

Post-GWAS using CropGalaxy

QTL of Interest: [qDTF2.2](#)

See the workflow here:

<http://cropgalaxy.excellenceinbreeding.org/u/vjuanillas/h/postgwas-hands-on>

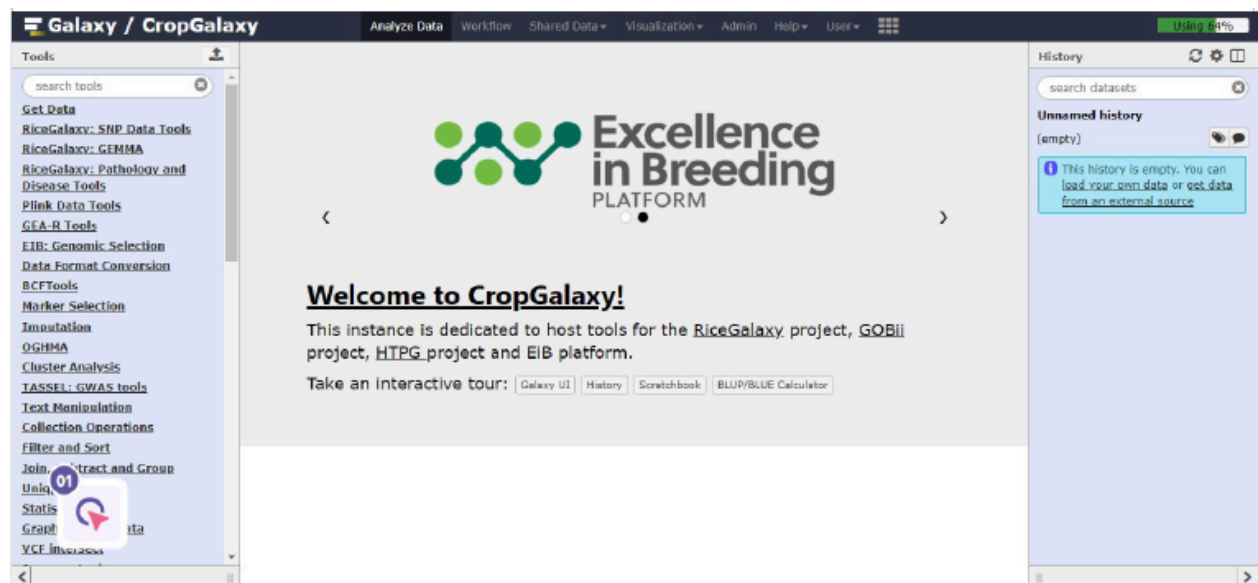
Part 1: Getting the candidate genes from a region of interest

1.) Go to <http://cropgalaxy.excellenceinbreeding.org>

On your browser, type “<http://cropgalaxy.excellenceinbreeding.org>”. This will bring you to the CropGalaxy landing page.

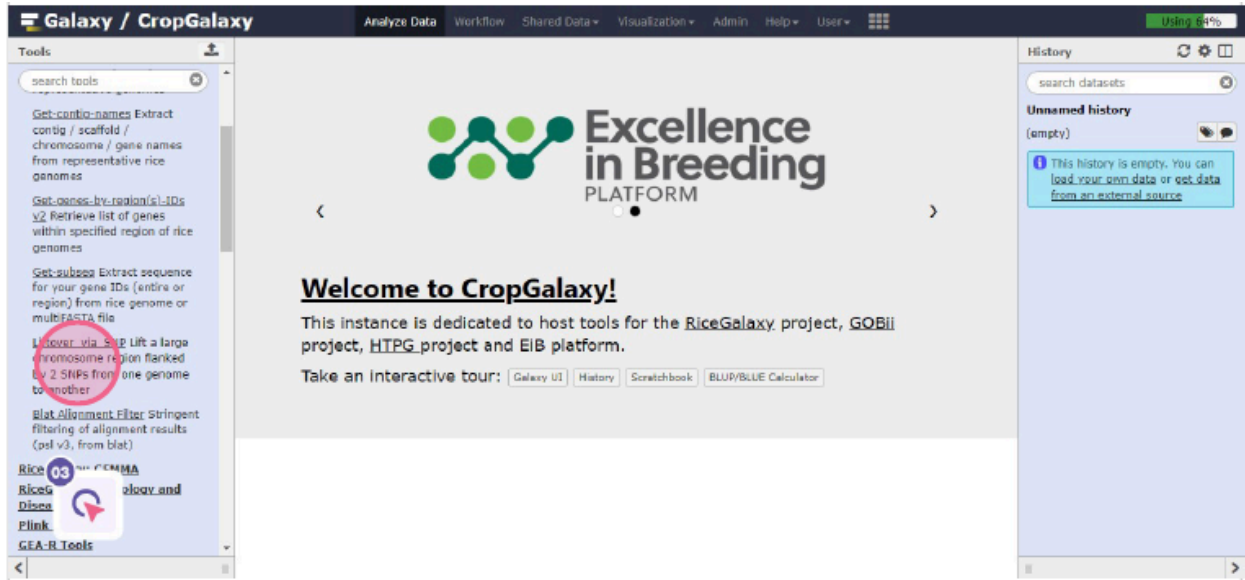
You will be greeted by the CropGalaxy landing page where you will see three main panels:

- ☐ The leftmost panel contains all the tools available for you to use.
- ☐ The rightmost panel displays the current history of the analysis. A history consists of datasets derived from each time a tool is run.
- ☐ The middle panel serves as an analysis panel which displays the tool parameters, as well as the contents of a dataset when the user selects a dataset to view.



