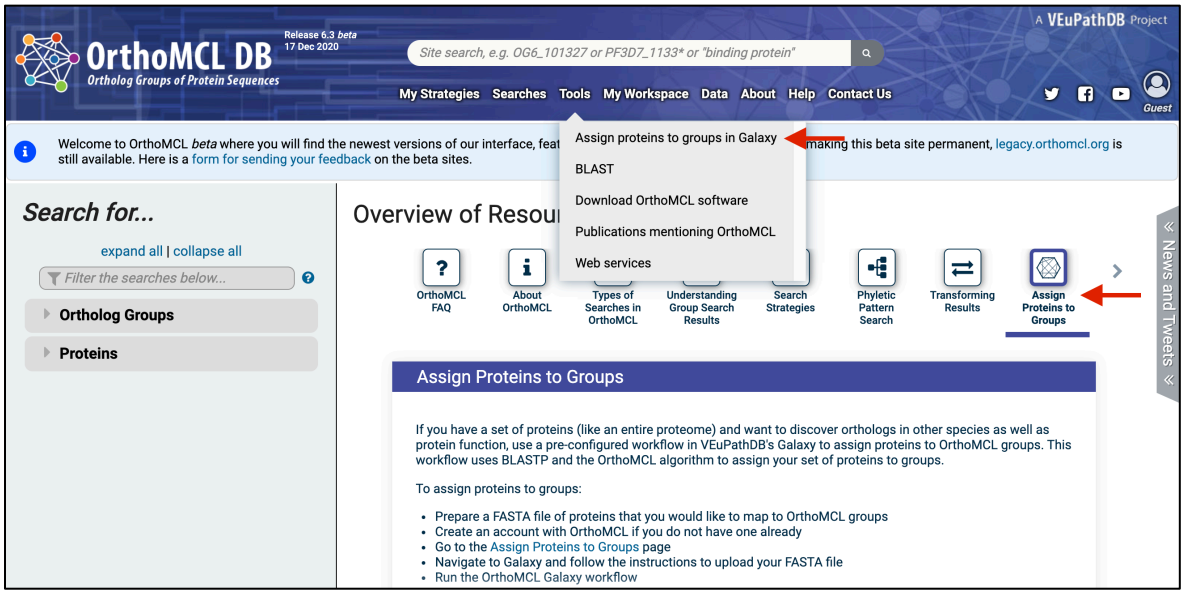
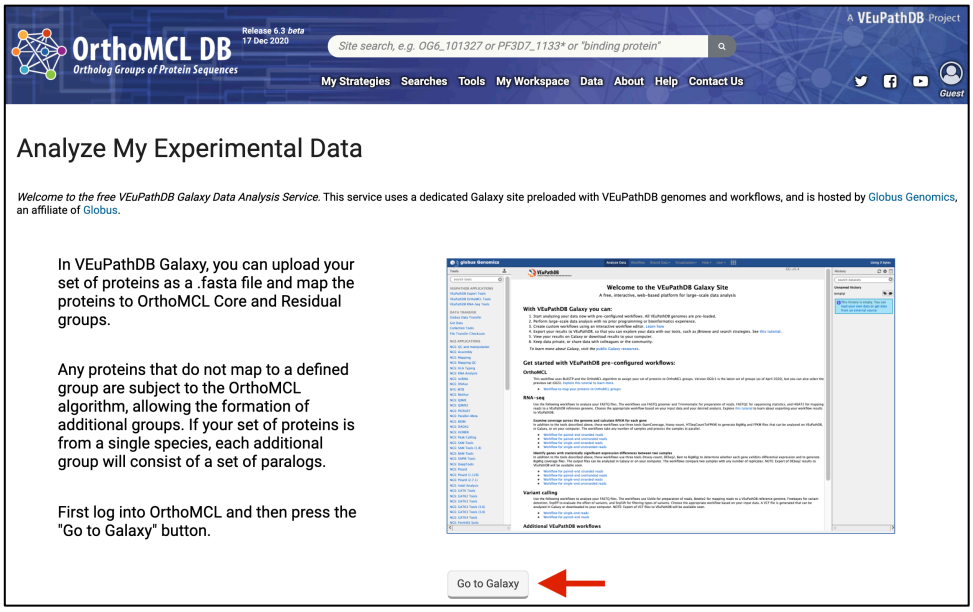
Map your proteins to OrthoMCL groups

