

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	sra-toolkit (v2.11.0) was used to collect public sequencing data from NCBI database
Data analysis	All softwares or online tools including R (v4.2.1), r/qlt (v1.66), TASSEL 5 (https://tassel.bitbucket.io/), CRISPR-P (http://crispr.hzau.edu.cn/CRISPR2/), STAR (v2.5.4b), cufflinks (v2.2.1), CleaveLand (v4.5), MEGA (7.0), vcftools (v0.1.16), RNAfold server (http://rna.tbi.univie.ac.at/) and psRNATarget (Schema V2 2017 release), bowtie2 (v2.5.1),Bismark (v0.23.1), Wheeler Aligner program (v0.7.15), Model-based Analysis of ChIP-Seq (MACS2), BiNGO (v3.0.3), ImageJ (v1.53k), clusterProfiler (v4.0) used for data analysis in the manuscript have been cited/referenced.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data are available in the main text, supplemental materials, public databases, or referenced studies. All the raw sequence data generated in this study have been deposited in NCBI database under the BioProject PRJNA876203. The genotypic data of the USDA soybean germplasm collection used for GWAS on pubescence form and leafhopper resistance in Extended Data Fig. 1c-d were downloaded from the SoyBase database (<https://soybase.org/snps/download.php>). The genotypic data of the re-sequenced soybean accessions used for GWAS on pubescence form in Extended Data Fig. 1a-b were downloaded from the Genome Variation Map (GVM) database in BIG Data Center (<http://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000063>). The RNA-seq, sRNA and WGBS data of the 45 highly diverse soybean accessions were download from the Sequence Read Archive (SRA) database in NCBI under Accession Number PRJNA432760 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA432760>). Source data are provided with this paper.

All software used in this study is publicly available as described in the Methods and Reporting summary. Detailed parameters used for analyzing each type of sequencing data have been described in the Method. An in-house Perl scrip used for creating SNP-corrected genomes is available at Zenodo (<https://doi.org/10.5281/zenodo.10801184>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes are indicated in the manuscript, figures or legends. No statistical method was employed to determine sample size; instead, sample sizes for each experiment were optimized based on practical feasibility and the availability of samples. The size of the RIL mapping population is 2,287. The sample sizes of association studies are 668 and 784 for pubescence form and leafhopper resistance, respectively. The sample size for the qRT-PCR experiments is 3 biologically independent samples. 20-30 plants were planted in each plot for each transgenic line or gene editing line (single and double mutants). 7 homozygous triple mutants were obtained by crossing the double mutants. Small RNA data from 45 soybean accessions were used to analyze the sRNA diversity at population level.
Data exclusions	small RNAs less than 17-nt or greater than 25-nt were excluded.
Replication	Three biologically independent samples (with three technical replicates for each sample) were performed in qRT-PCR experiments; Three biologically independent transgenic events and gene editing lines were created for functional validations. All attempts at replication for the qRT-PCR and functional validation were successful. Small RNAs were obtained from the parental lines and a pair of RILs as biologically independent replications. The RLM-RACE assay was repeated at least two times; The Bimolecular fluorescence complementation (BiFC) assay was repeated two times. All attempts at replication for the RLM-RACE and BiFC were successful.
Randomization	The plots of recombinants, transgenic lines, gene edited lines as well as the control lines were planted randomly in the field for phenotyping. qRT-PCR and RLM-RACE were performed to identify differences between genotypes, therefore, sample allocation is not relevant to these

experiments. BiFC and Y2H were performed to detect interaction among specific genes, therefore, sample allocation is not relevant to these experiments.

Blinding

Blinding was performed when phenotyping the recombinants, transgenic plants and gene edited lines. Genotype of the plant was not known when the phenotypic data collection and analysis. qRT-PCR and RLM-RACE were performed to identify differences between genotypes, therefore, blinding is not relevant to these experiments. BiFC and Y2H were performed to detect interaction among specific genes, therefore, blinding is not relevant to these experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used ANTI-FLAG® M2 Magnetic Beads (Sigma, M8823-1ML)

Validation ANTI-FLAG® M2 (sigma): The antibody has been validated by the manufacturer, <https://www.sigmaaldrich.com/catalog/product/sigma/fl804>

Plants

Seed stocks The soybean accession, PI 518671 (Williams 82), used for creating the transgenic lines were originally requested from the USDA soybean germplasm collection (<https://www.ars-grin.gov/>).

Novel plant genotypes Transgenic lines were generated for the ChIP-seq experiments by using the soybean transformation protocol described in the Method section. We fused the coding sequences of Glyma.01G051700 and Glyma.02G110000 from Wm82 with the FLAG epitope, separately, generating transgenic lines that overexpress each of the fused proteins by the 35S promoter.

Authentication Over-expression of the transgenes were validated by qRT-PCR.

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.* All the raw sequence data generated from ChIP-seq have been deposited in NCBI database under the BioProject PRJNA876203 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA876203>). All the called peaks from .BED files were provided in supplementary table 11 and 12.

Files in database submission BioSample: SAMN40263196; Sample name: Glyma.02G110000-Flag-input; SRA: SRS20659442
BioSample: SAMN40263195; Sample name: Glyma.02G110000-Flag-ChIP; SRA: SRS20659441
BioSample: SAMN40263194; Sample name: Glyma.01G051700-Flag-input; SRA: SRS20659440
BioSample: SAMN40263193; Sample name: Glyma.01G051700-Flag-ChIP; SRA: SRS20659439

Genome browser session *(e.g. [UCSC](#))* no longer applicable

Methodology

Replicates 1 replicate for Glyma.01G051700-Flag-ChIP and Glyma.02G110000-Flag-ChIP, only the peaks identified by both experiments were further analyzed.

Sequencing depth	Glyma.02G110000-Flag-input, depth: 8.1X, total reads: 29862028, uniquely mapped reads: 22600135,150bp paired-end Glyma.02G110000-Flag-ChIP, depth: 9.7X, total reads: 35745189, uniquely mapped reads: 27479899,150bp paired-end Glyma.01G051700-Flag-input, depth: 12.3X, total reads: 44938763, uniquely mapped reads: 33425686,150bp paired-end Glyma.01G051700-Flag-ChIP, depth: 11.0X, total reads: 40319731, uniquely mapped reads: 30302209,150bp paired-end
Antibodies	ANTI-FLAG® M2 Magnetic Beads (Sigma, M8823-1ML)
Peak calling parameters	macs2: callpeak -t Glyma.01G051700-Flag-ChIP.bam -c Glyma.01G051700-input-ChIP.bam -f BAM -g 11000000000 -n Glyma.01G051700-peak -B -q 0.01 macs2: callpeak -t Glyma.02G110000-Flag-ChIP.bam -c Glyma.02G110000-input-ChIP.bam -f BAM -g 11000000000 -n Glyma.02G110000-peak -B -q 0.01
Data quality	sequencing quality control were performed by Novogene Corporation Inc. (Sacramento, CA). 8727 and 9732 peaks with enrichment greater than 5 were identified from Glyma.01G051700-Flag-ChIP and Glyma.02G110000-Flag-ChIP, respectively.
Software	Burrows Wheeler Aligner program (v0.7.15) Model-based Analysis of ChIP-Seq (MACS2) (v2.1.0)