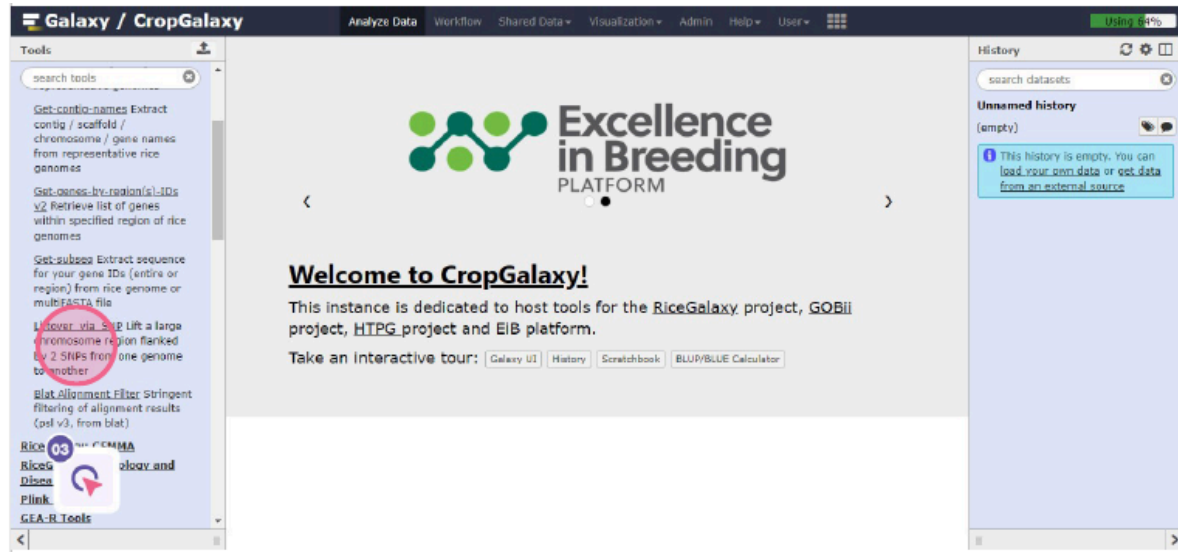
2.) SNP Data tool suite

On the tool panel, select RiceGalaxy: SNP Data tool suite. This set of tools contains the specific tool we are after. In the case where you do not know where to find the tool, you can always type the tool name on the “Search” box.



3.) LiftOver_via_SNP

LiftOver via SNP is a specialized alignment tool to find the location of large genome regions of interest from a source genome to another genome. It uses blat software as the alignment engine, and the output is blat **PSL** format (For more information about BLAT, please read the blat manual). The default parameters of blat should work fine for rice genomes.



4.) Set-up Lift-Over parameters

We now start our analysis by running lift-over from Nipponbare to N22 variety (aus), which, in this example, is more related to the donor variety than Nipponbare (temp japonica).

- a. Select Nipponbare as the source genome.
- b. Type in the coordinates of the QTL: chromosome (in this case, Chr2), left SNP position (22001414) and right SNP position (22831782) coordinates in the tool's form **EXACTLY** as it appears here.

Note that the Chromosome name must be correctly specified, otherwise, the tool will produce an error.

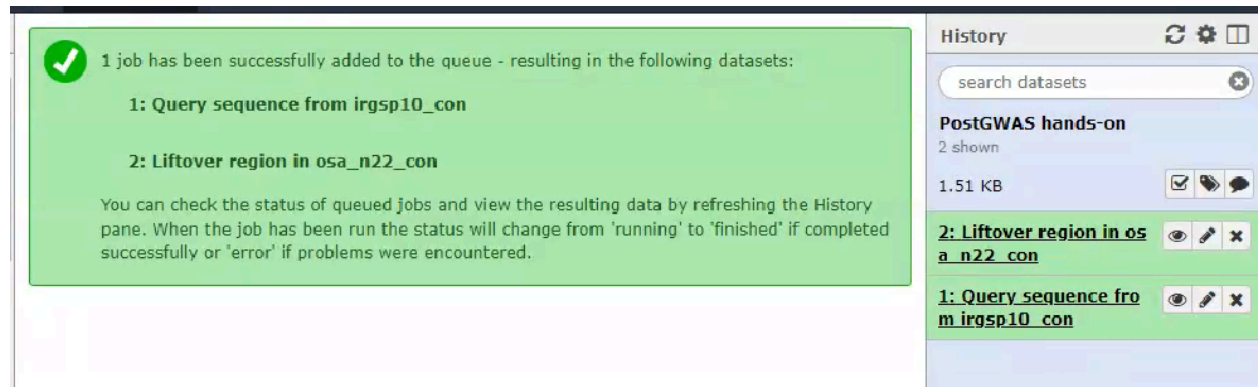
- c. Select N22 as the target genome.
- d. Leave the other settings to default for now.

5.) Name your history: PostGWAS hands-on

Before we proceed to running the tool, let us first annotate this analysis by naming our history as "PostGWAS hands-on".

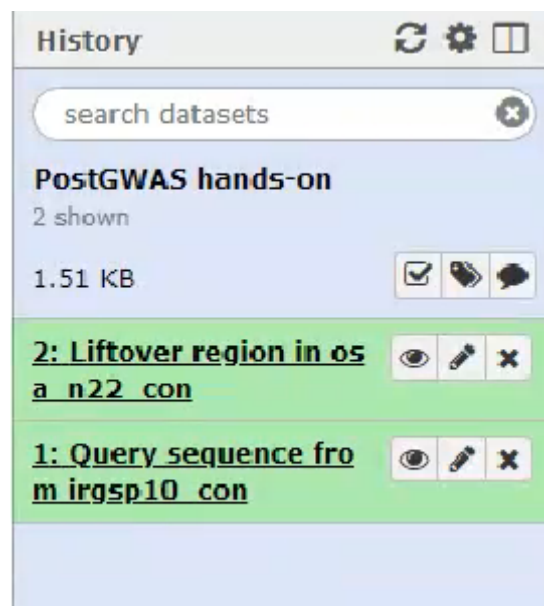
6.) Execute Lift-Over tool

Now we are ready. Just click on the "Execute" button to run the tool.



