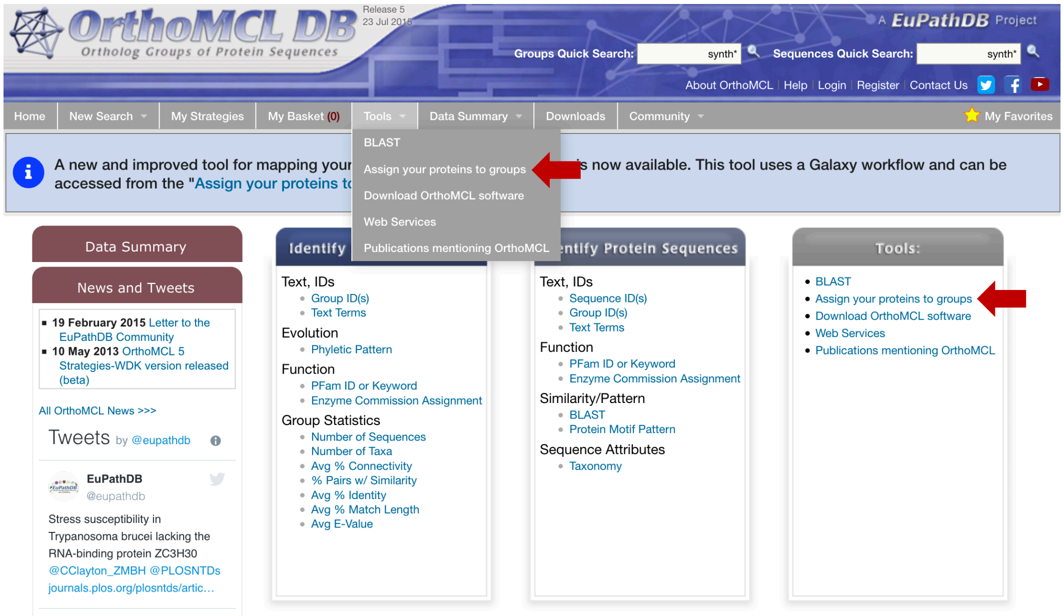
Map your proteins to OrthoMCL groups

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, you can follow these steps:

1. Click on the “Tools” item in the grey menu then select “Assign your proteins to groups” or click on the “Assign your proteins to groups” link from the right-hand side of the main OrthoMCL page.

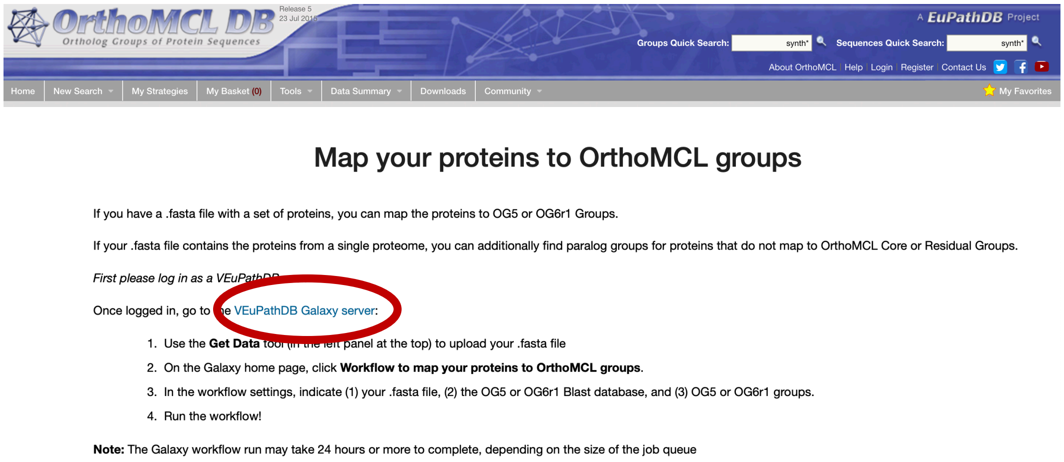


A new and improved tool for mapping your … s now available. This tool uses a Galaxy workflow and can be accessed from the "Assign your proteins t…

