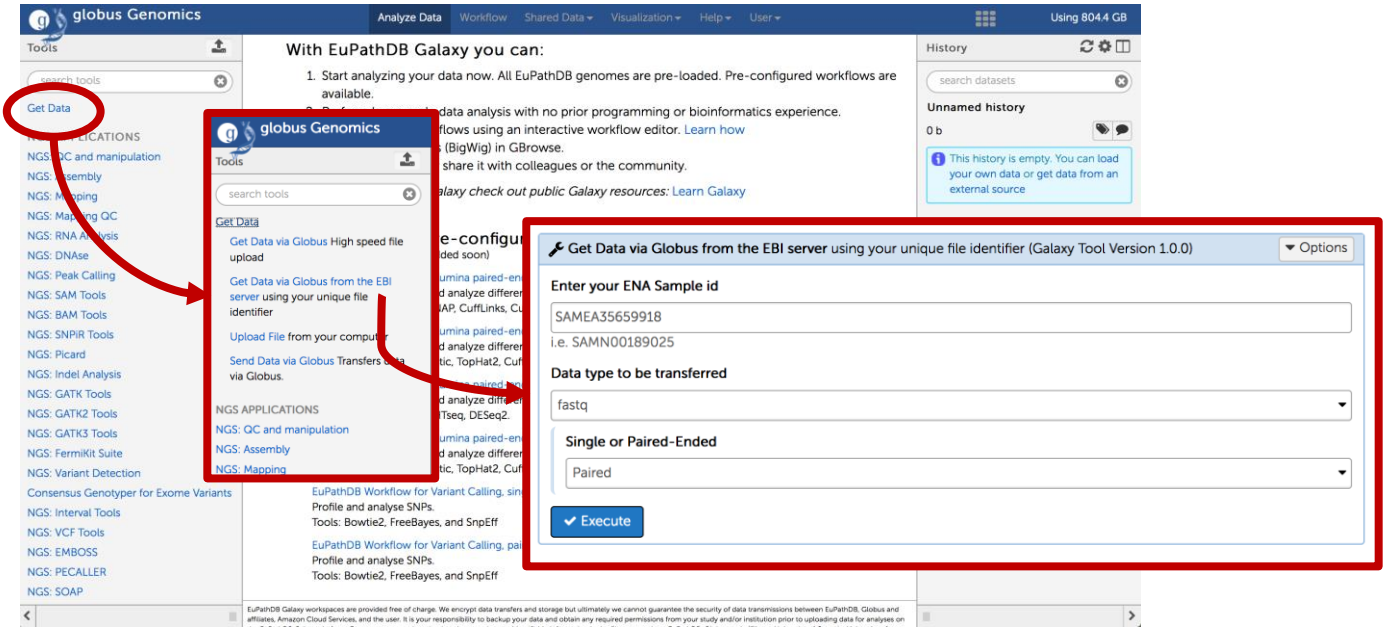
Variant Calling using EuPathDB Galaxy

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

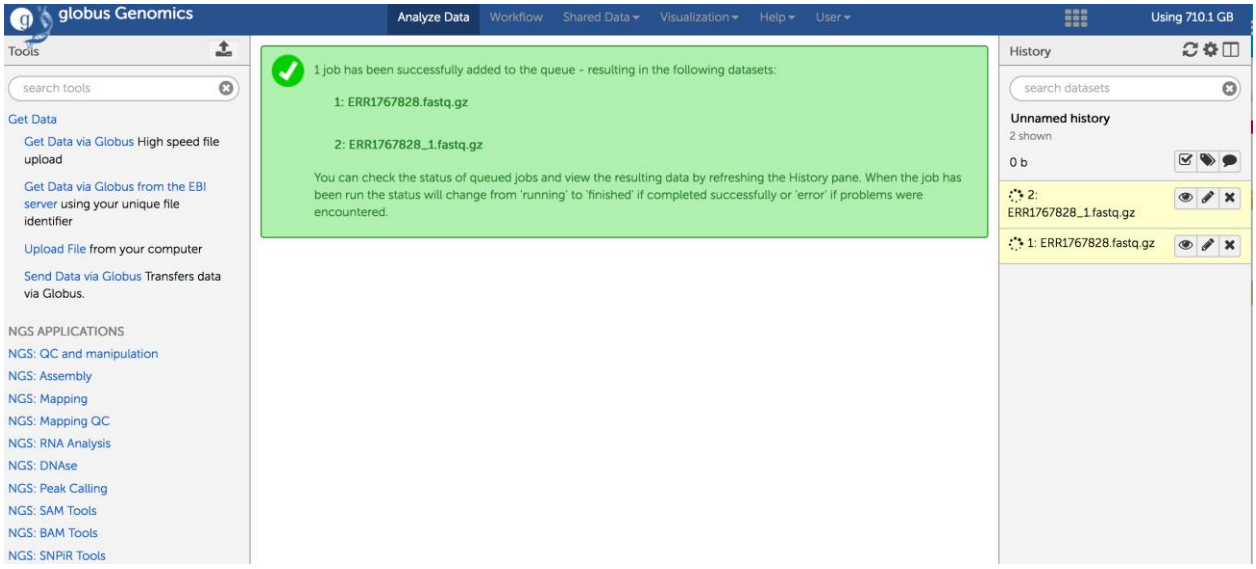
Group 2: *Plasmodium berghei* drug resistant mutant
Sample ID: SAMN04386825
https://www.ebi.ac.uk/ena/data/view/SAMN04386825

