

Reporting Summary

Nature Research 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 Research 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

No software was used to collect data.

Data analysis

"Online Method" section:

1. Reads of each rice accession were mapped to reference genome (MSUv7.0) using BOWTIE2 v2.3.2 with default settings. Consecutive steps using Samtools v1.9 and GATK v3.7 were applied for detection of variants.
2. The phylogenetic tree was constructed with using Fasttree. PCA was performed by SNPRelate v0.9.19. FastStructure v1.0 was applied to infer the ancestry of each rice accession.
3. GWAS was performed separately for each cohort by GCTA v7.93.2 with the mixed linear model.
4. The conservation scores and fitcon scores (*p*) of the QTNs were compared to the same type of variants (e.g., UTR, nonsynonymous SNP) annotated by SnpEff v3.6. In addition, we also used SIFT to measure the conservation level for nonsynonymous SNP sites for the rice QTGs.
5. The precise locations of QTNs were determined and anchored on the rice MSU v7.0 reference genome by BLAST (v2.7.1) alignment.
6. Genetic improvement of HHZ was performed by using RiceNavi (v1).

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 Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

"Data availability statement" section:

1. The raw DNA sequencing data of the QTN library are deposited in the GenBank under the bioproject accession number PRJNA623686.
2. A web-based version of RiceNavi is available in the website <http://www.xhhuanglab.cn/tool/RiceNavi.html> (supporting most browsers including Chrome, Firefox, and Safari, but not IE). In this web-based application, all functions in RiceNavi (QTNmap, QTNpick, Simulation and SampleSelect) can be accessed with user-friendly graphical interfaces.

The software package we developed is included in "Code availability statement" section:

3. The source code of RiceNavi is available in both our lab web (<http://www.xhhuanglab.cn/tool/RiceNavi.html>) and the GitHub repository (<https://github.com/xhhuanglab/RiceNavi.git>).

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](https://www.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	"GWAS in eight rice cohorts and QTG effect estimation" of "Online Methods" section: The genotype and phenotype data used was generated from this study and also collected from seven other cohorts. The eight rice cohorts included a total number of 17,376 accessions and 270 trait replicates.
Data exclusions	No samples were excluded in the analysis.
Replication	For the QTN library, phenotyping was conducted on three replicates for each line.
Randomization	Randomization is not involved in this study, because we used all the rice QTN information in public literatures, and we collected the genotypic and phenotypic data of all the rice accessions in the eight GWAS cohorts.
Blinding	Online Methods section: A double-blind test was performed, in which each sample was smelled and replicated two times. The fragrance level was recorded as 1 (non-basmati fragrance) and 2 (basmati-specific fragrance), respectively.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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