LETTER TO THE EDITOR

Population-scale multi-omics analysis reveals diverse phenotypic impacts of lncRNAs in rice

Cell Research (2026) 0:1–4; https://doi.org/10.1038/s41422-026-01247-3

Dear Editor,

Long noncoding RNAs (lncRNAs) are emerging as critical regulators of plant growth and development, and their expression orchestrates and integrates multiple processes underlying vital components of the phenotype, including stress adaptation, aspects of phenotypic evolution, and speciation.[1] In plants, lncRNA-mediated gene regulation remains poorly understood, even in rice, one of the most important staple foods and a model crop species. In this study, we performed strand-specific RNA sequencing of 188 accessions representing different developmental stages of cultivated (*Oryza sativa* ssp. *japonica* and *indica*) and wild rice (*O. rufipogon*). We identified a total of 57,162 lncRNAs in the rice genome, established a comprehensive lncRNA annotation resource, and analyzed the molecular mechanisms by which lncRNAs regulate gene expression and influence key agronomic traits. Subsequent functional validation suggested that the lncRNA *LBR5* may confer resistance to bacterial blight. This study advances our understanding of the role of the noncoding genome in shaping phenotypic diversity and evolutionary adaptation in rice.

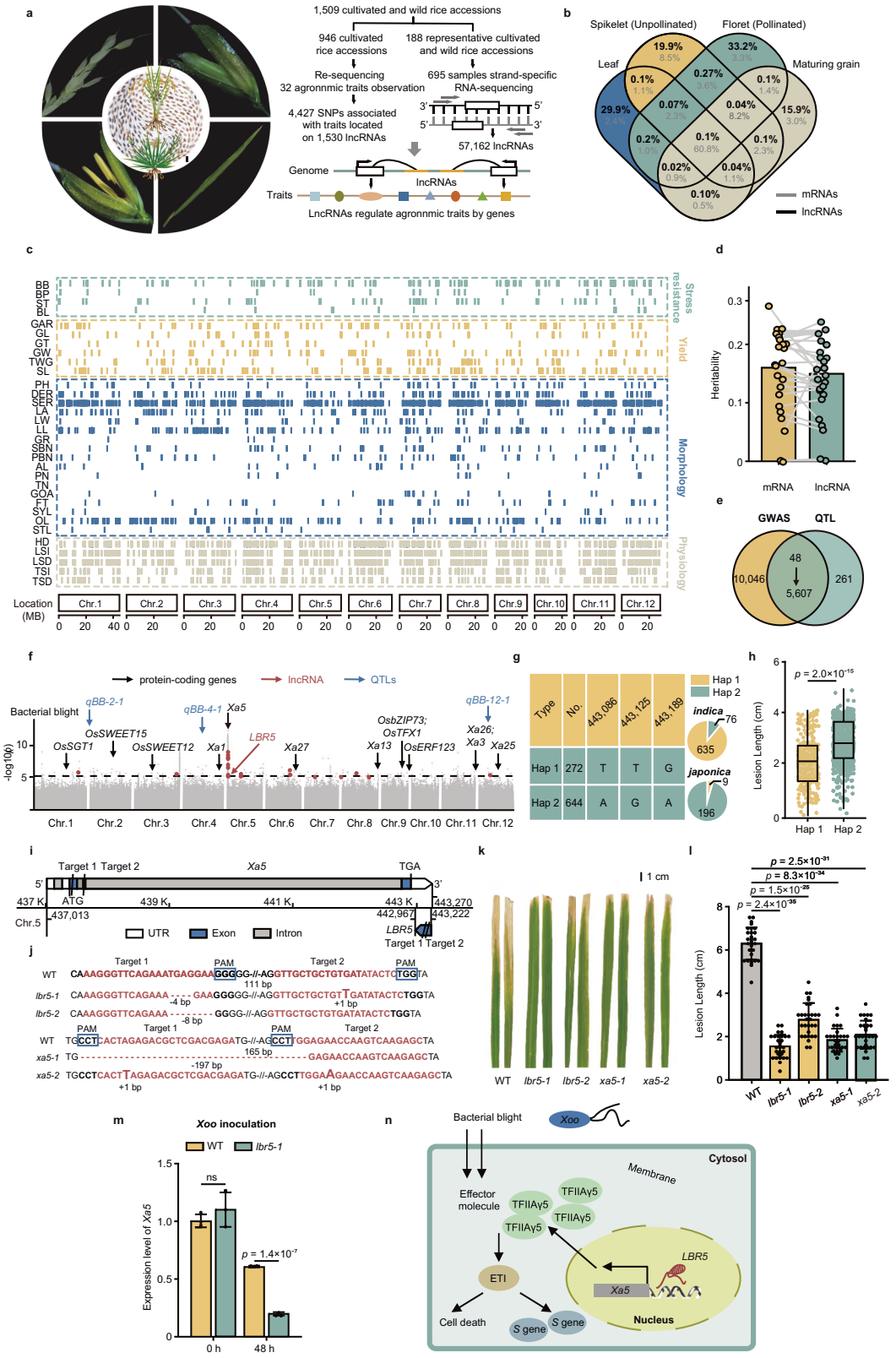
We performed strand-specific RNA sequencing of 695 samples representing four tissues and developmental stages from 188 rice accessions (Supplementary information, Table S1), including cultivated and wild rice (Fig. 1a, b). We identified 57,162 lncRNAs (Supplementary information, Table S2), 34.5% and 41.4% of which overlapped with lncRNAs in the existing databases PLncDB and CANTATAdb, respectively (Supplementary information, Fig. S1a–d and Materials and methods).[2,3] Some previously characterized lncRNAs (e.g., the disease resistance-associated lncRNA *TCONS_00089479*) were detected in our lncRNA database; however, some exhibited distinct transcription start sites and transcript lengths, likely owing to alternative splicing of lncRNAs.[4] We identified conserved vs lineage-/subspecies-specific lncRNAs at the population level using 1509 accessions grouped as *indica*, *japonica*, and the wild ancestor *O. rufipogon*. Selection signatures were evaluated by calculating sliding-window π and F_{st} values, with the top 5% of windows defined as putative sweeps. These putative sweep regions contained 1532 (*indica*/*O. rufipogon*) and 1024 (*japonica*/*O. rufipogon*) lncRNAs, 273 of which were shared. Expression divergence was quantified using DESeq2 ($|\log_2 FC| \geq 2$, FDR ≤ 0.05), identifying 8355 and 8414 differentially expressed lncRNAs, 3435 of which were shared. Jointly, 244 (*indica*/*O. rufipogon*) and 130 (*japonica*/*O. rufipogon*) candidates were supported by both signals, including 21 shared domestication-associated lncRNAs (Supplementary information, Table S3).

