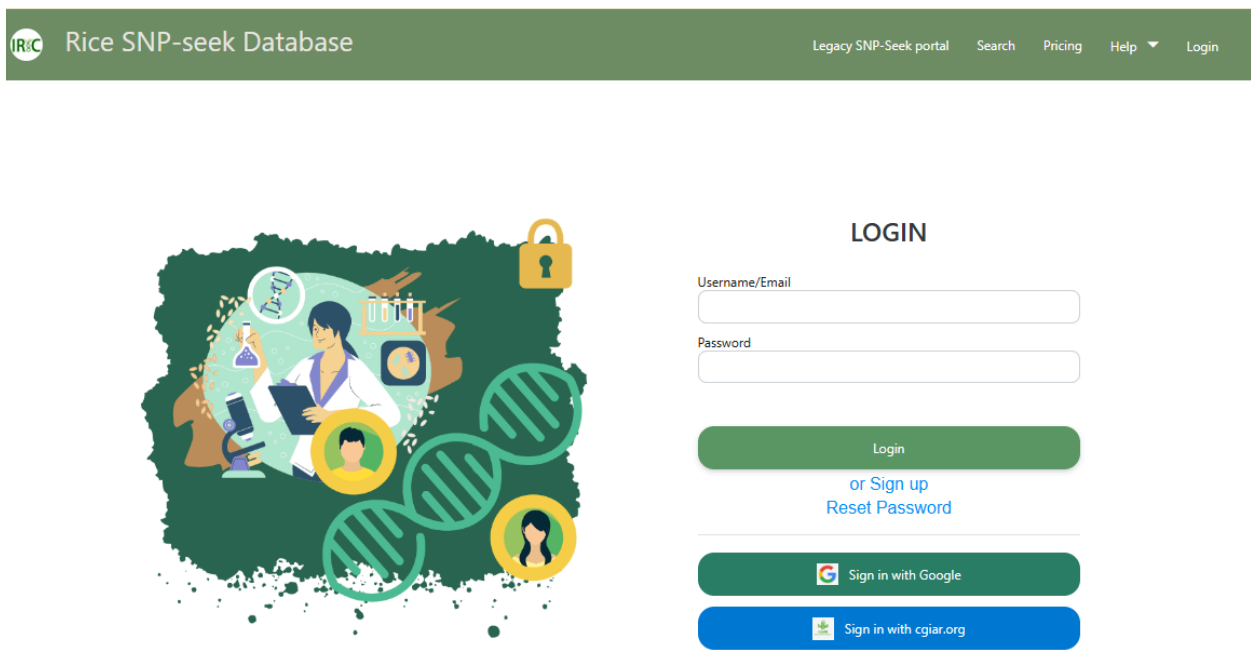# Post-GWAS using RiceSNPSeek (using v3)

QTL of Interest: [PHS](#)

Publication: Lee JS, Chebotarov D, McNally KL, Pede V, Setiyono TD, Raquid R, Hyun WJ, Jeung JU, Kohli A, Mo Y. *Novel Sources of Pre-Harvest Sprouting Resistance for Japonica Rice Improvement. Plants (Basel). 2021 Aug 19;10(8):1709. doi: 10.3390/plants10081709. PMID: 34451754; PMCID: PMC8401653.*

1.) Go to: https://snp-seek.irri.org/

a. Login using your SNP-seek credentials. There are three ways to do this:
    i. Username/email and Password
    ii. Sign in using your Google account
    iii. Sign in using [cgiar.org](#) (if you are part of this domain)
b. Sign up if you do not have an account in SNPSeek yet.

2.) Search Genotypes:

Note that our QTL/trait of interest is in this region: **Chr4:4662701-4670717**

You can begin by searching for SNPS that lie within this region using the 3k rice dataset. There are 4 SNP Sets available:

- **3K All**: 32 million full 3K RG SNPs Dataset

  This SNP Set contains the full set of 32 million biallelic & multiallelic SNP.

