

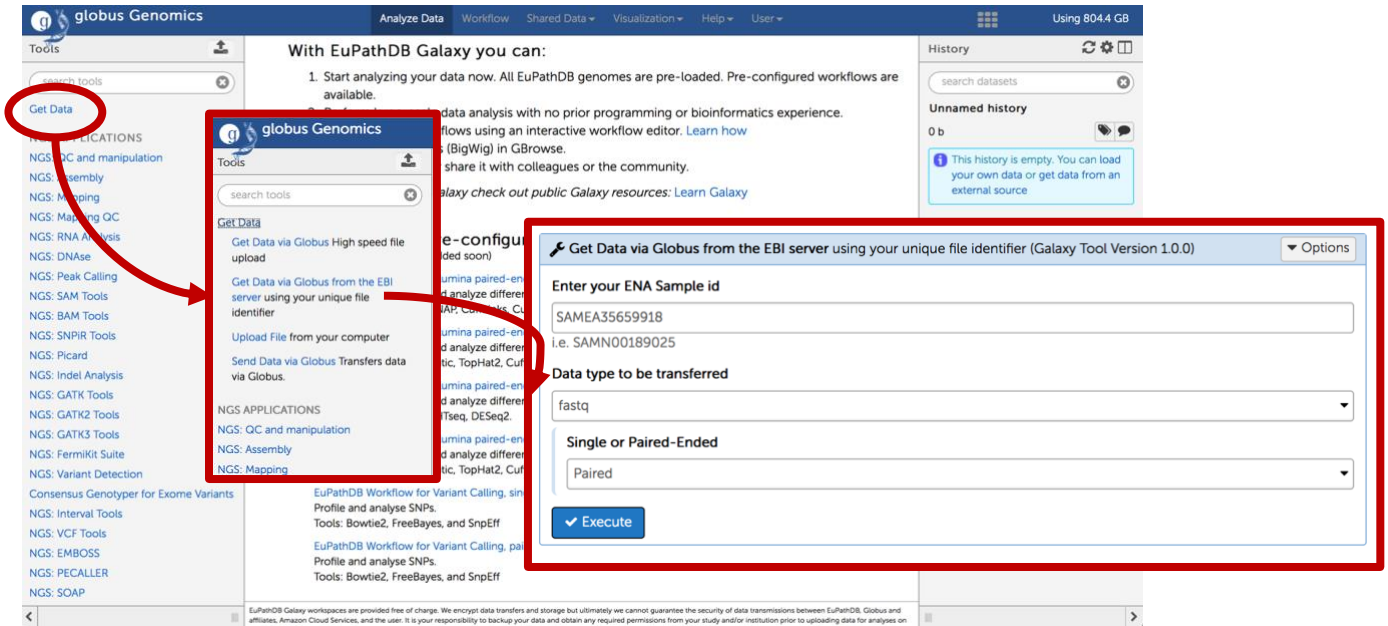
Variant Calling using EuPathDB Galaxy

In this exercise we will work in groups to retrieve DNA sequence data from the sequence repository and analyze it for variants using a workflow in EuPathDB Galaxy. For this workshop we will use the workshop specific galaxy site:

<https://eupathdbworkshop.globusgenomics.org/>

There are different ways to get data into Galaxy. Here we will use the sample ID and get the data using the “Get Data via Globus from the EBI server using your unique file identifier” link. Follow these steps:

1. Click on the “Get Data” link.
2. Click on the “Get Data via Globus from the EBI server” link.
3. The next window allows you to enter the sample ID. This ID starts with the letters ‘SAM’. Choose the sample ID for your group from the list below and use it in this form. **Note:** it is very important that you select whether the data is single or paired-end.
4. Once the form is properly filled, click on the ‘Execute’ button to start the data transfer process.



Groups:

Group 1: *Plasmodium berghei* wild type

Sample ID: SAMN04386828

<https://www.ebi.ac.uk/ena/data/view/SAMN04386828>

Group 2: *Plasmodium berghei* drug resistant mutant

Sample ID: SAMN04386825

<https://www.ebi.ac.uk/ena/data/view/SAMN04386825>

Group 3: *Cryptosporidium* field isolate (clinic visit sample)

Sample ID: SAMEA104459068

<https://www.ebi.ac.uk/ena/data/view/SAMEA104459068>

Group 4: *Cryptosporidium* field isolate (Diarrheal sample)

Sample ID: SAMEA104459070

<https://www.ebi.ac.uk/ena/data/view/SAMEA104459070>

Group 5: *Toxoplasma gondii* RH parental strain (type I strain)

Sample ID: SAMN06112744

<http://www.ebi.ac.uk/ena/data/view/SAMN06112744>

Group 6: *Toxoplasma gondii* RH IBET-151 resistant mutant (type I strain)