- **19 February 2015** Letter to the EuPathDB Community
- **10 May 2013** OrthoMCL 5 Strategies-WDK version released (beta)

Stress susceptibility in Trypanosoma brucei lacking the RNA-binding protein ZC3H30 @CClayton_ZMBH @PLOSNTDs journals.plos.org/plosntds/artic…

2. The next page provides some instruction and a link to the VEuPathDB Galaxy server. Click on the VEuPathDB Galaxy Server link to access this service. **Note:** to use this service you will have to create a VEuPathDB account.


OrthoMCL DB
Ortholog Groups of Protein Sequences
Release 5
23 Jul 2015
A EuPathDB Project
Groups Quick Search: synth*
Sequences Quick Search: synth*
About OrthoMCL | Help | Login | Register | Contact Us
Home
New Search
My Strategies
My Basket (0)
Tools
Data Summary
Downloads
Community
My Favorites

Map your proteins to OrthoMCL groups

If you have a .fasta file with a set of proteins, you can map the proteins to OG5 or OG6r1 Groups.

If your .fasta file contains the proteins from a single proteome, you can additionally find paralog groups for proteins that do not map to OrthoMCL Core or Residual Groups.

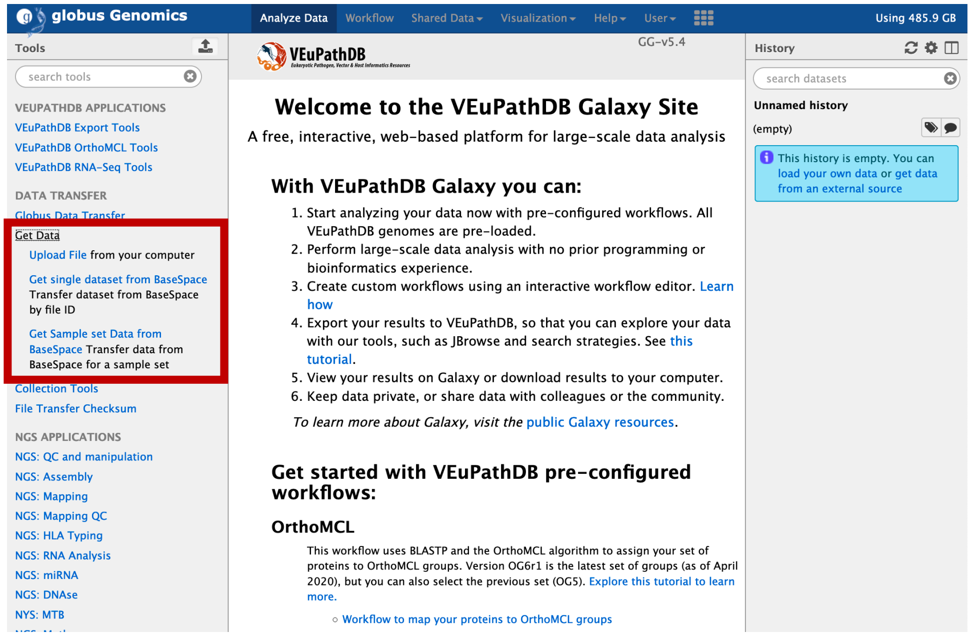
First please log in as a VEuPathDB

Once logged in, go to …e VEuPathDB Galaxy server:

1. Use the **Get Data** tool … panel at the top) to upload your .fasta file
2. On the Galaxy home page, click **Workflow to map your proteins to OrthoMCL groups**.
3. In the workflow settings, indicate (1) your .fasta file, (2) the OG5 or OG6r1 Blast database, and (3) OG5 or OG6r1 groups.
4. Run the workflow!