7.) Lift-over tool output files

The Lift Over tool outputs two files: the **query sequences from the source genome**, Nipponbare, and the **blat alignment output** where you can find the coordinates where these sequences are aligned in the target genome, N22.



8.) Viewing a dataset

To view a dataset in the history, click the "**View data**" (eye) icon.

Click on the first dataset. You can also resize the panel to adjust the size of the floating result window.

The screenshot shows the Galaxy / CropGalaxy web interface. The main panel displays a dataset titled "PostGWAS hands-on: Query sequence from irgsp10_con". The data is presented as a text-based genomic track with coordinates and sequence information. On the right, a "History" panel lists two datasets: "2: Liftover region in osa_n22_con" and "1: Query sequence from irgsp10_con". The interface includes a top navigation bar with options like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User".

Click on the second dataset. Now you can see two datasets side-by-side.

The screenshot shows the Galaxy / CropGalaxy web interface with two datasets displayed side-by-side. The left dataset, titled "PostGWAS hands-on: Liftover region in osa_n22_con", displays a table of genomic data. The right dataset, titled "PostGWAS hands-on: Query sequence from irgsp10_con", displays genomic coordinates and sequence information. The interface includes a top navigation bar with options like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User".

1	2	3	4	5	6	7	8	9	10
match	mis-match	rep.	N's	Q gap count	Q gap bases	T gap count	T gap bases	strand	Q name
117	4	0	0	0	0	0	0	+	Chr2: 22001354
110	11	0	0	0	0	1	18	+	Chr2: 22831722
119	2	0	0	0	0	0	0	+	Chr2: 22831722
100	11	0	0	0	0	0	0	+	Chr2: 22831722
90	9	0	0	0	0	0	0	+	Chr2: 22831722
59	2	0	0	2	8	1	31	-	Chr2: 22831722
76	5	0	0	1	21	1	18	-	Chr2: 22831722

To exit Scratchpad, just click on any gray space outside the floating window of any dataset displayed.

10.) Blat Alignment Filter

Examine the alignment result. Notice that it shows multiple hits on the probable location of the lift-over coordinates. So we need to filter this using the tool: **Blat Alignment Filter**

The screenshot shows the Galaxy / CropGalaxy web interface. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. On the left, a 'Tools' sidebar lists various genomic tools like Get-subseq, Liftover via SNP, Blat Alignment Filter, RiceGalaxy: GEMMA, RiceGalaxy: Pathology and Disease Tools, Plink Data Tools, GEA-R Tools, EIB: Genomic Selection, Data Format Conversion, BCFTools, Marker Selection, Imputation, and OGHMA. The main panel displays the 'Blat Alignment Filter Stringent filtering of alignment results (psl v3, from blat) (Galaxy Version v.01)' tool. The tool's configuration includes a dropdown for 'Alignment result to filter (psl format)' set to '5: Get-subseq on data 4 (as tabular)', a slider for 'Percent of length of query sequence that matches target sequence (usually 50 - 100%)' set to 80, and three integer input fields for 'maximum number of mismatch(es) allowed (integer)' (10), 'maximum number of gaps in query introduced (integer)' (3), and 'maximum number of gaps in target introduced (integer)' (3). An 'Execute' button is at the bottom. Below the tool form, a section titled 'blat-alignment-filter - what it does' explains that the tool parses alignment output from blat and find-seq, allowing for selection of alignments based on specified criteria (max mismatches, query length aligned). It notes that the tool uses 'awk' for filtering and provides a tip on how to use Find-seq alignment to identify scaffold/contig regions and extract subsequences with the Get-subseq tool.

The blat-alignment-filter parses alignment output of blat and find-seq (supports only blat psl format output), allowing for selection of alignments that pass specified criteria (# max mismatches, how much of query length is aligned).

It uses "awk" Linux utility in the background to filter the alignment results.




11.) Filtering parameters

In this exercise, we need to set a more stringent filtering criteria to keep only the sequences closest to the target genome. So we set the value "90" in Percent of length of query sequence that matches target sequence (usually 50 - 100%). We will use the default values of the other parameters.


Again, click on the "Execute" button to run the tool.

Blat Alignment Filter Stringent filtering of alignment results (psl v3, from blat) (Galaxy Version v.01) Options


Alignment result to filter (psl format)

   2: Liftover region in osa_n22_con

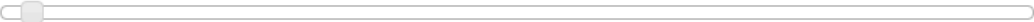
Percent of length of query sequence that matches target sequence (usually 50 - 100%)

90 


maximum number of mismatch(es) allowed (integer)


10 

maximum number of gaps in query introduced (integer)

3 

maximum number of gaps in target introduced (integer)

3 

 Execute

12.) Filtering output

You will see the dataset: "90 of query, mismatch 10, qgap 3, tgap 3 of data 2:Liftover region in osa_n22_con" added in your history.

Click on the "[View data](#)" button to see the result file. You will see that the resulting file now only has two entries.