lncRNAs serve as raw material for genome evolution, giving rise to new coding (de novo) genes through open reading frame (ORF) acquisition or exon sharing.[5] Using a rigorous screening pipeline, we identified 127 high-confidence de novo genes in cultivated rice from an initial set of 870 protein-coding genes (Supplementary information, Fig. S1e) that showed homology to intergenic lncRNAs in wild rice (Supplementary information, Table S4 and Materials and methods). To test their genetic contribution, we backcrossed *O. rufipogon* (donor) with *O. sativa* A8. In the BC_4F_8 population, approximately 52.8% (67 out of 127) of these lncRNA-derived de novo genes exhibited associations with quantitative trait loci (QTLs) (Supplementary information, Fig. S1f, g). For example, *LOC_Os02g48690* (from lncRNA *Os02Ln0440000*) fell within *qGPP-2-1*, a major QTL for filled grain number per spikelet (Supplementary information, Fig. S1h–j). These findings demonstrate the significant role of lncRNAs in the origin of novel functional genes that influence important agronomic traits.

lncRNAs play important roles in gene regulation at the epigenetic, transcriptional, and post-transcriptional levels, regulating gene expression through encoded micropeptides and various other mechanisms. We developed a pipeline to systematically analyze and predict the functions of lncRNAs and their underlying molecular mechanisms (Supplementary information, Fig. S2). At the pre-transcriptional and transcriptional levels, lncRNAs can regulate protein-coding genes by modulating transcription through *cis*-acting and *trans*-acting functions (Supplementary information, Materials and methods).[6] Our analysis identified 154 lncRNAs that exert potential *cis*-regulatory effects on 141 mRNAs (157 lncRNA-mRNA pairs). For *trans*-regulation, 15,048 lncRNAs targeted 13,927 mRNAs (495,599 lncRNA-mRNA pairs), with 876 lncRNAs and 967 mRNAs showing complementary relationships (Supplementary information, Fig. S2 and Table S5). We classified significant lncRNA-mRNA co-expression pairs (Spearman's ρ, BH-FDR < 0.05) into positive (23,758) and negative (162) groups, reflecting expression directionality rather than direct regulation. Positive associations predominated, likely due to shared transcriptional programs and chromatin states. Integration of RiceENCODE histone marks and ChromHMM (8 states, 200 bp) showed that positive-pair targets were enriched in active/enhancer contexts, whereas negative-pair targets were more often repressive/heterochromatic, providing supportive exploratory evidence (Supplementary information, Table S5).

Analysis of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways revealed that genes with *cis*-regulatory lncRNA relationships were enriched in cellular process, organelle, and organelle part terms, whereas those with *trans*-regulatory interactions were enriched in cellular process, cell, and cell part terms (Supplementary information, Fig. S3a, b and Table S6). In addition, genes with *cis*-regulatory relationships to lncRNAs were predominantly associated with mismatch repair, nucleotide excision repair, and photosynthesis pathways. Genes exhibiting *trans*-regulatory relationships with lncRNAs were associated with pathways such as metabolic processes, phenylpropanoid biosynthesis, and plant–pathogen interaction (Supplementary

Received: 9 June 2025 Accepted: 6 March 2026
Published online: 06 April 2026



information, Fig. S3c, d and Table S6). GO/KEGG enrichment analyses of lncRNA target genes (*cis* and *trans*) revealed significant clustering in stress/defense- and development-related pathways. Targets enriched in the plant–pathogen interaction pathway included *OsRbohB* (*LOC_Os01g25820*) and additional *NBS-LRR* genes (*LOC_Os01g16370* and *LOC_Os01g16390*), supporting a role for lncRNA-associated modules in stress responses (Supplementary information, Table S6). These lncRNAs and their target genes may play important roles in shaping rice traits and provide resources for subsequent lncRNA analyses.