Note: The Galaxy workflow run may take 24 hours or more to complete, depending on the size of the job queue

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



VEUPATHDB APPLICATIONS
VEuPathDB Export Tools
VEuPathDB OrthoMCL Tools
VEuPathDB RNA-Seq Tools

DATA TRANSFER
Globus Data Transfer
Get Data
Upload File from your computer
Get single dataset from BaseSpace Transfer dataset from BaseSpace by file ID
Get Sample set Data from BaseSpace Transfer data from BaseSpace for a sample set
Collection Tools
File Transfer Checksum

NGS APPLICATIONS
NGS: QC and manipulation
NGS: Assembly
NGS: Mapping
NGS: Mapping QC
NGS: HLA Typing
NGS: RNA Analysis
NGS: miRNA
NGS: DNAse
NYS: MTB

Welcome to the VEuPathDB Galaxy Site

A free, interactive, web-based platform for large-scale data analysis

With VEuPathDB Galaxy you can:

1. Start analyzing your data now with pre-configured workflows. All VEuPathDB genomes are pre-loaded.
2. Perform large-scale data analysis with no prior programming or bioinformatics experience.
3. Create custom workflows using an interactive workflow editor. Learn how
4. Export your results to VEuPathDB, so that you can explore your data with our tools, such as JBrowse and search strategies. See this tutorial.
5. View your results on Galaxy or download results to your computer.
6. Keep data private, or share data with colleagues or the community.

To learn more about Galaxy, visit the public Galaxy resources.

Get started with VEuPathDB pre-configured workflows:

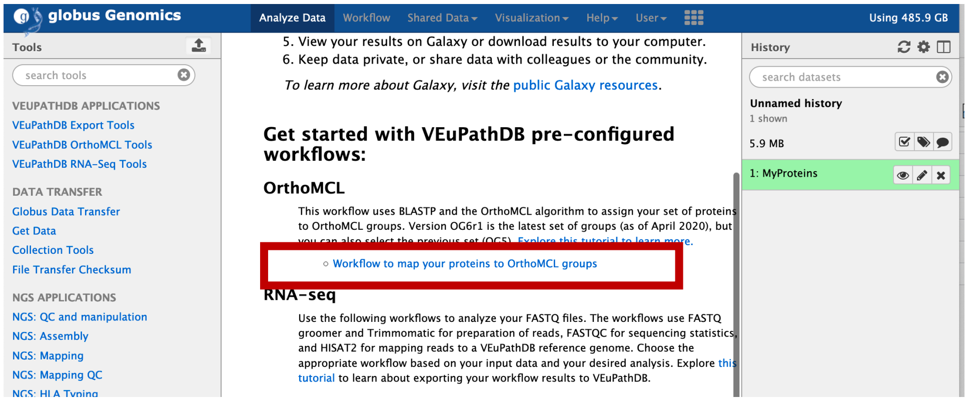
OrthoMCL

This workflow uses BLASTP and the OrthoMCL algorithm to assign your set of proteins to OrthoMCL groups. Version OG6r1 is the latest set of groups (as of April 2020), but you can also select the previous set (OG5). Explore this tutorial to learn more.

- Workflow to map your proteins to OrthoMCL groups

Unnamed history
(empty)
This history is empty. You can load your own data or get data from an external source

4. Your imported file will appear as a step in the right hand history pannel. The color of the step indicates its status. The step is grey when it is in queue, yellow when in process and green when completed.
5. Once your protein FASTA file has been successfully uploaded into Galaxy, select the “Workflow to map your proteins to OrthoMCL groups” from the middle section.



VEUPATHDB APPLICATIONS
VEuPathDB Export Tools
VEuPathDB OrthoMCL Tools
VEuPathDB RNA-Seq Tools

DATA TRANSFER
Globus Data Transfer
Get Data
Collection Tools
File Transfer Checksum

NGS APPLICATIONS
NGS: QC and manipulation
NGS: Assembly
NGS: Mapping
NGS: Mapping QC
NGS: HLA Typing

5. View your results on Galaxy or download results to your computer.
6. Keep data private, or share data with colleagues or the community.

