

# araGWAB tutorial

## 1. DATA SUBMISSION

Click the 'RUN' menu.

### araGWAB RUN

Submit files and run araGWAB.

(The test example may take up to 10 ~ 30 minutes depending on the status of the server)

**7** FILL TEST EXAMPLE   **8** SHOW EXPECTED RESULT

E-mail (Optional)

**1** you want to receive job status information and query results via email, please submit your email address

Phenotype name **2**

GWAS data set **3**

sample file

파일 선택   선택된 파일 없음

Reference phenotype gene set

FILE UPLOADING   MANUAL INPUT   **4**

Gene set file **4**

파일 선택   선택된 파일 없음

Genome Build   SNP-gene dist. range **5**

TAIR 10   10000

Test p-value threshold range: log<sub>10</sub>(P) **5**   Interval of log<sub>10</sub>(P) threshold **5**

-6 ~ -2 (1.00e-6 ~ 1.00e-2)   0.3

**6** SUBMIT   RESET

① (Optional) If you want to receive job status information and query results via email, please submit your email address here

② Enter a user-defined phenotype name. (ex. Flowering time)

③ Upload GWAS data set file with the following format.

- a. Input file name should have \*.txt extension.
- b. Input data format: [chromosome number] [tab] [basepair] [tab] [p-value] [new line]
  - 1-5, M and C are acceptable for chromosome number.
  - We support various formats of p-value (such as 0.00001, .00001 and 1e-05).

<b>CHR</b>	<b>BP</b>	<b>PVALUE</b>
Chromosome number (1-5, C, M)	Chromosomal position of each SNP (bp)	p-value of each SNP
1	1051029	.0089792
5	3312	0.0001251
C	99102	4.213000e-05

④ To find the optimal  $p$ -value threshold for boosting, a set of reference phenotype gene set is needed. Only genes with AGI locus IDs are accepted (e.g. AT1G01010). We provide two ways to upload a reference phenotype gene set (File uploading and Manual input).

A. File uploading

- a. Input file name should have a \*.txt extension.
- b. Input data format: [AGI locus ID] [new line]

B. Manual input

- a. If you have a list of phenotype genes, you can put them in a text box
- b. The list of genes must be separated with tabs, spaces or a new line

⑤ Configuration options.

Item	Description	Example
<b>Genome Build</b>	Arabidopsis genome build	default : TAIR10
<b>SNP-gene dist. range</b>	Chromosomal distance range from SNP in searching for genes to assign p-value	default : 10000 bp
<b>Test p-value threshold range: log<sub>10</sub>(P)</b>	Threshold range of SNP p-value for GWAB	default : -6 ~ -2
<b>Interval of log<sub>10</sub>(P) threshold</b>	Step of log <sub>10</sub> (P) threshold in the above test	default : 0.3

⑥ Click 'SUBMIT' button, then you can see the job status page

⑦ If you want to run araGWAB using a test example, click 'FILL TEST EXAMPLE' button.

(This job may take 10 ~ 30 minutes depending on the status of the server.)

⑧ You can see the expected results for the test example by clicking 'SHOW EXPECTED RESULT' button.

## 2. STATUS OF araGWAB JOB

To access araGWAB result of your query, users should keep a record of the URL for the STATUS OF GWAB JOB page. Also, this status page will be refreshed in every 10 seconds. Once the analysis is completed, the status page will move to the araGWAB result page automatically. If you have submitted your email as running parameter, you will receive the URL of Job status page via email.

# JOB STATUS

Bookmark this URL to access your query results in the future.  
In case you have submitted your email address, you will receive this URL via email  
This status page is refreshed in every 10 secs.