  Total SNPs: 32,064,217

  Samples : 3024

- **3kbase**: 18 million base SNP

  The Base SNP set of ~18 million SNPs was created from the ~29 million biallelic SNPs subset from the 32M full SNP set by removing SNPs with excess of heterozygous calls.

- **3K core**: 404k CoreSNP dataset

  The Core SNP set was obtained from the filtered SNP set by applying two-step LD pruning procedure as follows:

  1.) LD pruning with window size 10kb, step 1 SNP, R2 threshold 0.8
  2.) LD pruning with window size 50 SNPs, step 1 SNP, R2 threshold 0.8

- **3k filtered**: 4.8million filtered SNP dataset

  The filtered SNP set was obtained from the Base SNP set by applying the following filtering criteria:

  1.) alternative allele frequency at least 0.01
  2.) proportion of missing calls per SNP at most 0.2

You may use different SNP sets but for this hands-on, we will use the **"3Kbase" SNP set.**

In the Chromosome, put **"Chr4"**, in Start use **"4662701"** and end **"4670717".**

Click **"Search"** to retrieve the positions within this region within the 3KRG accessions.

3.) Query result

After querying the genotypes, you will have a table which displays the alleles for each 3KRG variety in a row, for all SNP positions within the specified region or gene. By default, the varieties are sorted by decreasing number of allele mismatches. The resulting table can be filtered using any of the column values, and sorted based on any column value by clicking the column header.

It will also report a graph of the allele frequencies where you can inspect the major and minor alleles per subpopulation for all SNP positions from the query.
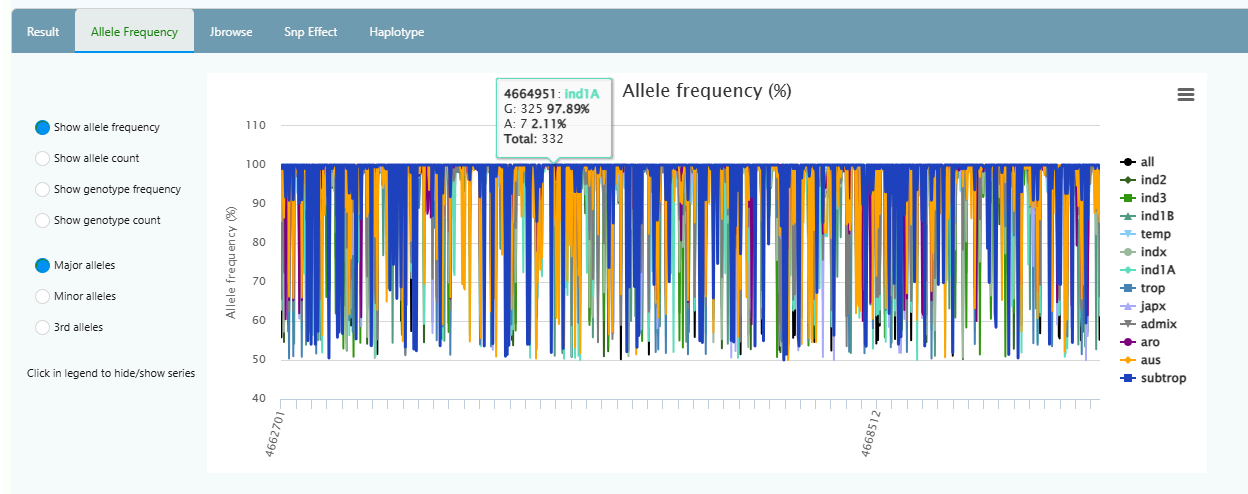
In another panel, you can also view the region in a genome browser.
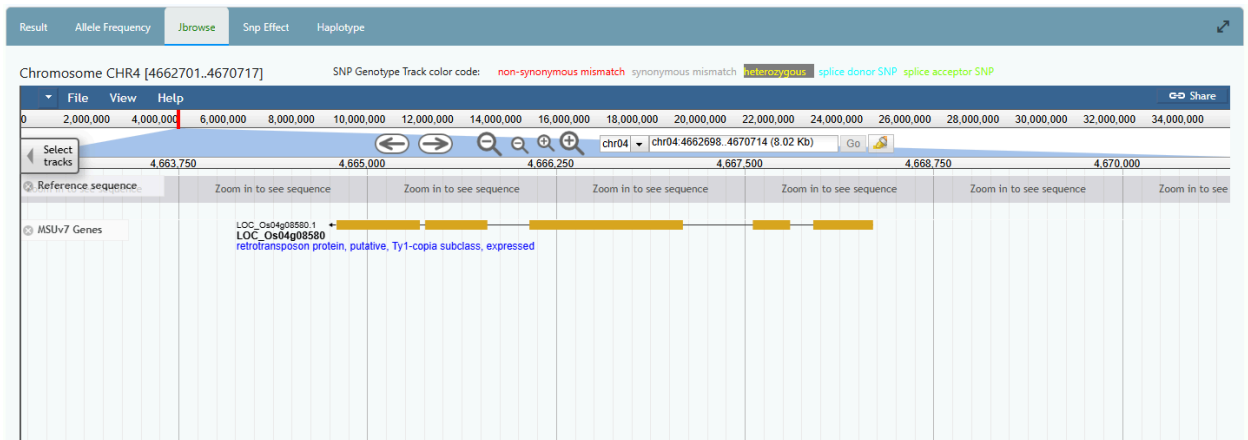
4.) SNP Effects

