

How to use this document

This document is a live copy of supplementary materials for the manuscript. It provides access to the exact analyses and workflows discussed in the paper, so you can play with them by re-running, changing parameters, or even applying them to your own data. Specifically, we provide the two histories and one workflow found below. You can view these items by clicking on their name to expand them. You can also import these items into your Galaxy workspace and start using them. To import workflows you must create a Galaxy account (unless you already have one) - a hassle-free procedure where you are only asked for a username and password.

This is the Galaxy history detailing the comparison of our pipeline to MEGAN.

Galaxy History | Galaxy vs MEGAN
Comparison of Galaxy vs. MEGAN pipeline.

8: s234 within 5% of max	This dataset contains blast hits +/- 5% of the maximum bitscore
9: Fetch taxonomic representation on data 7	Taxonomic representation for retained Sample 1 reads
10: Fetch taxonomic representation on data 8	Taxonomic representation for retained Sample 2 reads

```

1: line: 9: The 1st contig: Smp1contig1.kindred_0.01
2063077T Sample 1 Data S2063077T trimmed_to 17 1031496 root Eubaryota Fungi Dikarya n
2061667T Sample 1 Data S2061667T trimmed_to 25 991562 root Bacteria n n
2060227T Sample 1 Data S2060227T trimmed_to 196 929562 root Bacteria n n
2060749T Sample 1 Data S2060749T trimmed_to 64 991562 root Bacteria n n
2060229T Sample 1 Data S2060229T trimmed_to 77 992562 root Bacteria n n
20637549T Sample 1 Data S20637549T trimmed_to 63 985562 root Bacteria n n
            
```

This is the Galaxy history showing a generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and Figure 3A):

Galaxy History | metagenomic analysis

This is the Galaxy workflow for generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and Figure 3B):

Galaxy Workflow | metagenomic analysis
Generic workflow for performing a metagenomic analysis on NGS data.

Supplemental Analysis

Comparison between Galaxy pipeline and Megan

Go to "<http://main.g2.bx.psu.edu/user/create>"

<http://main.g2.bx.psu.edu/datasets/2be89fa329bc8b6/display/?review=1>

This dataset is large and only the first megabyte is shown below.
[Show all | Save](#)

S2063077T	Sample 1	Data	S2063077T	trimmed_to 17 1031496	root	Eubaryota	Fungi	Dikarya	n	n	Ascomycota
S2061667T	Sample 1	Data	S2061667T	trimmed_to 25 991562	root	Bacteria	n	n	n	n	Proteobacteria
S2060227T	Sample 1	Data	S2060227T	trimmed_to 196 929562	root	Bacteria	n	n	n	n	Proteobacteria
S2060749T	Sample 1	Data	S2060749T	trimmed_to 64 991562	root	Bacteria	n	n	n	n	Proteobacteria
S2060229T	Sample 1	Data	S2060229T	trimmed_to 77 992562	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 63 985562	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 27 9971061	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 61 9814539	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 18 1014292	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 24 876799	root	Bacteria	n	n	n	n	Bacteroidetes/Chlorobi
S20637549T	Sample 1	Data	S20637549T	trimmed_to 43 923663	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 19 973294	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 80 923294	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 19 807294	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 18 1014311	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 14 10061938	root	Bacteria	n	n	n	n	Actinobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 16 1020288	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 25 881106799	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 16 1020172049	root	Bacteria	n	n	n	n	Bacteroidetes/Chlorobi
S20637549T	Sample 1	Data	S20637549T	trimmed_to 17 994188	root	Bacteria	n	n	n	n	Proteobacteria
S20637549T	Sample 1	Data	S20637549T	trimmed_to 17 994387	root	Bacteria	n	n	n	n	Proteobacteria

<http://main.g2.bx.psu.edu/root>

Galaxy Analyze Data Workflow Data Libraries Lab Admin Help User

Tools

Get Data
Send Data
FNCODE Tools
ML-Over
Text Manipulation
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Display menu

History

Options

Imported: Galaxy vs MEGAN

- 14: Draw phylogeny on data 12
- 13: Draw phylogeny on data 11
- 12: Find lowest diagnostic rank on data 10
- 11: Find lowest diagnostic rank on data 9
- 10: Fetch taxonomic representation on data 8

<http://main.g2.bx.psu.edu/workflow/editor?id=6ad322b4c4ff7f6c>

Galaxy Analyze Data Workflow Data Libraries Lab Admin Help User

Workflow Canvas | Clone of 'metagenomic analysis' shared by 'anton@bx.psu.edu'

```

graph TD
    A[Input dataset] --> B[Select high quality segments]
    B --> C[Reads]
    C --> D[Quality scores]
    D --> E[output (fasta)]
    F[Input dataset] --> G[FASTA-to-Tabular]
    G --> H[Convert these sequences]
    H --> I[output (tabular)]
    I --> J[Megablast]
    J --> K[Compare these sequences]
    K --> L[output (tabular)]
    L --> M[Concatenate queries]
    M --> N[Concatenate Query]
    
```