Fig. 1 Characteristics and analysis of lncRNAs in rice. a Overview of methods for analyzing the genetic variation and functional roles of rice lncRNAs. **b** Expression profiles of lncRNAs (bold black) and mRNAs (gray) in four rice tissues. **c** Chromosome-wide distribution of significant SNPs within lncRNA loci associated with 32 agronomic traits in rice. Stress resistance-related traits (4) include bacterial blight (BB), brown planthopper (BP), salt tolerance (ST), and blast (BL); yield-related traits (6) include grain aspect ratio (GAR), grain length (GL), grain thickness (GT), grain width (GW), 1000-grain weight (TGW), and spikelet length (SL); morphology-related traits (17) include plant height (PH), double exposure rate (DER), single exposure rate (SER), leaf angle (LA), leaf width (LW), leaf length (LL), germination rate (GR), secondary branch number (SBN), primary branch number (PBN), anther length (AL), panicle number (PN), tillering number (TN), glume opening angle (GOA), flowering time (FT), style length (SYL), ovary length (OL), and stigma length (STL); physiological traits (5) include heading date (HD), light-sensitivity index (LSI), light-sensitive days (LSD), temperature-sensitivity index (TSI), and temperature-sensitive days (TSD). Lines on the chromosomes indicate SNP–GWAS association peaks identified at a significance threshold of $P \le 1 \times 10^{-6}$. **d** Heritability of traits contributed by variants in lncRNAs and mRNAs. **e** Venn diagram showing the overlap of significantly associated genomic regions identified through GWAS and QTL analyses. **f** Genome-wide Manhattan plot for bacterial blight based on a GWAS of 946 natural rice accessions. The dashed line indicates the significance threshold (–log10(P) = 5.1). Genes with known or putative roles in bacterial blight resistance are indicated by black arrows; significant SNPs within lncRNA loci are marked with red dots; QTL regions associated with bacterial blight are highlighted with blue arrows. **g** *LBR5* haplotype analysis, showing nucleotide differences between Hap1 ($n = 272$ accessions) and Hap2 ($n = 644$ accessions) at three significant SNP sites. **h** Statistical comparison of lesion length between Hap1 and Hap2. Boxes represent the 25th to 75th percentiles, and the median is indicated by a horizontal line. The P values from Student's t-test are provided. **i** Schematic diagram of the genomic location, transcription direction, and target sites of *LBR5* and *Xa5*. The slash indicates the CRISPR/Cas9 target site. **j** Sequence diagram of *LBR5* and *Xa5* mutants showing their respective sequence modifications. The protospacer adjacent motif (PAM) is highlighted in blue. **k** Phenotypic comparison of leaf lesion lengths between WT and knockout mutants after inoculation with *Xoo*. Scale bar, 1 cm. **l** Statistical comparison of lesion lengths between WT and knockout mutants after *Xoo* inoculation. Data are presented as mean ± S.D., with error bars representing the standard deviation and dots indicating individual samples ($n = 30$ per group). **m** Relative expression levels of *Xa5* in WT and *lbr5-1* mutants as determined by RT-qPCR. Dots represent individual expression values from three biological replicates, and ns indicates no significant difference. **n** Schematic model of the potential mechanism by which *LBR5* may *cis*-regulate *Xa5* expression in response to bacterial blight infection. Effector-triggered immunity (ETI): plants perceive effector-induced changes through resistance proteins, mounting more rapid and robust defense responses.

