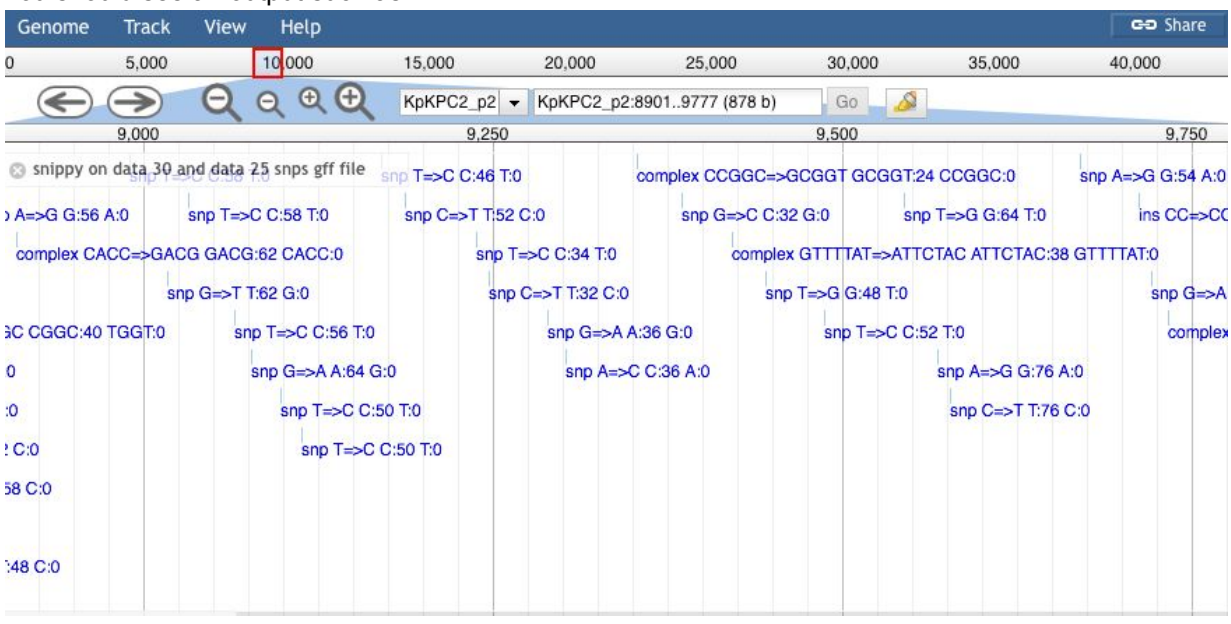
omics.data.edu.au

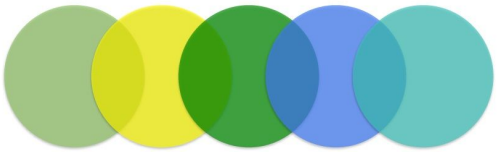
# How to identify genomic variation between a bacterial strain and a reference genome of that species

| STEP | Task |
|---|---|
| 1 | Go to http://omics.data.edu.au/use/ |
| 2 | Launch the platform and log in |
| 4 | Add the “ABPRI-Data” and “Galaxy-QLD” apps to the toolbar. |
| 5 | Open the “ABPRI-Data” app. |
| 6 | Log-into the “ABPRI-Data” app using the username and password given to you (you will need to enter **omics** as the domain). |
| Steps 7-10 are to obtain an already assembled genome from the ABPRI consortium that will be used as the ‘reference’ in this exercise: | |
| 7 | Search for “processed” “genomic” data from “Klebsiella pneumoniae”. |
| 8 | Send the resultant data files to the GenomeSpace part of the OMICs platform. |
| 9 | Locate the data you just sent to GenomeSpace. |
| 10 | Drag the 3 fasta (`.fsa`) genomic assembly files onto the “Galaxy-QLD” app *(you added this app to the toolbar in Step 4)*. You will be asked to log into Galaxy. Use the username and password given to you. |
| Steps 11-13 are to obtain some NGS read data from the ABPRI consortium that will be used as ‘sample’ data in this usability test (i.e. raw illumina data derived from another sample of the same species). | |
| 11 | Search for “raw” “genomic” data from “Klebsiella pneumoniae” sample ID “25730” that was generated on an Illumina “MiSeq” instrument. |
| 12 | Select and drag the `*_R1.fastq.gz` and `*_R2.fastq.gz` files onto the analysis tool ABPRI-Galaxy. |
| 13 | For each, specify its data type as **fastqsanger** (instead of fastq which it will be defined as by default). To do this, click on the pencil icon, locate the Datatype tab and then select “fastqsanger” from the list. |

omics.data.edu.au

| | |
|---|---|
| Steps 14-16 outline the variant analysis between the 'reference' genome and the 'sample' paired read raw data files. | |
| 14 | Now look for variants in the 26730 sample compared to the genome we are using as a reference ( i.e. sample 25728), and generate a table listing the variants.<br><br>Tutorials describing how to undertake tasks and use tools within the OMICS platform and Galaxy-QLD are found under the Training menu link on this page: http://omics.data.edu.au<br><br>Hint: From the training website menus, select **Genomics > Variant finding**.<br><br>Note that some of the tutorial content specifically refers to a tutorial training dataset and not the data used in this example, so you will need to modify the steps as outlined in the tutorial material.<br><br>Note also that performing the variant calling on the larger chromosome (chr) file will take much longer than on either of the smaller plasmid (p1, p2) files. |
| 15 | Download the table describing the variants |
| 16 | Now visually examine the outputs from Snippy using the JBrowse tool within Galaxy-QLD.<br><br>Hint: When following the tutorial, there are two version of JBrowse available in the Galaxy-QLD - choose Galaxy Version 0.5.2. Also only follow the steps to (a) read in your fasta (.fsa) sequence as the 'reference', and to (b) add one track for the variants (where you can use the `.gff` snippy output file as input) |
| You should see an output such as:<br>  | |

If you need further assistance, please contact us at omicsdataservices@lists.unimelb.edu.au



omics.data.edu.au