To learn more about Galaxy, visit the public Galaxy resources.

Get started with VEuPathDB pre-configured workflows:

OrthoMCL

This workflow uses BLASTP and the OrthoMCL algorithm to assign your set of proteins to OrthoMCL groups. Version OG6r1 is the latest set of groups (as of April 2020), but you can also select the previous set (OG5). Explore this tutorial to learn more.

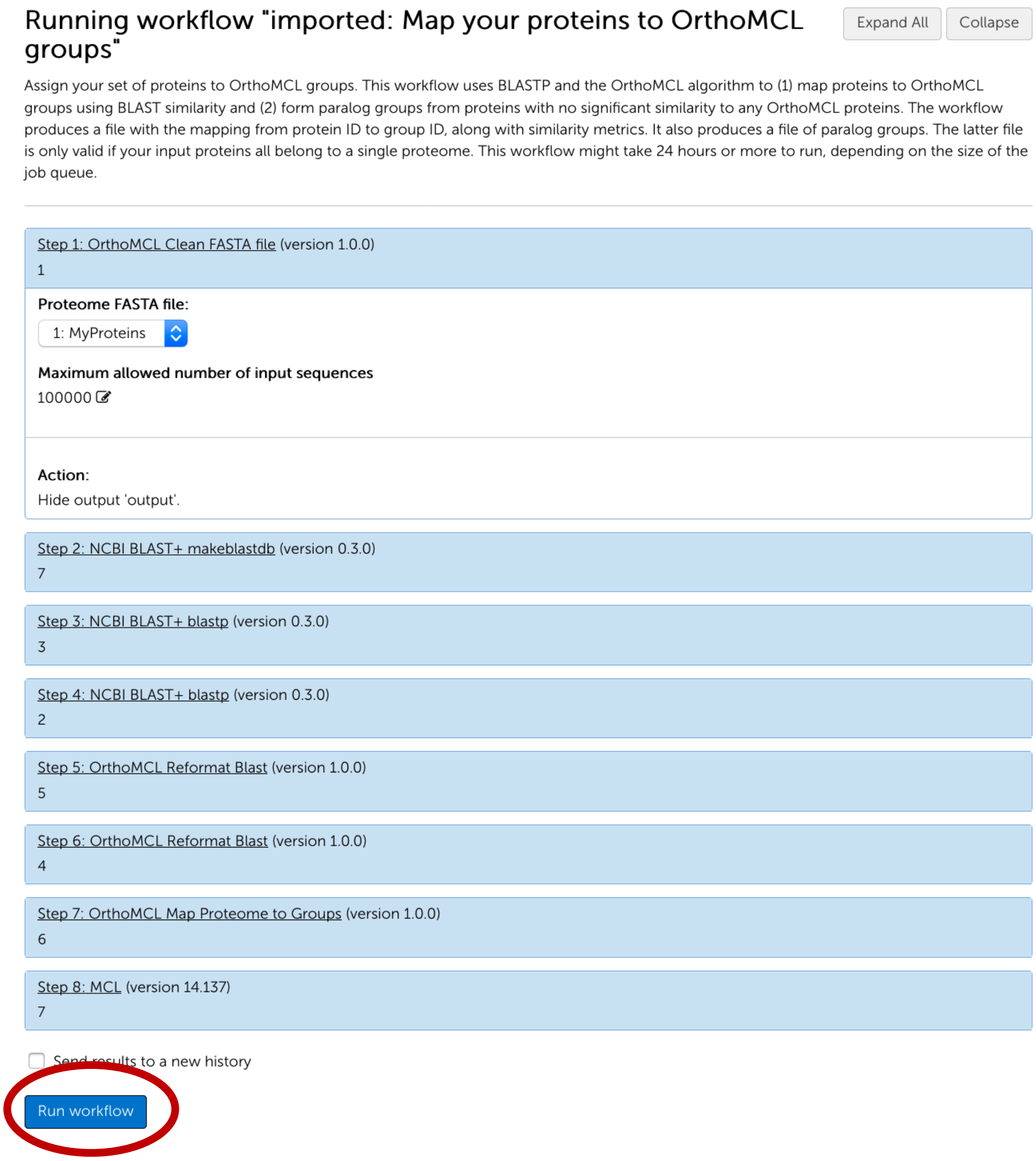
- Workflow to map your proteins to OrthoMCL groups