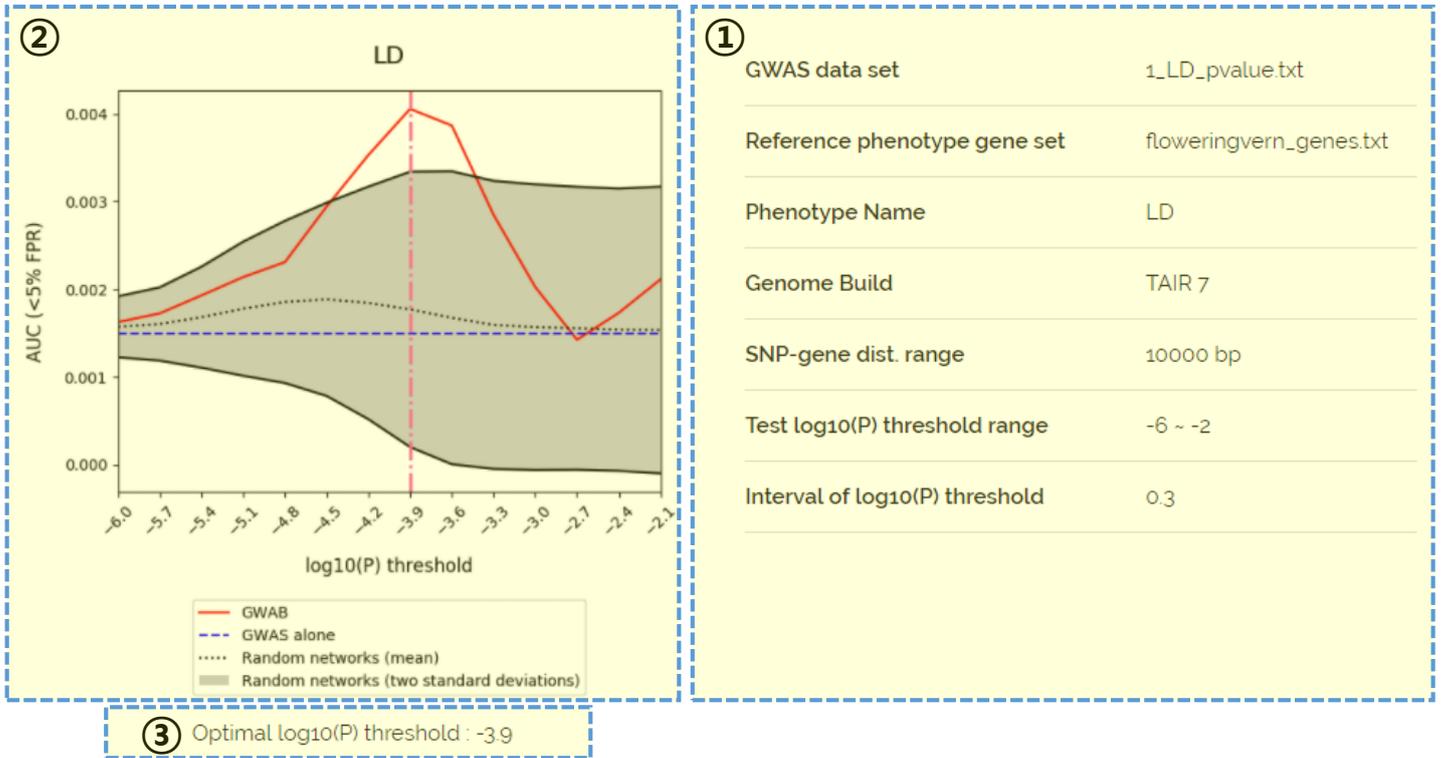
<a href="#">GWAB job running status</a>			<a href="#">GWAB running parameters</a>	
SNP Mapping	P-value Extracting	Boosting & Validation	GWAS data set	LD_sample_input.txt
② Done	③ Done	④ Running	Reference phenotype gene set	floweringvern_genes.txt
			Phenotype Name	Long day flowering time
			Genome Build	TAIR 7
			SNP-gene dist. range	10000
			Test log <sub>10</sub> (P) threshold range	-6 ~ -2
			Interval of log <sub>10</sub> (P) threshold	0.3

- ① This panel shows information for running parameters.
- ② To assign SNP to genes, araGWAB performs distance-based gene mapping. Here, araGWAB supports three versions of genome build, TAIR7, TAIR8, and TAIR10.
- ③ araGWAB extracts p-values from the GWAS data set.
- ④ araGWAB conducts boosting of GWAS p-values using co-functional links between genes. In addition, araGWAB calculates the accuracy of retrieved reference disease gene set by 100 randomized networks by edge shuffling.

Each job status will be marked as either 'Not Yet', 'Running' or 'Done'.

# RESULT

④ [Click here to download ranked genelist of your GWAB result](#)



## 3. INTERPRETATION OF THE RESULT

- ① This panel shows the information of the parameters used for running araGWAB.
- ② This plot shows the AUC score for FPR 5% (y-axis), indicating how much araGWAB improved the retrieval rate of reference disease gene set with the given log10(P) threshold (x-axis). The plot also shows the baseline performance by GWAS alone and the range of 100 random performances to evaluate effectiveness and significance of the boosting, respectively.

### a. GWAS alone

The baseline represents AUC score of the disease gene candidates prioritized by original GWAS p-value

b. Random networks

The mean and range of two standard deviations of AUC scores for candidate genes with boosting using 100 randomized AraNet V2 (Lee et. al, Nucleic Acids Research Database issue, 2015) by edge shuffling. If the AUC score by araGWAB is higher than that by shuffled networks, as well as araGWAS baseline, novel candidate genes are also likely to be related with corresponding disease.

- ③ To present the candidate genes prioritized by optimal boosting conditions, araGWAB use  $\log_{10}(P)$  threshold for the highest AUC score.
- ④ For the optimal  $\log_{10}(P)$  threshold, users can download prioritized genes with their AGI locus ID, gene symbol, and araGWAB score.

#### 4. EXAMPLES

### Pre-calculated araGWAB

Data obtained from [Atwell et al. Nature 2010](#)

Phenotype	① Reference Genes and GWAS Data	② Result	③ Phenotype Description
As2CFU2	<a href="#">Download</a>	<a href="#">View Result</a>	<i>In planta</i> bacterial growth (number of CFU / leaf area) of the 5 <i>P. viridiflava</i> strains were individually measured as described in Goss and Bergelson 2006
At1CFU2	<a href="#">Download</a>	<a href="#">View Result</a>	<i>In planta</i> bacterial growth (number of CFU / leaf area) of the 5 <i>P. viridiflava</i> strains were individually measured as described in Goss and Bergelson 2006
Bacterial titer	<a href="#">Download</a>	<a href="#">View Result</a>	Bacterial titers of <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000
FLC	<a href="#">Download</a>	<a href="#">View Result</a>	FLC gene expression
FRI	<a href="#">Download</a>	<a href="#">View Result</a>	FRI gene expression
LD	<a href="#">Download</a>	<a href="#">View Result</a>	Days to flowering time under 16hrs daylight, 18°C
LDV	<a href="#">Download</a>	<a href="#">View Result</a>	Days to flowering time under 16hrs daylight, 18°C, vernalized (5wks, 4°C)

- ① Users can download GWAS data sets and reference disease gene set for the nine pre-calculated phenotypes.
- ② Users can see pre-calculated araGWAB results and prioritized genes for all phenotypes.
- ③ The description for each GWAS experiment