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

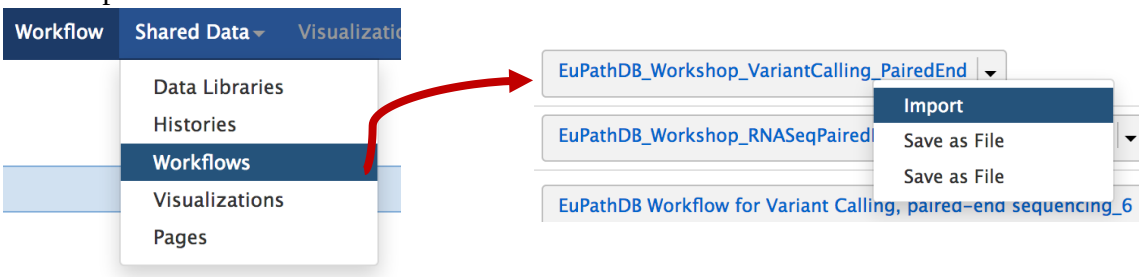
Group 5: *Toxoplasma gondii* RH parental strain (type I strain)
Sample ID: SAMN06112744
http://www.ebi.ac.uk/ena/data/view/SAMN06112744

Group 6: *Toxoplasma gondii* RH IBET-151 resistant mutant (type I strain)
Sample ID: SAMN06112745
http://www.ebi.ac.uk/ena/data/view/SAMN06112745

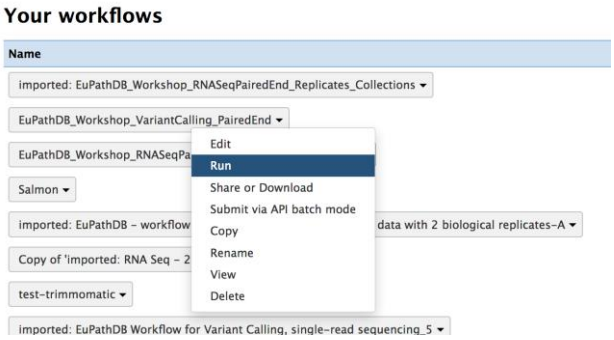


Running a variant calling workflow:

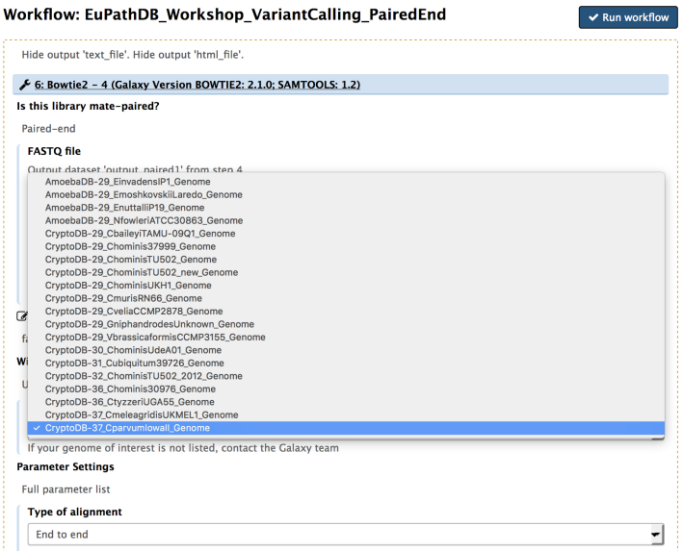
a. Use the shared workflow called: EuPathDB_Workshop_VariantCalling_PairedEnd. To use a shared workflow, first you must import it by going to 'workflows' under the shared data tab, clicking on the desired workflow an d selecting 'import' from the dropdown menu.



b. To run this workflow, go to the workflow section, click on the workflow with the correct name and select "Run"



c. 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)



d. Click on the 'Run Workflow' button.