At the post-transcriptional level, cytoplasm-localized lncRNAs regulate mRNA expression by competitively binding to microRNA (miRNA) sequences (Supplementary information, Materials and methods). For example, *IPS1* acts as a sponge for *miR399* via a partially complementary sequence, thereby preventing complete *miR399*-mediated silencing of the *PHO2* gene.[7] Co-expression analysis of lncRNAs and mRNAs based on the miRNA database miRBase (https://www.mirbase.org/) identified 273 lncRNA-miRNA-mRNA regulatory networks among 139 lncRNAs, 79 miRNAs, and 206 mRNAs (Supplementary information, Fig. S2 and Table S7). We also observed partial sequence complementarity between *miR399* and both the lncRNA *Os01Ln0482700* and the blast-resistance gene *LOC_Os01g57280*, suggesting a potential competing endogenous RNA (ceRNA)-like regulatory mechanism.

Some lncRNAs contain small open reading frames (sORFs) encoding micropeptides that function in developmental processes.[8] We used ribosome footprinting (RF) technology combined with RNA sequencing data to search for lncRNAs that encode micropeptides in noncoding regions of the rice genome (Supplementary information, Materials and methods), and we identified 44 sORFs within 39 lncRNAs that were expressed simultaneously across four developmental stages (Supplementary information, Fig. S2). These sORFs averaged 45 bp in length and encoded peptides of 10–20 amino acids. The associated lncRNAs were distributed across all chromosomes except for chromosomes 6 and 11. Through sequence alignment analysis, we observed sequence homology among some sORFs and rigorously narrowed them down to 11 core sORFs with independent coding potential. Among these, two sORFs encoded micropeptides with significant similarity to micropeptides from *Gossypium arboreum* in the NCPbook database of non-canonical peptides (https://ncp.wiki/ncpbook/) (Supplementary information, Table S8). Our study thus reveals a robust transcriptomic landscape for the regulatory analysis of lncRNAs in rice and enhances our understanding of their multifaceted roles in gene regulation and expression.