Galaxy / CropGalaxy																				
PostGWAS hands-on: 90 of query, mismatch 10, qgap 3, tgap 3 of data 2:Liftover region in osa_n22_con																				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
psLayout version 3																				
match	mis-	rep.	N's	Q gap	Q gap	T gap	T gap	strand	Q	Q	Q	Q	T	T	T	T	block	block	qStarts	tStarts
match	match	match	count	count	bases	count	bases		name	size	start	end	name	size	start	end	count	count		
117	4	0	0	0	0	0	0	+	Chr2:22001354-22001474	121	0	121	Chr02	37316900	22947842	22947963	1	121,	0,	22947963
119	2	0	0	0	0	0	0	+	Chr2:22831722-22831842	121	0	121	Chr02	37316900	23768915	23769036	1	121,	0,	23769036

13.) Lifted-over coordinates on N22

You now have the new coordinates in N22 genome.

Copy the chromosome "Chr02".

Copy the value of the first entry in "T start" (Column 16). This will become your lower bound coordinate: "22947842"

Copy the value of the second entry in "T end" (Column 17). This will become your upper bound coordinate: "23769036"

The N22 genome coordinate of the lifted-over QTL is: Chr02:22947842-23769036

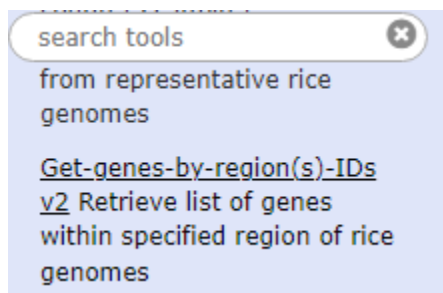
14	15	16	17
T	T	T	T
name	size	start	end

Chr02	37316900	22947842	22947963
Chr02	37316900	23768915	23769036

14.) Get-genes-by-region(s) IDs v2

Now, we want to look for candidate genes that lie in this region. To do that, we will use the tool "Get Genes from Region IDs v2".

Go to RiceGalaxy: SNP Data tools suite again and select "Get Genes from Region IDs v2".



15.) Get Genes parameter set-up

- Set the Reference genomes to get genes to : "Oryza sativa N22 (aus) new annotation"
- Set Gene or mRNA to: "mRNA"

- Set Type in a region directly (chr:start-end), OR use regions list from history? to: **"Directly type-in region"**
- To extract the genes in the QTL from N22 mRNA; use the lower & upper bound of the lift-over results. We will use here the one we got from the previous step: **"Chr02:22947842-23769036"**.
- Once the parameters are set, click **"Execute"**

Get-genes-by-region(s)-IDs v2 Retrieve list of genes within specified region of rice genomes (Galaxy) Options
Version v.0.20)

Reference genome to get genes..
 Oryza sativa N22 (aus) new annotation
 The info is derived from gene annotation efforts independent of the genome assembly (from GFF or other info sources).

Gene or mRNA
 mRNA
 Display gene or mRNA in the region?

Type in a region directly (chr:start-end), OR use regions list from history?
 Directly type-in region
 You can either type the region directly to search, or use multiple regions in a file from history.

Search region
 Chr02:22947842-23769036
 Directly type in your region (the chromosome/contig name SHOULD BE PRESENT in the genome) to list the genes (example in Nipponbare: Chr1:1-500000).

16.) Get Genes output

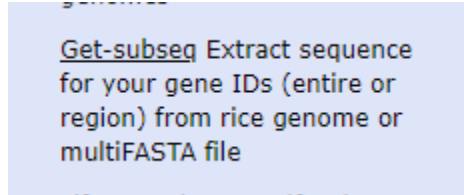
Click on **"View data"** to see the list of genes.

contig	start	end	strand	geneID	meta...
Chr02	22947201	22955608	+	OsN22RS2_02T0360100.3	
Chr02	22947201	22949065	+	OsN22RS2_02T0360100.4	
Chr02	22955540	22964416	+	OsN22RS2_02T0360100.1	
Chr02	22964321	22971731	+	OsN22RS2_02T0360100.2	
Chr02	22991728	22995923	-	OsN22RS2_02T0360500.1	
Chr02	22996987	23000953	-	OsN22RS2_02T0148200.1	
Chr02	22997710	23000953	-	OsN22RS2_02T0148200.2	
Chr02	23005106	23006373	-	OsN22RS2_02T0360600.1	
Chr02	23012554	23013471	-	OsN22RS2_02T0036300.1	
Chr02	23015537	23017220	+	OsN22RS2_02T0036200.1	
Chr02	23017514	23017902	+	OsN22RS2_02T0360200.1	
Chr02	23019238	23021541	+	OsN22RS2_02T0360300.1	
Chr02	23024207	23026193	+	OsN22RS2_02T0360400.1	
Chr02	23026523	23029188	-	OsN22RS2_02T0148300.1	
Chr02	23026523	23029188	-	OsN22RS2_02T0148300.2	

You will see that there are 149 genes (mRNA) in the lift-over region from N22 reference genome. The next thing we want to do is get the sequences of these genes. To do that, we will use the "Get-subseq" tool.

17.) Get-subset tool

This tool extracts a subsequence of interest from important rice reference genomes.



18.) Get-subset parameter set-up

- Set the output of the previous tool as input.
- Set gene ID column to : "Column 5"
- Set "Will you select a multiFASTA sequence from history or use a built-in rice genome database?" to: "Use a built -in gene database"
- Set Select reference database to: "O.sativa N22 aus mRNA v2.2"
- Click "**Execute**" afterwards.

Get-subseq Extract sequence for your gene IDs (entire or region) from rice genome or multiFASTA file (Galaxy Version 0.1.0) Options

Select (gene) ID/region list (tabular format)

4: mRNA : Chr02:22947842-23769036 in Oryza sativa N22 (aus) new annotation

Your gene ID list SHOULD BE (1) tabular data type and (2) present in your target sequence(s), else the tool fails!

gene ID column

Column: 5

Select column number where your gene ID/region(s) are (if your list is a multicolumn table).

Will you select a multiFASTA sequence from history or use a built-in rice genome database?