The Genotype query also returns a table containing the SNP effects detected on MSU gene models. This is calculated using the SNPEff program.
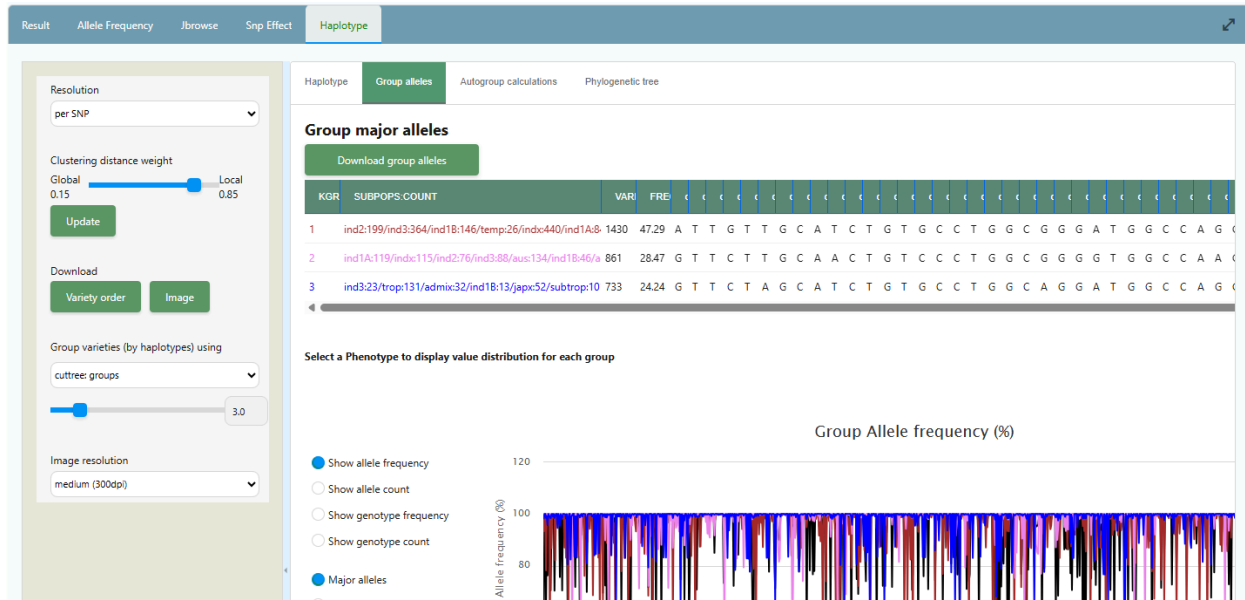
POSITION	"DELETERIOUSNESS"	Alt (ANN)	Effect (ANN)	Putative_impact (ANN)	Gene Name (ANN)	Gene ID (ANN)	Feature type (ANN)	Feature ID (ANN)	Transcript B
chr04-4662701	30	A A	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662730	30	A A	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662732	30	G G	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662735	30	G G	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662743	30	G G	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662767	30	T T	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662795	30	C C	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662822	30	G G	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -
chr04-4662864	30	G G	downstream_gene intergenic_region	MODIFIER MODIFIER	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	LOC_Os04g08580 LOC_Os04g08570- LOC_Os04g08580	transcript intergenic_region	LOC_Os04g08580.1 LOC_Os04g08570- LOC_Os04g08580	Coding -

