Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication

X. M. Zheng1, J. Chen1, H. B. Pang2, S. Liu1, Q. Gao1, J. R. Wang1, W. H. Qiao1, H. Wang3, J. Liu1*, K. M. Olsen4*, Q. W. Yang1*

Genomes carry millions of noncoding variants, and identifying the tiny fraction with functional consequences is a major challenge for genomics. We assessed the role of selection on long noncoding RNAs (lncRNAs) for domestication-related changes in rice grains. Among 3363 lncRNA transcripts identified in early developing panicles, 95% of those with differential expression (329 lncRNAs) between Oryza sativa ssp. japonica and wild rice were significantly down-regulated in the domestication event. Joint genome and transcriptome analyses reveal that directional selection on lncRNAs altered the expression of energy metabolism genes during domestication. Transgenic experiments and population analyses with three focal lncRNAs illustrate that selection on these loci led to increased starch content and grain weight. Together, our findings indicate that genome-wide selection for lncRNA down-regulation was an important mechanism for the emergence of rice domestication traits.

INTRODUCTION

The domestication of plants from their wild progenitors has provided an important model system for investigating the genetics and evolution of complex traits (1, 2). Selection during domestication generates major morphological changes over evolutionarily short periods of time, yielding vital sources of food, medicine, and energy (1, 2). Over the past decade, large-scale transcriptomics studies have documented that genome-wide alterations of gene expression play important roles in many phenotypic changes during crop domestication (3–6). Deep sequencing of transcriptomes has further revealed that the number of RNAs produced in organisms far exceeds the number of protein-coding genes (7). Among these RNAs, long noncoding RNAs (lncRNAs) are especially numerous in plants and a major regulator of the transcriptional process. On the basis of their genomic positions relative to adjacent protein-coding genes, lncRNAs can be classified into long intergenic noncoding RNAs (lincRNAs), natural antisense transcripts (NATs), and intronic noncoding RNAs (incRNAs) (5). lncRNAs not only play important roles in plant development and adaptation to environmental change but also are essential molecular drivers of lineage-specific morphological evolution (5, 8, 9). Several studies of vertebrates and invertebrates have already pointed to the power of complete genome comparisons for understanding the potential function of lncRNAs during species evolution (10–12). However, the role of changes at lncRNA loci for morphological and developmental changes during crop domestication remains almost entirely uncharacterized.

In this study, we combined large-scale genome resequencing, transcriptomic analyses, and transgenic experiments to assess the role of selection on lncRNAs in rice domestication. Asian rice consists of two subspecies, Oryza sativa ssp. indica and O. sativa ssp. japonica, which were domesticated in Southern Asia from the wild species Oryza rufipogon about 10,000 years ago (13, 14). Indica varieties are grown primarily in tropical regions, including South and Southeast Asia, while japonica cultivation occurs primarily in subtropical and temperate climates. Population genetic studies using diverse genetic markers [including simple sequence repeat (SSR), transposons, neutral gene sequences, and whole-genome single-nucleotide polymorphisms (SNPs)] have indicated that japonica rice is derived from O. rufipogon populations in southern China (13, 15, 16). In contrast, indica rice is proposed to be derived from japonica × wild rice hybridization that occurred as early crop varieties that were introduced into South and Southeast Asia (15). Whereas indica and wild rice have a broadly sympatric contemporary distribution in Southern Asia, allowing for both historic and contemporary gene flow, japonica rice is largely allopatric with O. rufipogon except for some regions of Southeast Asia, where tropical japonica varieties are grown. Because of its clearer phylogenetic relationship to wild rice and the decreased influence of post-domestication crop-wild introgression, japonica rice can provide a superior research system to indica for studying the genomic impact of domestication (15, 17). Here, we used comparative population genomics between japonica rice and accessions of its closest wild relative in China to assess lncRNA expression pattern differences and the role of lncRNAs in domestication, particularly as related to improved grain quality.

RESULTS

Genome-wide identification and characteristics of lncRNAs

We first surveyed genetic diversity in a collection of 858 Oryza accessions, representing wild rice (O. rufipogon) sampled across its geographical distribution and all major variety groups within domesticated O. sativa (Fig. 1, A and B). From these, we identified a representative set of domesticated rice varieties (10 O. sativa ssp. japonica accessions) and closely related wild rice (nine O. rufipogon accessions) for lncRNA transcriptome analyses (table S1 and fig. S1). Details on criteria for accession selection are provided in the Materials and Methods. We collected early developing panicle tissue, which plays a critical role in determining the grain size, grain number, grain shattering, and grain quality (1, 18, 19), to systematically characterize the lncRNA transcriptome and explore its significance for rice domestication traits.
Whole transcriptome strand-specific RNA sequencing (ssRNA-seq) using the Illumina HiSeq 2500 platform yielded >1.9 billion paired-end reads (~100 million per sample); approximately 84 to 93% of filtered reads could be aligned to the Nipponbare Rice Genome Annotation Project (RGAP) 7.0 reference genome sequence (fig. S2, A to E). Average unique mapping rates ranged from 71 to 86% for *O. japonica* accessions and 74 to 82% for wild accessions (fig. S2, B and C). Only reads that uniquely mapped to the reference genome were used for IncRNA expression analysis. We calculated the mapping rate of *O. japonica* and *O. rufipogon* reads to the reference genome (fig.S2F) and found no significant difference in noncoding regions ($P = 0.18$), which suggests that any impact of mapping bias is trivial in our analyses.