Use a built-in gene database

Built-in genomes and genes for representative rice variety groups are included.

Select reference database

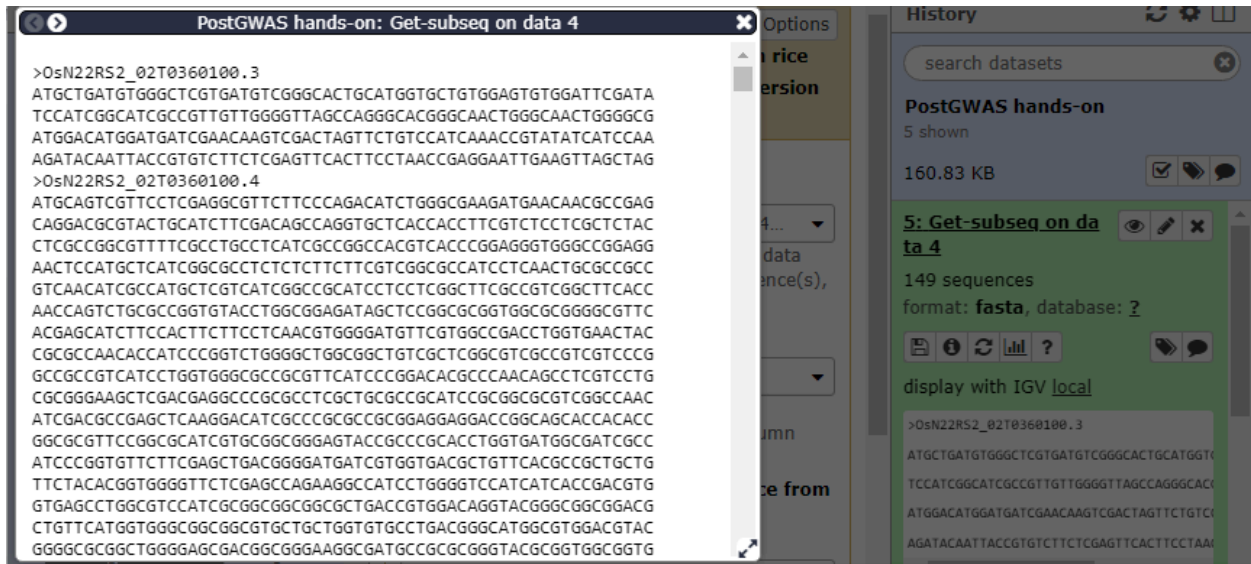
O. sativa N22 (aus) mRNA v2.2

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

Execute

19.) Get-subset output

Click "View data" to display output.



You now have a multi-FASTA file which contains the sequences of the genes within the lifted-over QTL in the N22 reference genome. You may use these sequences to look for further annotations (putative functions) from other data sources such as NCBI database, Gramene, or Rice SNP Seek.

---End of part 1 ---

Part 2: Using Nippobare annotation to learn more about the putative functions of these genes in N22

Recall: We generated a gene list and their sequences anchored on N22, an aus-type genome, for the QTL, qDTF2.2. We did this because our donor is closely related to N22.

Our goal now is to use available annotations to learn about the putative functions of the genes in our list. We will now begin another series of analysis.

1.)Find-seq

Use the sequences derived from Part 1 step 19 to align back to Nipponbare.

Set-up the parameters as shown in the image below:

The screenshot shows the Galaxy / CropGalaxy web interface. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. The left sidebar contains a 'Tools' section with a search bar and a list of tools under 'RiceGalaxy: SNP Data Tools'. The main panel displays the 'Find-seq' tool configuration. The tool title is 'Find-seq Align your nucleotide sequences to rice variety-representative genomes (Galaxy Version v.01)'. The configuration includes a 'Query FASTA file to align against rice genome of choice' dropdown set to '5: Get-subseq on data 4'. Below this is a section 'Select a built-in rice genome/gene database or use a multi-FASTA data from your history?' with a dropdown set to 'Use a built-in gene database'. A note states 'Built-in rice genomes are representative of important rice variety groups'. The 'Select reference database' dropdown is set to 'O. sativa Nipponbare (japonica) cDNA RGAP7 IRGSP1.0'. A note below this says 'If your genome of interest is not listed, contact Rice Galaxy team!'. The 'size of match that triggers an alignment(between 8 - 12 allowed)' is set to 11. The 'maximum intron size allowed' is set to 5000. The 'Format for alignment output' dropdown is set to 'Default blat tabular format,no sequence'. An 'Execute' button is at the bottom.

Galaxy / CropGalaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data

RiceGalaxy: SNP Data Tools

[FGSEA-2](#) - fast preranked gene set enrichment analysis

[Find-gene-by-terms](#) Find gene(s) using a single or list of text terms/phrases

[Find-seq](#) Align your nucleotide sequences to rice variety-representative genomes

[Get-contig-names](#) Extract contig / scaffold / chromosome / gene names from representative rice genomes

[Get-genes-by-region\(s\)-IDs v2](#) Retrieve list of genes within specified region of rice genomes

Find-seq Align your nucleotide sequences to rice variety-representative genomes (Galaxy Version v.01) Options

Query FASTA file to align against rice genome of choice

5: Get-subseq on data 4

Select a built-in rice genome/gene database or use a multi-FASTA data from your history?

Use a built-in gene database

Built-in rice genomes are representative of important rice variety groups

Select reference database

O. sativa Nipponbare (japonica) cDNA RGAP7 IRGSP1.0

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

size of match that triggers an alignment(between 8 - 12 allowed)

11

maximum intron size allowed

5000

Format for alignment output

Default blat tabular format,no sequence

Execute

After running the tool, you will get an alignment output.