5.) Haplotype View

Another view for the result of the genotype query is the Haplotype View. This allows us to know whether our haplotypes make sense.
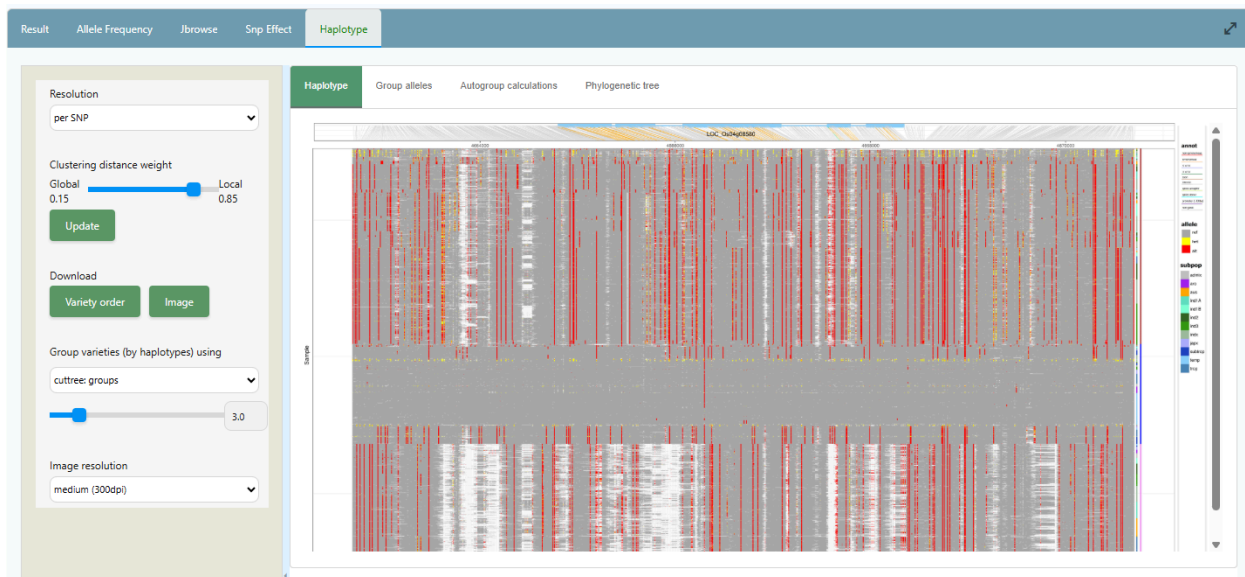
You should also see major grouping of alleles which contributes to the haplotype.

In this hands-on, you should see **3 major haplotype groups**.

6.) Downloading the results table

Download the table in "Plink" format by clicking on the "Plink" button.

This will give you two files in a zipped folder: a ped and map file that by now you should be familiar with.