This tool allows you to map proteins from a FASTA file to OrthoMCL groups. The tool has been implemented as a workflow in the VEuPathDB Galaxy workspace. To use this tool, follow these steps:

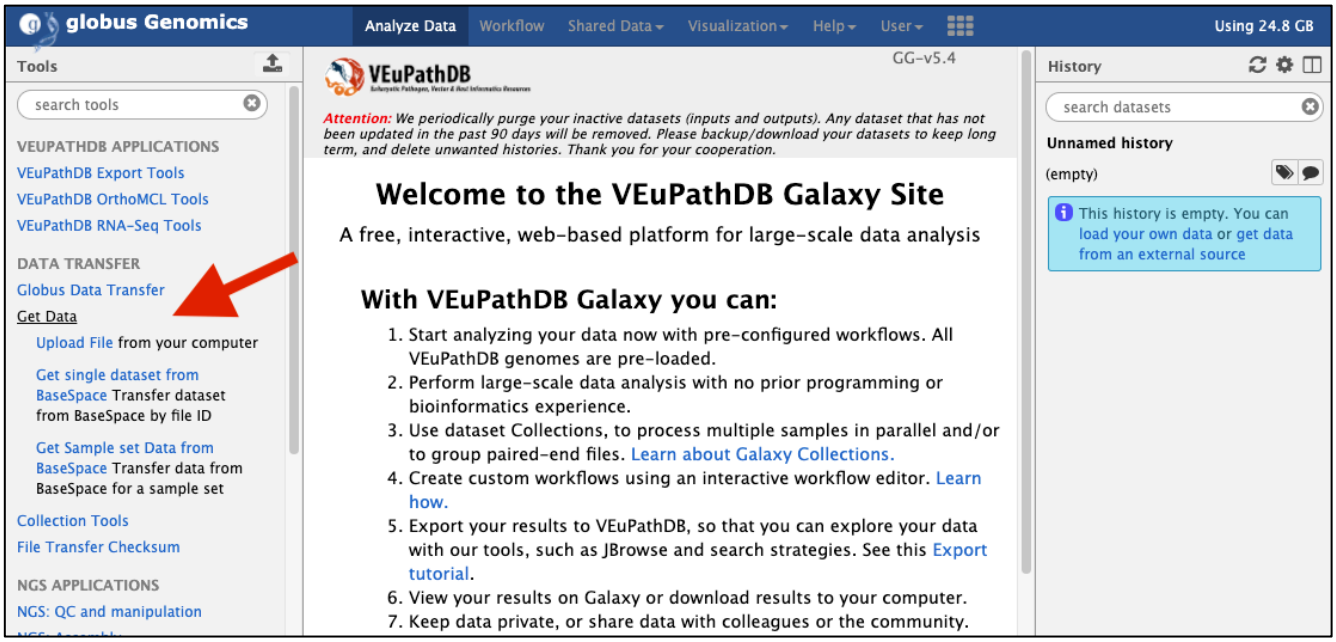
1. There are two ways to access this feature: (a) Click on the “Tools” item in the grey menu and then select “Assign your proteins to groups in Galaxy”. (b) On the home page, click on the “Assign Proteins to Groups” entry in the Overview of Resources and Tools section. Then, select the link entitled “Assign Proteins to Groups”.