To assess the effects of lncRNAs on key agronomic traits in rice, we analyzed 946 genetically diverse rice accessions by phenotyping and genome resequencing. Through genome-wide association studies (GWAS), we identified 81,680 SNPs and 851 QTLs associated with 32 agronomic traits (significance threshold, $P \le 1 \times 10^{-6}$) (Supplementary information, Fig. S4a). A total of 4427 SNPs were associated with traits that overlapped with 1530 lncRNAs (Fig. 1c; Supplementary information, Table S9). Among these, 2781 SNPs overlapped with 519 pleiotropic lncRNAs affecting multiple traits. For example, the SNP Chr3_5170840, located in *Os03Ln0094100*, had phenotypic variance explained (PVE) values of 39.93% for grain aspect ratio and 39.63% for grain width. Another lncRNA, *Os03Ln0313600*, contained three trait-associated SNPs: Chr3_18851267 (double exposure rate, PVE 13.37%), Chr3_18870886 (flowering time, 7.41%), and Chr3_18856999 (temperature-sensitivity index, 1.18%) (Supplementary information, Table S10). Our analysis indicated that variants in lncRNAs contributed to the heritability of traits at a level nearly equal to that of mRNAs on average ($h^2 = 0.15$ and 0.16, respectively), demonstrating that variants in non-coding regions can affect gene expression and thus alter important agronomic traits (Fig. 1d). We used a population of chromosome segment substitution lines (CSSLs) with cultivated rice A8 as the recipient parent and wild rice as the donor parent to identify 456 potential QTLs related to 28 agronomic traits (16 traits overlapped with GWAS) (Supplementary information, Fig. S4b). For overlapping traits, we identified 15,653 SNPs in 309 QTLs (Fig. 1e), including 48 major-effect QTLs containing 5607 SNPs. We also found 23,564 lncRNAs located in 231 QTL regions. One QTL, *qGL-7-1*, was narrowed down to a 0.38-Mb region. CSSLs X152 and X216, which carried this region, showed significantly longer grains than A8 (Supplementary information, Fig. S4c–g). Within this region, a SNP in lncRNA *Os07Ln0401000* defined two haplotypes, and Hap2 was associated with longer grains (Supplementary information, Fig. S4h).

Xanthomonas oryzae pv. *oryzae* (*Xoo*) causes bacterial blight, a major threat to rice production. Several significant SNP associations for bacterial blight ($P \le 1 \times 10^{-5.1}$) were identified with multiple overlapping QTLs (Fig. 1f). Notably, the lncRNA *Os05Ln0007600* was located near *Xa5* on chromosome 5, which encodes the gamma subunit of transcription factor IIA (TFIIAγ) and contributes to bacterial blight resistance.[9] Three significant SNPs formed two major haplotype groups in the *LBR5-Xa5* genomic block (Fig. 1g). Rice accessions with Hap1 showed shorter lesion lengths than those with Hap2, and haplotype frequencies differed between *indica* and *japonica* (Fig. 1h). *LBR5* is located in the 3′ UTR region of *Xa5* and is predicted to influence *Xa5* expression (Fig. 1i). Both *LBR5* and *Xa5* were significantly downregulated after *Xoo* inoculation (Supplementary information, Fig. S5a, b). To confirm the role of *LBR5* in resistance, we created two independent knockout lines for

each gene and analyzed these lines (*lbr5-1*, *lbr5-2* for *LBR5* and *xa5-1*, *xa5-2* for *Xa5*) in the T2 generation (Fig. 1j; Supplementary information, Fig. S5c, d). The lesion lengths of *lbr5* mutants were similar to those of *xa5* mutants, and both were significantly shorter than those of the wild-type (WT) (Fig. 1k, l). *Xa5* expression was significantly reduced in *LBR5* knockout lines after inoculation, suggesting that *LBR5* may influence *Xa5* during resistance to bacterial blight through a *cis*-regulatory function (Fig. 1m). In addition, Cleavage Under Targets and Tagmentation (CUT&Tag) analysis showed reduced enrichment of Pol II (0.73; *ns*) and H3K4me3 (12.96; $P = 5.0 \times 10^{-4}$) across the gene body and promoter region of *Xa5* in *lbr5-1* compared with WT (Pol II: 1.27; H3K4me3: 17.26) after *Xoo* inoculation (Supplementary information, Fig. S5e). Collectively, these findings provide evidence supporting the GWAS results and functional predictions described above and offer new insights and perspectives into the regulatory function of rice lncRNAs in resistance to bacterial blight (Fig. 1n).

In conclusion, we comprehensively characterized rice lncRNAs through genome resequencing, analysis of de novo gene evolution, and trait association mapping, revealing their expression patterns, molecular interactions, and relevance to agronomic traits. These findings informed an upgraded LncPheDB database, establishing a complete paradigm from multi-omics discovery to functional validation. This resource enables targeted lncRNA editing via CRISPR/Cas9, opening new avenues for improving yield and disease resistance.