Sample ID: SAMN06112745

<http://www.ebi.ac.uk/ena/data/view/SAMN06112745>

The screenshot displays the Globus Genomics web interface. At the top, the navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The main content area features a green notification box with a checkmark icon, stating: '1 job has been successfully added to the queue - resulting in the following datasets: 1: ERR1767828.fastq.gz 2: ERR1767828_1.fastq.gz'. Below this, a message reads: 'You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the left, the 'Tools' sidebar lists various NGS applications such as 'NGS: QC and manipulation', 'NGS: Assembly', and 'NGS: Mapping'. On the right, the 'History' panel shows a search bar and a list of datasets under the heading 'Unnamed history', with 2 items shown: 'ERR1767828_1.fastq.gz' and 'ERR1767828.fastq.gz'. The top right corner indicates 'Using 710.1 GB' of storage.

Running a variant calling workflow:

- Use the EuPathDB preconfigured workflow from the Galaxy home page. The workflow you want to use is called “
- Remember to select the correct reference genome. Check with the other group using a sample from the same experiment and make sure you both agree on which reference genome to use (Bowtie2, FreeBayes, SnpEff)
- Click on the ‘Run Workflow’ button.

Workflow: **EuPathDB_Workshop_VariantCalling_PairedEnd** Run workflow

Hide output 'text_file'. Hide output 'html_file'.

6: Bowtie2 - 4 (Galaxy Version BOWTIE2: 2.1.0; SAMTOOLS: 1.2)

Is this library mate-paired?

Paired-end

FASTQ file

Output dataset 'outdir_naired1' from stan.4

- AmoebaDB-29_EinvadensIP1_Genome
- AmoebaDB-29_Emoshkovskil.aredo_Genome
- AmoebaDB-29_EnuttaIP19_Genome
- AmoebaDB-29_HowlerITCC30863_Genome
- CryptoDB-29_CbaileyTAMU-09Q1_Genome
- CryptoDB-29_Chominis37999_Genome
- CryptoDB-29_ChominisTU502_Genome
- CryptoDB-29_ChominisTU502_new_Genome
- CryptoDB-29_ChominisUKH1_Genome
- CryptoDB-29_CmurisRN66_Genome
- CryptoDB-29_CvellaCCMP2878_Genome
- CryptoDB-29_GniphandrosesUnknown_Genome
- CryptoDB-29_VbrassicaformisCCMP3155_Genome
- CryptoDB-30_ChominisUdeA01_Genome
- CryptoDB-31_Cubiquitum39728_Genome
- CryptoDB-32_ChominisTU502_2012_Genome
- CryptoDB-36_Chominis30976_Genome
- CryptoDB-36_CtyzzeriUGA55_Genome
- CryptoDB-37_CmelegridsUKMEL1_Genome
- CryptoDB-37_Csairumhwal_Genome**

If your genome of interest is not listed, contact the Galaxy team

Parameter Settings

Full parameter list

Type of alignment

End to end

RNA Sequencing

[EuPathDB RNA-Seq paired-end: for RNAseq Export Tool](#)

This workflow generates BigWig and Expression files that are compatible with the RNA-Seq Export Tool. [Explore this tutorial to learn more.](#)

Tools: FASTQ Groomer, Trimmomatic, HISAT2, Cufflinks, BAM to BigWig

[EuPathDB Workflow for Illumina paired-end RNA-seq, without replicates](#)

Profile a transcriptome and analyze differential gene expression.

Tools: FastQC, Sickle, GSNAP, CuffLinks, CuffDiff.

[EuPathDB Workflow for Illumina paired-end RNA-seq, without replicates](#)

Profile a transcriptome and analyze differential gene expression.

Tools: FastQC, Trimmomatic, TopHat2, CuffLinks, CuffDiff.

[EuPathDB Workflow for Illumina paired-end RNA-seq, biological replicates](#)

Profile a transcriptome and analyze differential gene expression.

Tools: FastQC, Trimmomatic, TopHat2, HTseq, DESeq2.

[EuPathDB Workflow for Illumina paired-end RNA-seq, biological replicates](#)

Profile a transcriptome and analyze differential gene expression.

Tools: FastQC, Trimmomatic, TopHat2, CuffLinks, CuffDiff.

[EuPathDB Workflow for Illumina paired-end RNA-seq, biological replicates](#)

Profile a transcriptome and analyze differential gene expression.

Tools: Collections, FastQC, Trimmomatic, HISAT2, HTseq, DESeq2.

Variant Calling

[EuPathDB Workflow for Variant Calling, single-read sequencing](#)

Profile and analyse SNPs.

Tools: Sickle, Bowtie2, FreeBayes, and SnpEff

[EuPathDB Workflow for Variant Calling, paired-end sequencing](#)

Profile and analyse SNPs.

Tools: Sickle, Bowtie2, FreeBayes, SnpEff and SnpSift

Click to run the workflow