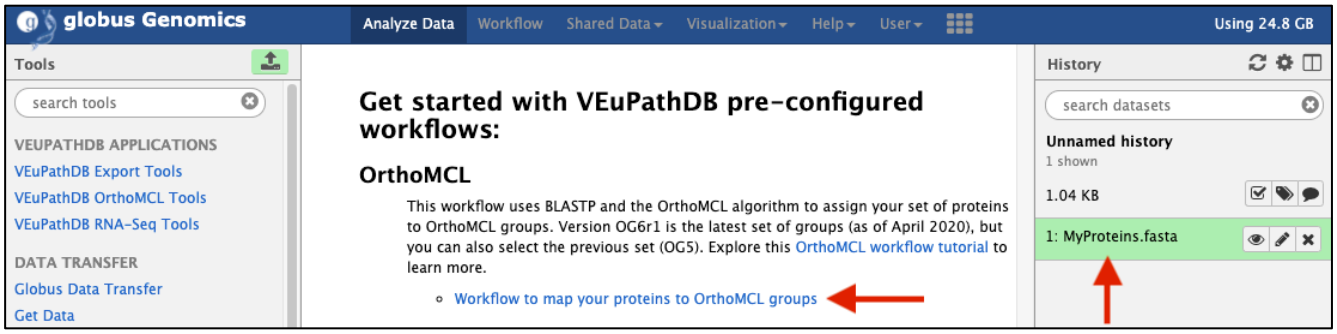
2. The next page provides instruction and a link to the VEuPathDB Galaxy server. Click on the “Go to Galaxy” button. **Note:** to use this service, you must create a VEuPathDB account.



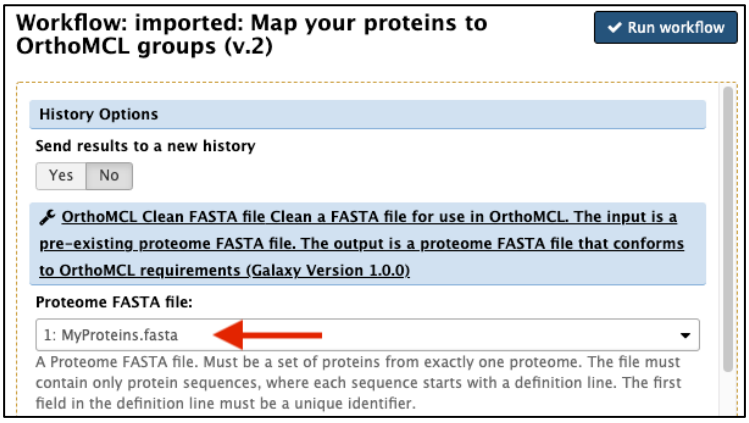
3. Once in Galaxy, import your FASTA file containing protein sequences by clicking on the “Get Data” option in the left panel and selecting an import option, such as “Upload file from your computer”.

