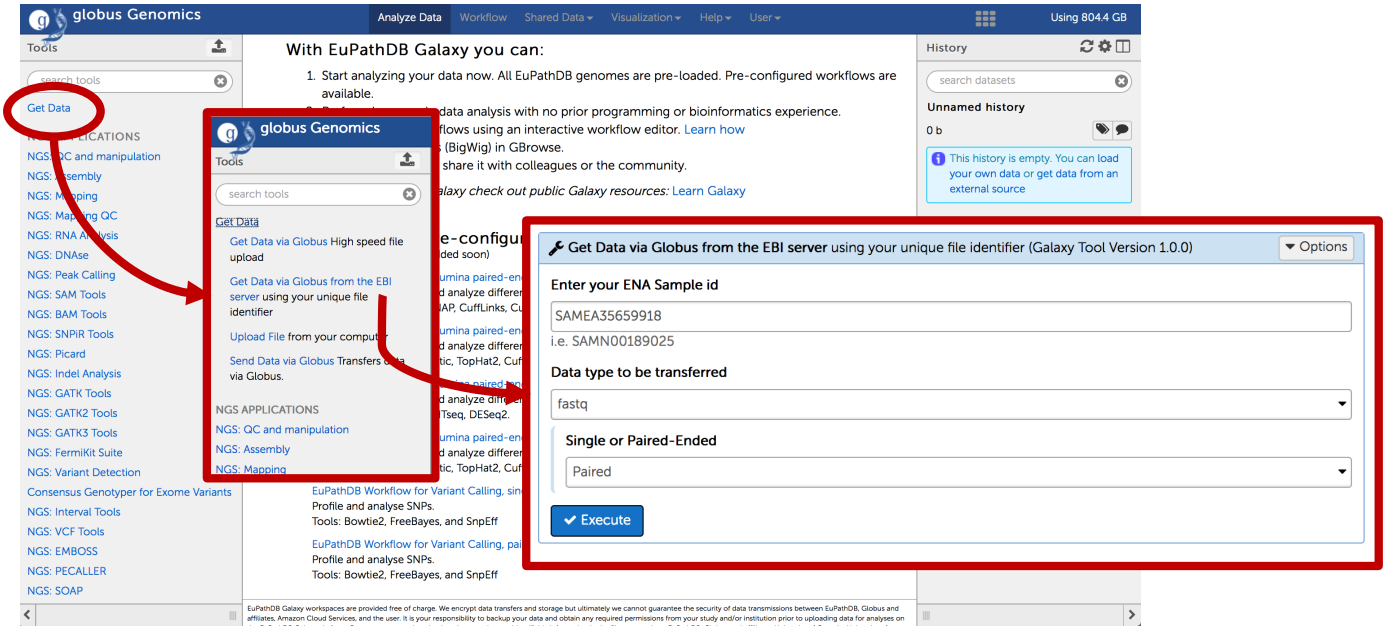
Variant Calling using EuPathDB Galaxy

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 falciparum* drug resistant field isolate
Sample ID: SAMN01087919
http://www.ebi.ac.uk/ena/data/view/SAMN01087919

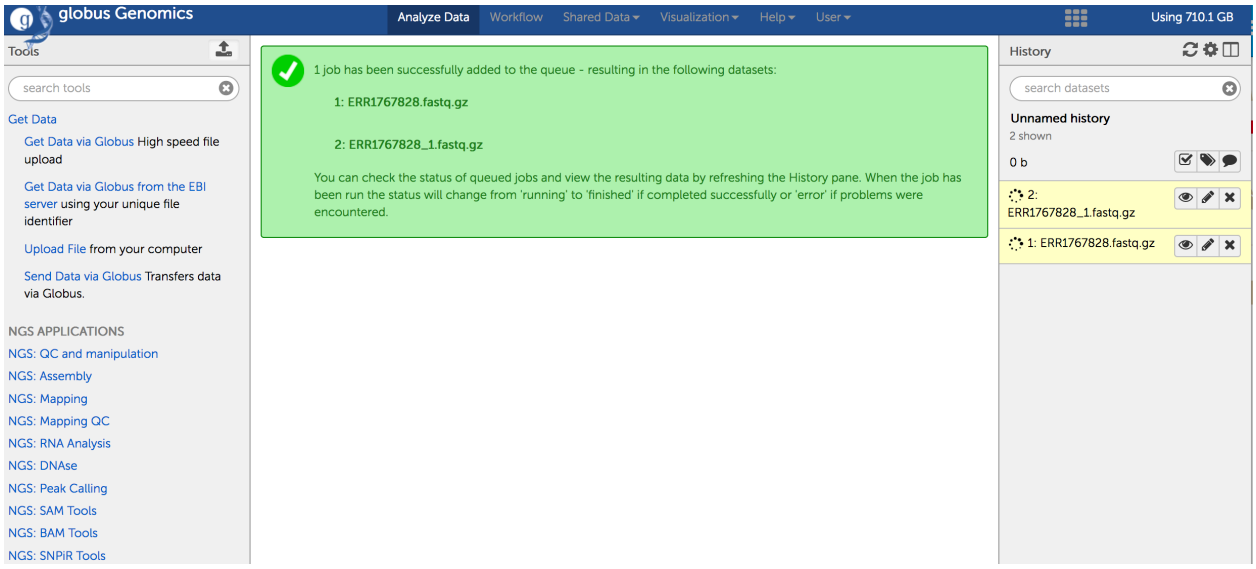
Group 2: *Babesia microti* field isolate (Rhode Island)
Sample ID: SAMEA3918179
http://www.ebi.ac.uk/ena/data/view/SAMEA3918179

