RicePilaf hands-on

List of QTL regions for analysis will be from the paper : Genotyping-by-sequencing based QTL mapping for rice grain yield under reproductive stage drought stress tolerance - https://doi.org/10.1038/s41598-019-50880-z (Yadav et al.)

- Table 5 has a list of QTLs. Column 3 (**Stable QTLs**) with 5 QTLs, with 2 intervals per stable QTL. **Marker interval** column states the chromosome and bp start-end of a marker interval, the coordinates of which are based on the Nipponbare reference genome.
- As example, qDTY 2.4 marker interval **S2_17630922–S2_17731936** is from (start) chromosome 2 position 17,630,922 to (end) chromosome 2 position 17,731,936 in Nipponbare.
- To use this interval on RicePilaf as a query , the marker interval should be written as **Chr02:17630922-17731936** .

General hands-on flow

A - Select a marker interval from the **stable QTLs** only, as listed in the paper. **DO NOT COPY-PASTE** directly from the paper, it likely will cause an error due to hidden or weird characters from the pdf.

B - Run an end-to-end analysis of the QTL in RicePilaf for these tools.

1. **Gene List and lift-over**: include **1 genome** from the selection (your decision). Inspect the output tables (All genes, common genes, nipponbare, unique to your chosen genome). **Download** the **csv** file for the tab "**unique to *your chosen genome***".
2. **Gene retrieval by text mining**: Use any of these three as key search phrase: "days to flowering" , "plant height", "grain yield". Inspect the output table, **download as a csv file**.
3. **Coexpression network analysis**: Inspect the **top 2 modules** and the connectivity of the GWAS genes with other genes in the network.
    a. Select different coexpression networks and module detection algorithms until you get an enriched module result. Inspect the graphs for the top 2 modules.
    b. For the **top module**, generate the figure of a graph network showing the connection (edges) to other genes of ONE GWAS-region gene.
    Capture this graph and **save it as an image file**.

4. **Regulatory feature enrichment** : NOTE that if your interval is too large (i.e. in megabases), results may not be generated. Consider chopping up a long interval to multiple 500kb sized overlapping intervals.
   Inspect the output table of the enrichment analysis. **Download** the output CSV file

5. **Epigenomic information** : generate the IGV genome browser view and inspect the output.
   Select different tissues and note the active genome regions (ATAC-Seq and FAIRE-Seq tracks have NO track data), and inactive regions (ATAC-Seq and FAIRE-Seq tracks have track data) based on the results in the genome browser.
   Choose an IGV you wish to submit, right click on the image, and **save as an image file** (either PNG or SVG, PNG is preferred for this exercise).

C - Download the output **summary** in csv format.

E: Submit the following as a compressed file bundle (ZIP/RAR/7Z/GZ or whatever you are familiar with), containing the following:

1 - In a separate document, a table that mentions the tool settings used in your analysis for each tool. Example , for gene list and lift-over if you used IR64 as genome for lift-over, mention it explicitly. Each selectable parameter of the tool should be explicitly mentioned. Use this table as a guide/template to follow:

Tool	**Parameters used** *(example only, use your own parameters)*
Interval	chr01:10000-600000
Gene List and lift-over	Additional genome(s): IR64
Gene retrieval by text mining	Keyphrase: days to flowering filter the results to display only genes overlapping your input intervals: NO
Coexpression network analysis	coexpression network: RCRN module detection algorithm: COACH parameter for running the algorithm: 2
Regulatory Feature Enrichment	transcription factor binding site prediction technique: MotifScan transcription factor binding sites in the following regions: Promoters
Epigenomic Information	Tissue: leaf tracks to be displayed: FAIRE-Seq

2 - CSV output of Gene List and lift-over for list of genes **unique to** {*your chosen additional genome}*.
3- CSV - text mining
4 - image file of chosen network graph (jpg/png)
5 - CSV of Regulatory feature enrichment output
6 - image file of chosen Epigenomic Information output (png)
7- CSV, summary output

Further exploration

As an independent exercise, explore the tool using different QTL intervals and using other options in each tool (e.g. different reference genomes to include in Gene List and liftover, different parameters / algorithm selections in other tools).