The first few lines of your downloaded .map file should look like this:

```
chr04    10404662701 0    4662701
chr04    10404662730 0    4662730
chr04    10404662732 0    4662732
chr04    10404662735 0    4662735
chr04    10404662743 0    4662743
chr04    10404662767 0    4662767
chr04    10404662795 0    4662795
chr04    10404662822 0    4662822
chr04    10404662864 0    4662864
chr04    10404662867 0    4662867
chr04    10404662873 0    4662873
chr04    10404662892 0    4662892
chr04    10404662906 0    4662906
chr04    10404662920 0    4662920
chr04    10404662927 0    4662927
chr04    10404662930 0    4662930    ...
```

Likewise, the first few lines of your .ped file should be like this:

```
IRIS_313-11645  IRIS_313-11645  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-8493   IRIS_313-8493   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-9066   IRIS_313-9066   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11453  IRIS_313-11453  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11098  IRIS_313-11098  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11381  IRIS_313-11381  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
CX317   CX317   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C  C  A  A  T
IRIS_313-8703   IRIS_313-8703   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-10114  IRIS_313-10114  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-8407   IRIS_313-8407   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11700  IRIS_313-11700  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11761  IRIS_313-11761  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11085  IRIS_313-11085  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-9023   IRIS_313-9023   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11705  IRIS_313-11705  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-10863  IRIS_313-10863  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
B264    B264    0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C  C  G  G  T
IRIS_313-10177  IRIS_313-10177  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11677  IRIS_313-11677  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11683  IRIS_313-11683  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11709  IRIS_313-11709  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11762  IRIS_313-11762  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11789  IRIS_313-11789  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-12259  IRIS_313-12259  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-11273  IRIS_313-11273  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
CX306   CX306   0  0  0  -9  A  A  T  T  T  C  C  T  T  T  T  G  G  C  C  A  A  T
IRIS_313-8356   IRIS_313-8356   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-10337  IRIS_313-10337  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-9160   IRIS_313-9160   0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
IRIS_313-10332  IRIS_313-10332  0  0  0  -9  A  A  T  T  T  T  G  G  T  T  T  T  G  G  C
```

7.) Add positions in my list

From the map file, remove the middle columns such that only the 1st column and the last columns remain (chr, position).

Go to "My lists". Select "SNP" in List Type. Click on "Create" to make a new list.

Paste the chr-position table from the map file.

You may tick the box on "Verify if SNP is in Variant set". However, in this hands-on, we will not do that since we know that these SNPs already came as a result of the genotype query.

You will be able to see this new list on the left side of the panel.

8.) Annotate SNP List

Next thing we want to do is to look for candidate genes within the QTL region in Nipponbare, go to "Search"-> "Gene Loci".

Set "Japonica nipponbare" as reference genome.

Set "SNP Annotator (My SNP List)" in Search.

Set "All" to search all Gene models available.

Tick the box to "Include gene interactions (RiceNetv2, PRIN)" and "Gene Ontology" annotations. You may choose to add more to gather more annotations for this region.

The result should be a table which contains the annotations that can be grouped by marker position, gene locus, or by QTL.

9.) Saving the Gene List

Set the List name as "PHS-genelist" and "click Create Gene List".

Go to "My Lists" and you should see the gene locus/loci when you search under "Locus".

10.) Get Gene set/network

Using the gene list we generated, we will now do a gene enrichment analysis to know if there are other genes previously reported that are associated with our QTL of interest.

To do this, go to "Search"-> "Gene Loci.

Set as Reference genome: "Japonica Nipponbare".

Set Search to: "Gene set/networks".

Set Gene Models to: "All".

Set the My Locus List to: "PHS-genegenes".

Tick "Ricenetv2 direct neighbors" and "AgriGO".

Click on "Gene set/enrichment".

Unfortunately, this query will not return any result. Click the "Go" button beside the Ricenet option and it will give us this message:

"No ROC analysis, because the valid query set size < 4."

In this case, our query size is very small. You can explore nearby regions and see if there are existing gene networks reported by increasing the upper and lower bound positions.

To know more about SNPSeek, visit the site: https://snpseek.irri.org/.