Ge Gao[1,2,3,9], Danjing Lou[1,4,5,9], Yunpeng Li[1,4,6,9], Chao Zhang[2,3], Baolin Kan[2,3], Hao Wen[4], Siyu Wen[2], Jun Shu[2,4,6], Wenlong Guo[1,2], Weihua Qiao[1], Qingwen Yang[1], Youlin Peng[2], Kenneth M. Olsen[7]✉, Qian Qian[1,2,3]✉ and Xiaoming Zheng[1,4,8]✉

[1]*State Key Laboratory of Crop Gene Resources and Breeding, Key Laboratory of Grain Crop Genetic Resources Evaluation and Utilization Ministry of Agriculture and Rural Affairs, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China.* [2]*Wild Rice Germplasm Resources Innovation Team, Yazhouwan National Laboratory, Sanya, Hainan, China.* [3]*College of Plant Science, Huazhong Agricultural University, Wuhan, Hubei, China.* [4]*National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya, Hainan, China.* [5]*Institute of Botany, Chinese Academy of Sciences, Beijing, China.* [6]*China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, Zhejiang, China.* [7]*Biology Department, Washington University, St. Louis, MO, USA.* [8]*International Rice Research Institute, Metro Manila, Philippines.* [9]*These authors contributed equally: Ge Gao, Danjing Lou, Yunpeng Li.* ✉*email: kolsen@wustl.edu; qianqian188@hotmail.com; zhengxiaoming@caas.cn*

REFERENCES

1. Goede, O. M. et al. *Cell* **184**, 2633–2648 (2021).
2. Jin, J. J. et al. *Nucleic Acids Res.* **49**, D1489–D1495 (2020).
3. Szczesniak, M. W. *Plant Cell Physiol.* **57**, 1–7 (2016).
4. Zhang, Y. et al. *Viruses* **12**, 551 (2020).
5. Ruiz-Orera, J. et al. *PLoS Genet.* **11**, e1005721 (2015).
6. Palos, K. et al. *Plant Cell* **35**, 1762–1786 (2023).
7. Franco-Zorrilla, J. et al. *Nat. Genet.* **39**, 1033–1037 (2007).
8. Wang, S. X. et al. *Mol. Plant* **13**, 1078–1093 (2020).
9. Jiang, G. H. et al. *Mol. Genet. Genomics* **275**, 354–366 (2006).

ACKNOWLEDGEMENTS

This work was supported by the Hainan Province Science and Technology Special Fund (ZDYF2022XDNY260), the National Key R&D Program of China (2021YFD1200101), the Sanya Yazhouwan Science and Technology City (SKJC-2023-02-001, SCKJ-JYRC-2023-47), the National Natural Science Foundation of China (32261143465, 32350710198), the Crop Germplasm Innovation and Creation Program (32188102), the Nanfan special project, CAAS (YBXM2404, YBXM2403), the Hainan Province Nature and Science Fund (2021JJLH0075), the Hainan Province Science and Technology Innovation (KJRC2023A01), and the Hainan Province International Scientific and Technological Cooperation Talent and Exchange Project (Foreign Expert Program) Plan (G20241024007E).

AUTHOR CONTRIBUTIONS

X.Z. and Q.Q. designed the study. K.M.O. co-wrote the manuscript. G.G. conducted analyses, visualized data, performed most of the experiments, and wrote the manuscript. D.L. performed RNA-seq and genome resequencing assays and conducted GWAS analyses. Y.L. conducted de novo gene analysis and contributed to data visualization. C.Z. contributed to GWAS analysis. B.K. and S.W. contributed to lncRNA regulatory mechanism analysis. H.W. and W.G. contributed to the prediction of lncRNA function. J.S. contributed to CSSL identification and QTL mapping. W.Q., Q.Y., and Y.P. contributed to the writing and editing of the manuscript. All authors read and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41422-026-01247-3.

Correspondence and requests for materials should be addressed to Kenneth M. Olsen, Qian Qian or Xiaoming Zheng.