RNA-seq

Use the following workflows to analyze your FASTQ files. The workflows use FASTQ groomer and Trimmomatic for preparation of reads, FASTQC for sequencing statistics, and HISAT2 for mapping reads to a VEuPathDB reference genome. Choose the appropriate workflow based on your input data and your desired analysis. Explore this tutorial to learn about exporting your workflow results to VEuPathDB.

Unnamed history
1 shown
5.9 MB
1: MyProteins

6. This will import the workflow into your workspace and allow you to run it on your protein file. In most cases you will not need to change any of the default parameters. Simply click on the “Run workflow” button at the bottom of the middle section. Note that there is a few second lag between clicking on the button and the workflow starting to run – please be patient.



Running workflow "imported: Map your proteins to OrthoMCL groups"

Assign your set of proteins to OrthoMCL groups. This workflow uses BLASTP and the OrthoMCL algorithm to (1) map proteins to OrthoMCL groups using BLAST similarity and (2) form paralog groups from proteins with no significant similarity to any OrthoMCL proteins. The workflow produces a file with the mapping from protein ID to group ID, along with similarity metrics. It also produces a file of paralog groups. The latter file is only valid if your input proteins all belong to a single proteome. This workflow might take 24 hours or more to run, depending on the size of the job queue.

Step 1: OrthoMCL Clean FASTA file (version 1.0.0)
1

Proteome FASTA file:
1: MyProteins

Maximum allowed number of input sequences
100000

Action:
Hide output 'output'.

Step 2: NCBI BLAST+ makeblastdb (version 0.3.0)
7

Step 3: NCBI BLAST+ blastp (version 0.3.0)
3

Step 4: NCBI BLAST+ blastp (version 0.3.0)
2

Step 5: OrthoMCL Reformat Blast (version 1.0.0)
5

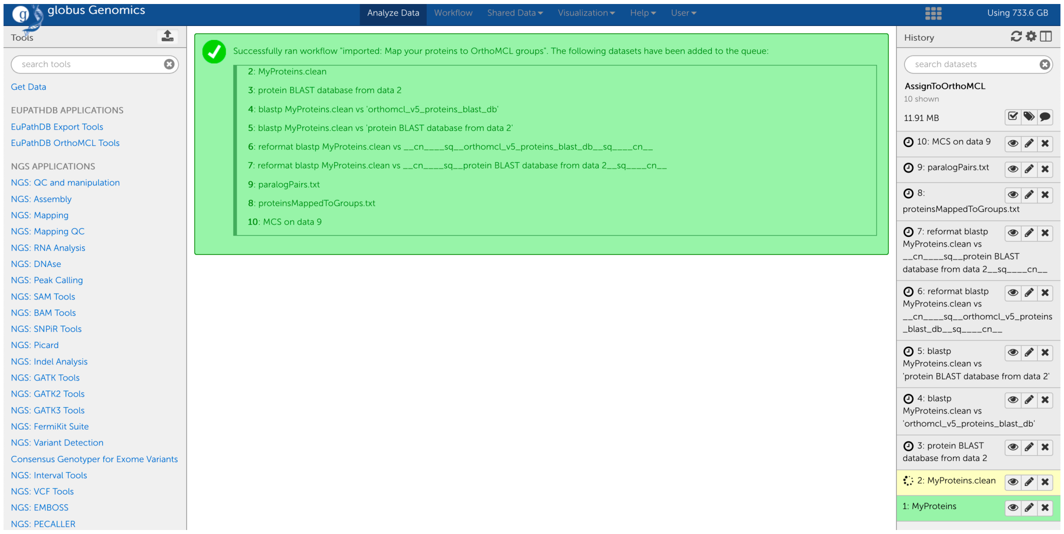
Step 6: OrthoMCL Reformat Blast (version 1.0.0)
4

Step 7: OrthoMCL Map Proteome to Groups (version 1.0.0)
6

Step 8: MCL (version 14.137)
7

☐ Send results to a new history

7. 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 size.