Group 3: *Babesia microti* field isolate (Wisconsin)
Sample ID: SAMEA3918185
http://www.ebi.ac.uk/ena/data/view/SAMEA3918185

Group 4: *Candida albicans* CHN1
Sample ID: SAMN00974105
http://www.ebi.ac.uk/ena/data/view/SAMN00974105

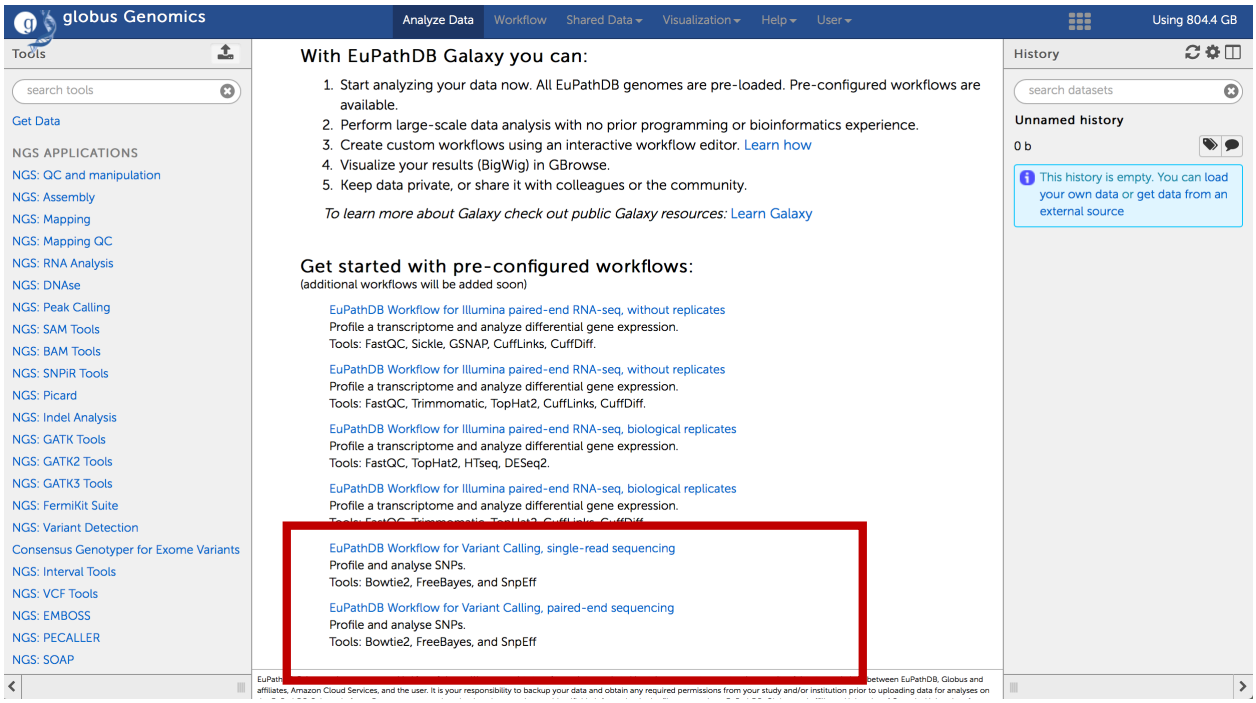
Group 5: *Toxoplasma gondii* RH parental strain (type I strain)
Sample ID: SAMN06112744
http://www.ebi.ac.uk/ena/data/view/SAMN06112744

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:

- Once the data files have been transferred into your galaxy history you need to choose an appropriate workflow. EuPathDB provides some preconfigured workflows on the EuPathDB Galaxy instance home page.
- Remember to choose the appropriate workflow – Single ended or paired ended.



- Set workflow parameters. **Note that the trimming step "Sickle" has a parameter to select the "quality type". The default is often "Illumina". This will not work and has to be changed to Sanger.**
- Select the correct reference genome (Bowtie2, FreeBayes, SnpEff)
- Click on the 'Run Workflow' button.