PostGWAS hands-on: Alignment of data 5: Get-subseq on data 4																		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
psLayout version 3																		
match	mis-match	rep. match	N's	Q gap count	Q gap bases	T gap count	T gap bases	strand	Q name	Q size	Q start	Q end	T name	T size	T start	T end	block count	block sizes
33	2	0	0	0	0	0	0	+	OsN22RS2_02T0360100.3	240	171	206	LOC_Os02g36450.2	1753	0	35	1	35,
33	2	0	0	0	0	0	0	+	OsN22RS2_02T0360100.3	240	171	206	LOC_Os02g36450.1	1866	0	35	1	35,
33	2	0	0	0	0	0	0	+	OsN22RS2_02T0360100.3	240	171	206	LOC_Os02g36450	1866	0	35	1	35,
158	3	0	0	2	5	3	40	+	OsN22RS2_02T0360100.3	240	0	166	LOC_Os02g36414.1	2308	1860	2061	6	7,26,57,17,46,8,
158	3	0	0	2	5	3	40	+	OsN22RS2_02T0360100.3	240	0	166	LOC_Os02g36414	2308	1860	2061	6	7,26,57,17,46,8,
45	4	0	0	0	0	0	0	+	OsN22RS2_02T0360100.4	1422	1084	1133	LOC_Os07g01560.2	2120	833	882	1	49,
1404	9	0	0	3	9	0	0	+	OsN22RS2_02T0360100.4	1422	0	1422	LOC_Os02g36414.1	2308	393	1806	4	1351,30,17,15,
1404	9	0	0	3	9	0	0	+	OsN22RS2_02T0360100.4	1422	0	1422	LOC_Os02g36414	2308	393	1806	4	1351,30,17,15,
45	4	0	0	0	0	0	0	-	OsN22RS2_02T0360100.4	1422	1084	1133	LOC_Os07g01550.1	4010	3619	3668	1	49,
45	4	0	0	0	0	0	0	-	OsN22RS2_02T0360100.4	1422	1084	1133	LOC_Os07g01550	4010	3619	3668	1	49,
32	1	0	0	0	0	0	0	+	OsN22RS2_02T0360100.1	1620	1428	1461	LOC_Os08g08070.1	2060	1485	1518	1	33,
32	1	0	0	0	0	0	0	+	OsN22RS2_02T0360100.1	1620	1428	1461	LOC_Os08g08070	2060	1485	1518	1	33,
39	2	0	0	0	0	1	24	+	OsN22RS2_02T0360100.1	1620	1254	1295	LOC_Os04g37970.1	1563	1272	1337	2	31,10,
39	2	0	0	0	0	1	24	+	OsN22RS2_02T0360100.1	1620	1254	1295	LOC_Os04g37970	1563	1272	1337	2	31,10,
1278	71	0	0	4	156	5	54	+	OsN22RS2_02T0360100.1	1620	0	1505	LOC_Os02g36450.2	1753	66	1469	6	46,9,574,268,12,440,
62	1	0	0	0	0	0	0	+	OsN22RS2_02T0360100.1	1620	1557	1620	LOC_Os02g36450.2	1753	1	64	1	63,
1383	76	0	0	3	46	5	55	+	OsN22RS2_02T0360100.1	1620	0	1505	LOC_Os02g36450.1	1866	66	1580	6	38,12,689,268,12,440,

Upon closer look on the output, you will notice that there are a lot of gene duplications and genome rearrangements given the chromosome location of the Nipponbare genes. We will need to filter this using “**Blat Alignment Filter**”.

2.) Blat_alignment_filter

We will filter out PSL table to get only those whose sequence matches “90%” with the Nipponbare gene sequence.

Tools

from representative rice
search tools

[Get-genes-by-region\(s\)-IDs v2](#) Retrieve list of genes within specified region of rice genomes

[Get-subseq](#) Extract sequence for your gene IDs (entire or region) from rice genome or multifasta file

[Liftover via SNP](#) Lift a large chromosome region flanked by 2 SNPs from one genome to another

[Blat Alignment Filter](#) Stringent filtering of alignment results (psl v3, from blat)

[RiceGalaxy: GEMMA](#)

[RiceGalaxy: Pathology and](#)

Blat Alignment Filter Stringent filtering of alignment results (psl v3, from blat) (Galaxy Version v.01)
Options

Alignment result to filter (psl format)

10: Alignment of data 5: Get-subseq on data 4

Percent of length of query sequence that matches target sequence (usually 50 - 100%)

90

maximum number of mismatch(es) allowed (integer)

10

maximum number of gaps in query introduced (integer)