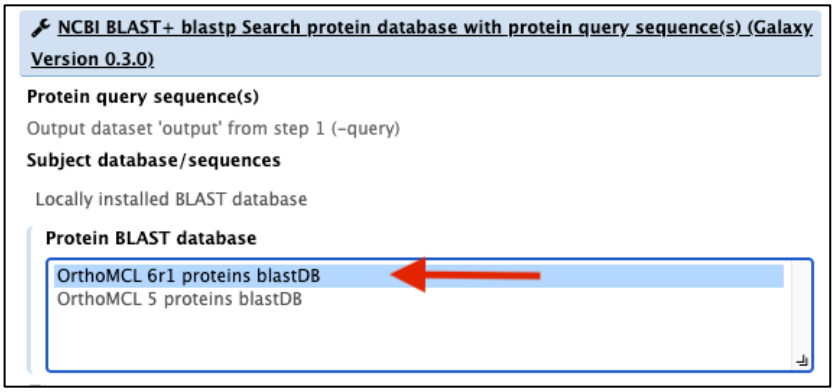


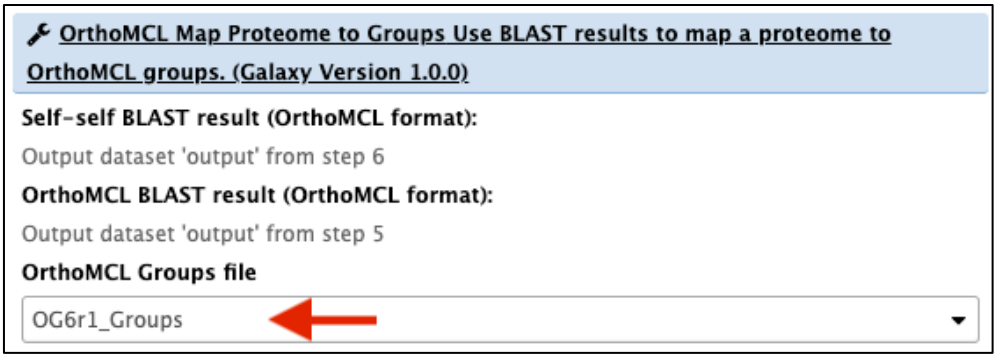
4. Your imported file will appear as a step in the History panel located in the right-hand side of the page. The color of the step indicates its status: grey indicates *in queue*, yellow *in process*, and green *completed*.
5. Once your file has been successfully uploaded into Galaxy, select the “Workflow to map your proteins to OrthoMCL groups” from the middle section of the page.



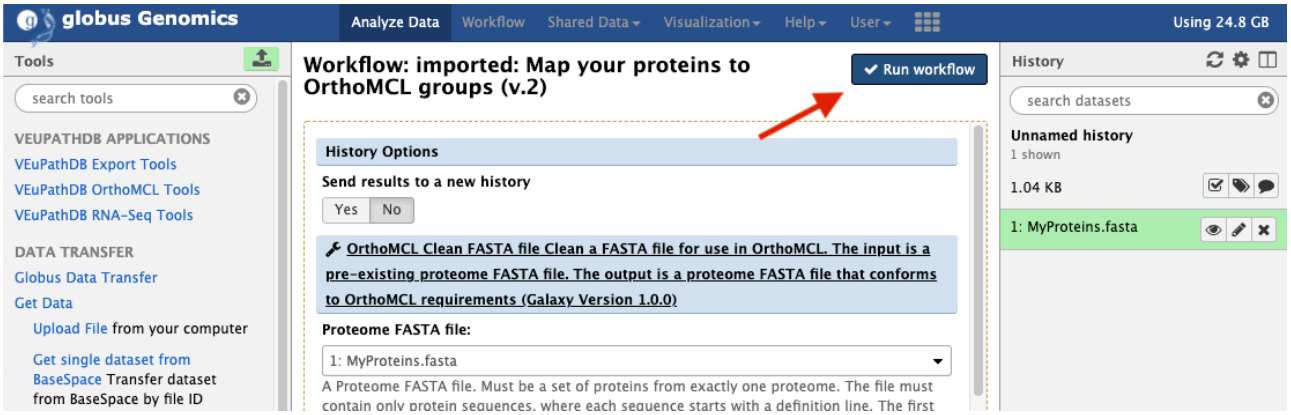
6. The pre-written workflow will be imported into your workspace, where you can process your protein file. On the workflow page, there are three options that you need to set: (a) Proteome FASTA file, (b) Protein BLAST database, and (c) OrthoMCL Groups file. For (b) and (c), it is recommended to choose the most recent version, which is 6r1 or later; version 5 is an older version that is kept for those that would like to compare their current analysis with their old analyses.