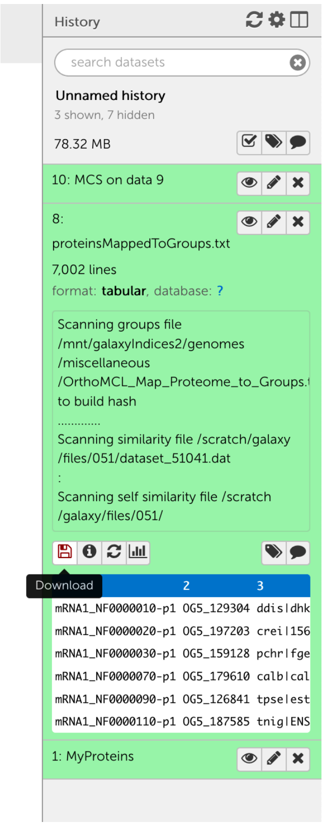
Successfully ran workflow "imported: Map your proteins to OrthoMCL groups". The following datasets have been added to the queue:

2: MyProteins.clean
3: protein BLAST database from data 2
4: blastp MyProteins.clean vs 'orthomcl_v5_proteins_blast_db'
5: blastp MyProteins.clean vs 'protein BLAST database from data 2'
6: reformat blastp MyProteins.clean vs __cn____sq__orthomcl_v5_proteins_blast_db__sq____cn__
7: reformat blastp MyProteins.clean vs __cn____sq__protein BLAST database from data 2__sq____cn__
9: paralogPairs.txt
8: proteinsMappedToGroups.txt
10: MCS on data 9

AssignToOrthoMCL
10 shown
11.91 MB

10: MCS on data 9
9: paralogPairs.txt
8: proteinsMappedToGroups.txt
7: reformat blastp MyProteins.clean vs __cn____sq__protein BLAST database from data 2__sq____cn__
6: reformat blastp MyProteins.clean vs __cn____sq__orthomcl_v5_proteins_blast_db__sq____cn__
5: blastp MyProteins.clean vs 'protein BLAST database from data 2'
4: blastp MyProteins.clean vs 'orthomcl_v5_proteins_blast_db'
3: protein BLAST database from data 2
2: MyProteins.clean
1: MyProteins

8. Once the workflow has completed, you can work with the output files. Click on the name of the step in your history in the right-hand panel to expand it. This provides additional options including the option to download file.



Unnamed history
3 shown, 7 hidden
78.32 MB

10: MCS on data 9

8: proteinsMappedToGroups.txt
7,002 lines
format: **tabular**, database: ?

```
Scanning groups file
/mnt/galaxyIndices2/genomes
/miscellaneous
/OrthoMCL_Map_Proteome_to_Groups.t
to build hash
............
Scanning similarity file /scratch/galaxy
/files/051/dataset_51041.dat
:
Scanning self similarity file /scratch
/galaxy/files/051/
```

```
                    2         3
mRNA1_NF0000010-p1 OG5_129304 ddis|dhk
mRNA1_NF0000020-p1 OG5_197203 crei|156
mRNA1_NF0000030-p1 OG5_159128 pchr|fge
mRNA1_NF0000070-p1 OG5_179610 calb|cal
mRNA1_NF0000090-p1 OG5_126841 tpse|est
mRNA1_NF0000110-p1 OG5_187585 tnig|ENS
```

1: MyProteins