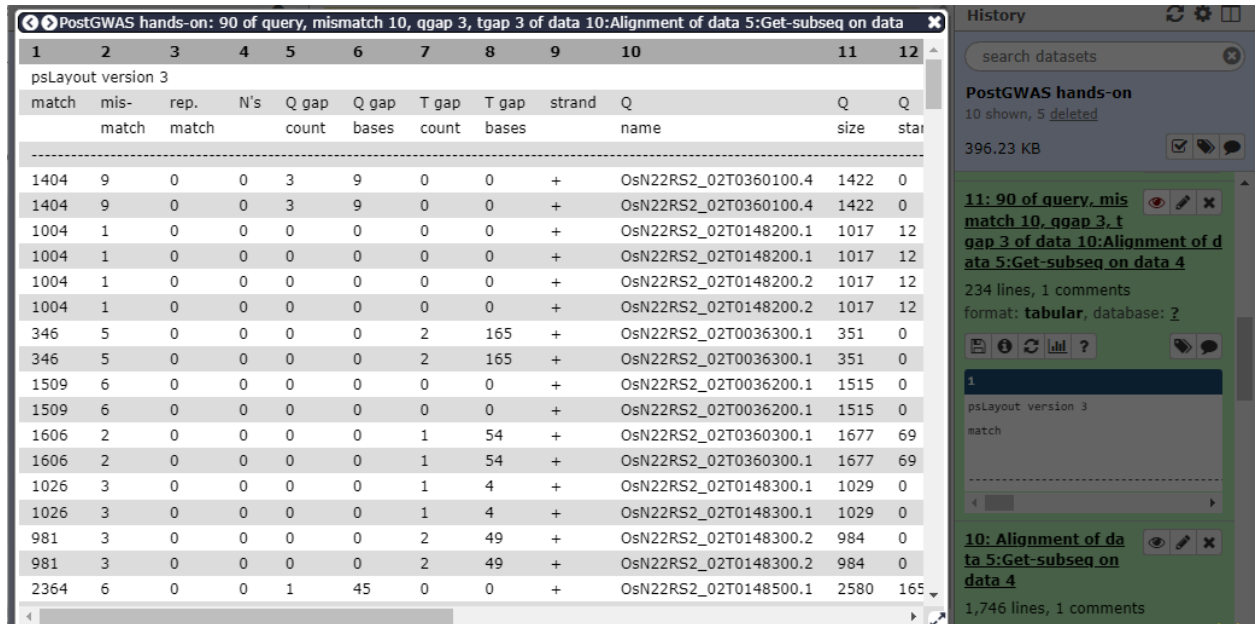
3

maximum number of gaps in target introduced (integer)

3

Execute

After running the tool, you will see this output. Notice the reduction of alignment results from 1742 records to 230.



The screenshot shows a Galaxy interface with a table of alignment results and a history panel on the right.

Table:

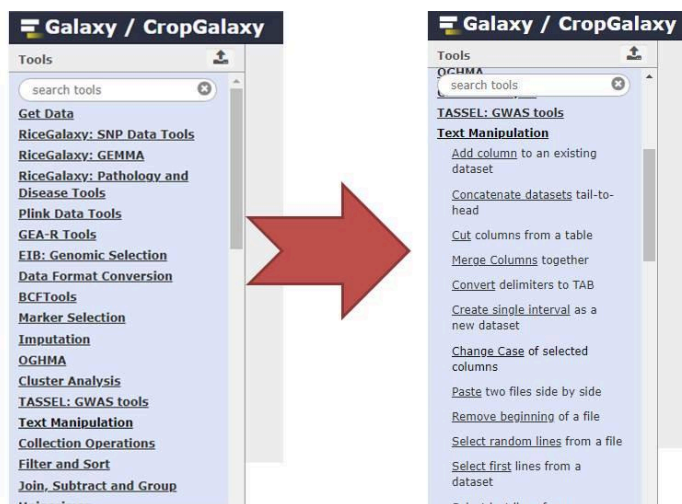
1	2	3	4	5	6	7	8	9	10	11	12
match	mis-match	rep. match	N's	Q gap count	Q gap bases	T gap count	T gap bases	strand	Q name	Q size	Q star
1404	9	0	0	3	9	0	0	+	OsN22RS2_02T0360100.4	1422	0
1404	9	0	0	3	9	0	0	+	OsN22RS2_02T0360100.4	1422	0
1004	1	0	0	0	0	0	0	+	OsN22RS2_02T0148200.1	1017	12
1004	1	0	0	0	0	0	0	+	OsN22RS2_02T0148200.1	1017	12
1004	1	0	0	0	0	0	0	+	OsN22RS2_02T0148200.2	1017	12
1004	1	0	0	0	0	0	0	+	OsN22RS2_02T0148200.2	1017	12
346	5	0	0	0	0	2	165	+	OsN22RS2_02T0036300.1	351	0
346	5	0	0	0	0	2	165	+	OsN22RS2_02T0036300.1	351	0
1509	6	0	0	0	0	0	0	+	OsN22RS2_02T0036200.1	1515	0
1509	6	0	0	0	0	0	0	+	OsN22RS2_02T0036200.1	1515	0
1606	2	0	0	0	0	1	54	+	OsN22RS2_02T0360300.1	1677	69
1606	2	0	0	0	0	1	54	+	OsN22RS2_02T0360300.1	1677	69
1026	3	0	0	0	0	1	4	+	OsN22RS2_02T0148300.1	1029	0
1026	3	0	0	0	0	1	4	+	OsN22RS2_02T0148300.1	1029	0
981	3	0	0	0	0	2	49	+	OsN22RS2_02T0148300.2	984	0
981	3	0	0	0	0	2	49	+	OsN22RS2_02T0148300.2	984	0
2364	6	0	0	1	45	0	0	+	OsN22RS2_02T0148500.1	2580	165

History Panel:

- PostGWAS hands-on: 90 of query, mismatch 10, qqap 3, tgap 3 of data 10:Alignment of data 5:Get-subseq on data 4
- 11: 90 of query, mismatch 10, qqap 3, tgap 3 of data 10:Alignment of data 5:Get-subseq on data 4
- 10: Alignment of data 5:Get-subseq on data 4

3.) Post-process the result

We will now extract the genes in column 14 from the output file of the blat filtering process. For this, we will use the Text Manipulation tools.






- a. Take out column 14 (c14) from the table using “Cut columns from a table” tool.

Cut columns from a table (Galaxy Version 1.0.2) Options

Cut columns
c14

Delimited by
Tab




From
   11: 90 of query, mismatch 10, qgap 3, tgap 3 of dat...
Execute

WARNING: This tool breaks column assignments. To re-establish column assignments run the tools and click on the pencil icon in the latest history item.

- b. Remove the headers and comments using “Remove beginning of a file”

Remove beginning of a file (Galaxy Version 1.0.0) Options

Remove first
5
lines

from
   12: Cut on data 11
Execute

What it does

This tool removes a specified number of lines from the beginning of a dataset.