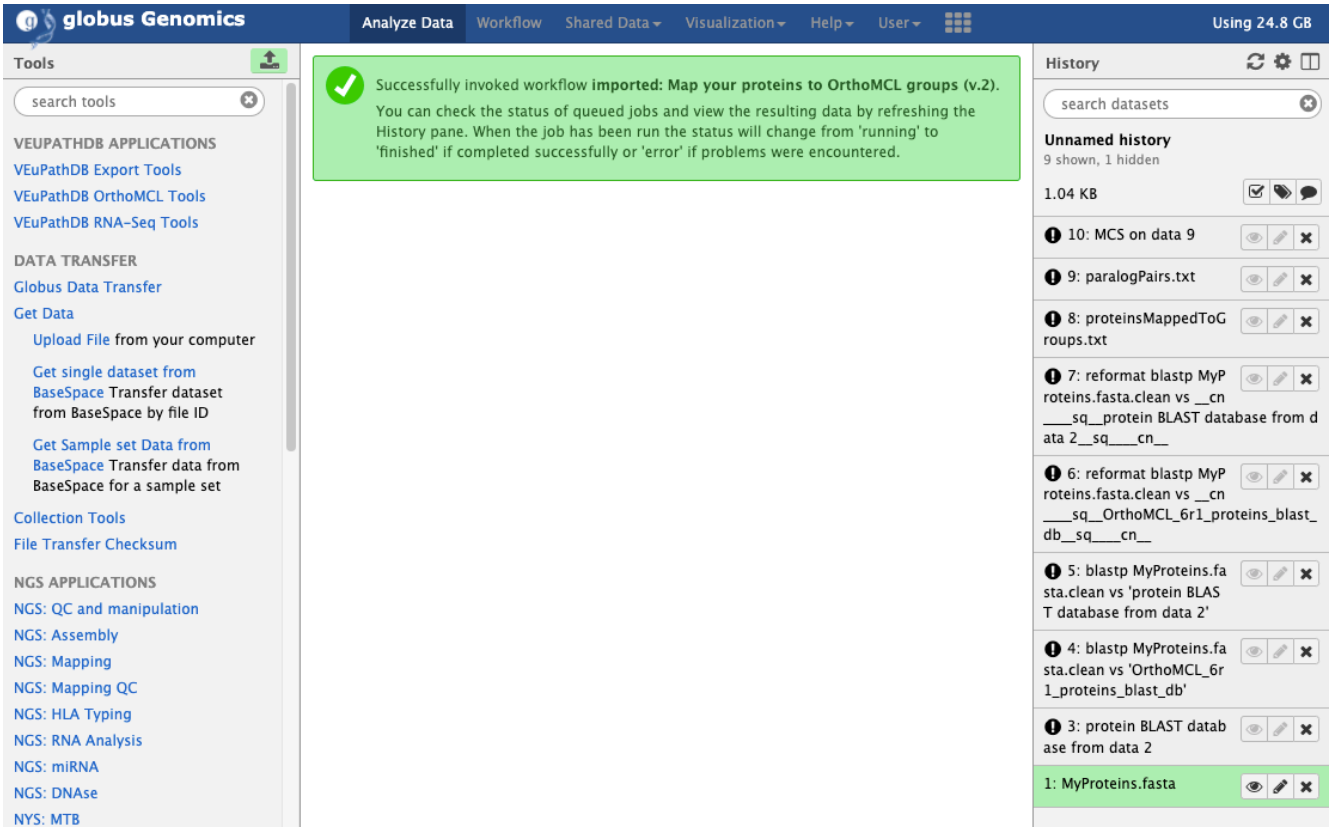




7. After setting up the workflow, press the “Run workflow” button at the top of the page. It may take a few minutes before the workflow starts to run.



8. Workflow steps will queue up in the right-hand panel. The entire workflow may take up to 24 hours or more to run, depending on the number of proteins in your FASTA file.



9. Once the workflow has completed, there are two important tab-delimited output files that can be downloaded to your computer: (a) proteinsMappedToGroups.txt and (b) MCS.
10. File (a) contains your proteins that mapped to OrthoMCL groups. The columns are: your_protein_id, orthomcl_group_id, most_similar_orthomcl_protein_id, evalue_mantissa, evalue_exponent, percent_identity, percent_match.
11. Those proteins that do not map to OrthoMCL group are processed by the OrthoMCL algorithm to determine whether there is any pairwise similarity which would indicate paralogous groups. Each row of file (b) contains the proteins that make up a paralogous group.