After these two steps, you will be left with a one-column tabular file, which consists of 230 lines of genes.

1
LOC_Os02g36414.1
LOC_Os02g36414
LOC_Os02g36500.1
LOC_Os02g36500
LOC_Os02g36500.1
LOC_Os02g36500
LOC_Os02g36505.1
LOC_Os02g36505
LOC_Os02g36510.1
LOC_Os02g36510
LOC_Os02g36520.1
LOC_Os02g36520
LOC_Os02g36550.1
LOC_Os02g36550
LOC_Os02g36550.1
LOC_Os02g36550
LOC_Os02g36570.1
LOC_Os02g36570
LOC_Os02g36570.1
LOC_Os02g36570
LOC_Os02g36595.1
LOC_Os02g36595
LOC_Os02g36590.3
LOC_Os02g36590.2
LOC_Os02g36590.1

4.) Find genes by term (using the gene list)

We will now get putative functions information on these genes using an annotation from Nipponbare.

To do that, use the “**Find-genes-by-terms**” tool.

Find-gene-by-terms Find gene(s) using a single or list of text terms/phrases

Options

(Galaxy Version v.0.3.0)

Which genome do you wish to search?

O. sativa Nipponbare cDNA RGAP7 annotation

The info is obtained from gene annotations for several rice genomes (from GFF or other info sources).

Type in term(s) directly, OR use terms list from history?

Use a list of terms/phrases from the history

You can either type the term(s) directly to search, or use multiple terms in a file from history. Search includes partial matches.

Select list of terms to search (search is OR operation)

13: Remove beginning on data 12

UPLOAD a list file with the terms to search, 1 search term per line only, NO BLANKS!

Case-sensitive search or not?

Case-insensitive

Upper/lower case distinction enforced or not?

Whole word or partial match search?

Partial match

Match only whole words enforced or not?

Execute

Find-gene rice - what it does

Retrieves list of genes that match the search terms (either directly typed in, or from a list), using the annotation description for the selected rice genomes:

After running the tool, you should be able to see on **column 7** the **annotations** regarding the putative function of the genes.

1	2	3	4	5	6	7
Chr1	34425818	34427049	+	LOC_Os01g59540	LOC_Os01g59540.1	GRF zinc finger family protein
Chr1	36317205	36318022	+	LOC_Os01g62720	LOC_Os01g62720.1	GRF zinc finger family protein
Chr1	42830103	42830920	-	LOC_Os01g73920	LOC_Os01g73920.1	hypothetical protein
Chr2	956628	960792	-	LOC_Os02g02610	LOC_Os02g02610.1	transposon protein, putative, unclassified, expressed
Chr2	21995376	22002735	+	LOC_Os02g36414	LOC_Os02g36414.1	transporter family protein, putative, expressed
Chr2	22043694	22046900	-	LOC_Os02g36500	LOC_Os02g36500.1	expressed protein
Chr2	22053938	22054860	-	LOC_Os02g36505	LOC_Os02g36505.1	expressed protein
Chr2	22056928	22058611	+	LOC_Os02g36510	LOC_Os02g36510.1	ethylene-insensitive 3, putative, expressed
Chr2	22060800	22064176	+	LOC_Os02g36520	LOC_Os02g36520.1	OsFBK9 - F-box domain and kelch repeat containing
Chr2	22068014	22070683	-	LOC_Os02g36550	LOC_Os02g36550.1	phosphopantothenate--cysteine ligase, putative, expressed
Chr2	22077489	22085283	+	LOC_Os02g36570	LOC_Os02g36570.1	ABC1 family domain containing protein, putative, expressed
Chr2	22096006	22100897	+	LOC_Os02g36590	LOC_Os02g36590.1	CPuORF19 - conserved peptide uORF-containing transmembrane
Chr2	22098509	22100476	+	LOC_Os02g36595	LOC_Os02g36595.1	expressed protein
Chr2	22102488	22105124	+	LOC_Os02g36600	LOC_Os02g36600.1	aldose 1-epimerase, putative, expressed
Chr2	22110121	22117660	+	LOC_Os02g36619	LOC_Os02g36619.1	exo70 exocyst complex subunit, putative, expressed
Chr2	22135516	22136944	-	LOC_Os02g36670	LOC_Os02g36670.1	expressed protein
Chr2	22140602	22141996	-	LOC_Os02g36680	LOC_Os02g36680.1	expressed protein
Chr2	22144257	22144998	-	LOC_Os02g36690	LOC_Os02g36690.1	expressed protein
Chr2	22146420	22150147	-	LOC_Os02g36700	LOC_Os02g36700.1	sucrose transporter BoSUT1, putative, expressed
Chr2	22151432	22156042	-	LOC_Os02g36710	LOC_Os02g36710.1	SET, putative, expressed
Chr2	22163299	22167383	-	LOC_Os02g36740	LOC_Os02g36740.1	zinc finger, C3HC4 type, putative, expressed
Chr2	22215057	22217951	+	LOC_Os02g36830	LOC_Os02g36830.1	cytokinin-O-glucosyltransferase 2, putative, expressed
Chr2	22229412	22231714	+	LOC_Os02g36840	LOC_Os02g36840.1	cytokinin-O-glucosyltransferase 2, putative, expressed
Chr2	22234374	22235445	-	LOC_Os02g36850	LOC_Os02g36850.1	oxygen evolving enhancer protein 3, identical, putative

You may also use other resources such as RicePilaf, Rice SNPSeek, Rice Gene Index to gain more insight on the function of the genes as well as gene networks.

---End of part 2---

Hands-on prepared by:

Ramil Mauleon

Venice Margarette Juanillas