To identify IncRNAs, we used a modified computational pipeline with three key steps (fig. S2A) (19): assembling, basic filtering, and protein filtering. A set of 3363 high-confidence IncRNA transcripts (mapping to 2873 genomic loci) was thereby obtained from a total of 9296 unannotated transcripts (fig. S2G). These IncRNAs could be grouped into three categories based on their positional relationships to nearby protein-coding genes (Fig. 1C): 2517 lincRNAs, 200 IncRNAs, and 646 NATs. The rice IncRNA transcripts were relatively short and have lower conservation scores compared with those of mRNAs (fig. S2, I to K), which is consistent with the known IncRNA features in rice (8, 19), *Arabidopsis* (20), and humans (21). We further verified the RNA-seq expression levels of a subset of IncRNAs showing different FPKM (fragments per kilobase of transcript per million mapped reads) values in the 19 accessions using quantitative reverse transcription polymerase chain reaction (qRT-PCR; table S2). We found a strong correlation between the fold-change ratios derived from RNA-seq and those obtained using qRT-PCR ($P = 2 \times 10^{-4}$; Fig. 1D and fig. S3), confirming that the RNA-seq data provide an accurate assessment of gene expression variation. The IncRNA genes were expressed at lower relative levels than mRNAs ($t$ test, $P < 2.2 \times 10^{-16}$; fig. S2H), and the expression variation of IncRNAs is substantially greater than that of mRNAs (Fig. 1E).

### Divergence of IncRNA expression between *japonica* and wild rice

To further assess IncRNA expression variation across samples, we calculated the Pearson correlation coefficients between all pairs of accessions (fig. S3A) and generated a principal variance components plot of expression variation among the 19 accessions based on equivalently development tissue (Fig. 1F). These results indicate clear differences in IncRNA expression between domesticated *japonica* rice and its wild relative. Next, we performed differential expression analysis for IncRNAs using a generalized linear model...
Effects of directional selection on differentially expressed IncRNAs

If selection during domestication has led to differences in IncRNA expression between wild and cultivated rice, then an expected pattern would be that within-species expression variation should be low for cultivated rice and for wild rice accessions but with a significant difference in expression between the two groups. To test this hypothesis, we used a rank-based method to identify the IncRNA loci showing the greatest evidence of selection. We detected that most of the 329 differentially expressed IncRNAs (64%, 208 loci) exhibited signatures consistent with evolution under selection (see Materials and Methods). All of these IncRNAs showing signatures of selection are expressed at lower levels in japonica rice than in wild rice.

As a complementary method to test for selection on the IncRNAs, we also examined the patterns of sequence diversity in the set of IncRNA loci with significantly reduced expression levels. We specifically investigated whether the promoter regions of these loci [estimated as the 200–base pair (bp) upstream region immediately flanking the transcription start site] and IncRNA regions showed significant differences between japonica and wild rice in the following parameters: nucleotide diversity (π), $F_{ST}$, $D_{xy}$, and the proportion of fixed SNPs. For nucleotide diversity, the promoter and IncRNA regions of the putatively selected IncRNAs showed a significantly greater drop in diversity between the crop and wild species than the nonputatively selected loci; the ratio between wild rice (πw) and japonica rice (πj) for the putatively selected IncRNAs (12.1) was approximately twice the value of the ratio for nonputatively selected IncRNAs (5.7) (Fig. 2C). Similarly, the value of $F_{ST}$ and $D_{xy}$ between japonica and wild rice for putatively selected IncRNAs (average $F_{ST} = 0.68; D_{xy} = 0.0132$) was significantly higher than that of nonputatively selected IncRNAs (average $F_{ST} = 0.37; D_{xy} = 0.0061$) based on a rank-based method (Fig. 2, D and E). The statistical significance of these diversity and differentiation differences was also confirmed by permutation tests (Fig. 2, C to E). For fixed SNPs, 8.4% of SNPs in japonica and 2.3% in wild rice were fixed in the promoter and IncRNA regions of putatively selected IncRNAs compared to 5.2% of SNPs in japonica and 1.9% in wild rice in the nonputatively selected IncRNAs. The average number of fixed SNPs is 4.9% in japonica and 2.0% in wild rice based on genome variations. Thus, for the crop, the proportion of fixed SNPs at putatively selected loci is >50% greater than the proportion for nonputatively selected loci and randomly selected genomic loci. In addition, coalescent analyses were conducted to test whether or not the observed differences in diversity pattern can be explained solely as an artifact of demographic bottlenecks during domestication. We calculated the Akaike information criterion (AIC) for each model, and the five smallest AIC values correspond to the bottleneck plus selection model ($s > 0$) (fig. S4). This result indicates that a model assuming a bottleneck plus selection ($s > 0$) fits significantly better than a model assuming a bottleneck alone ($s = 0$). Together, these sequence diversity analyses suggest that there has been selection on IncRNAs during the history of japonica rice cultivation.

Functional roles of differentially expressed IncRNAs

We then therefore examined Gene Ontology (GO) data to test whether the target genes of the 208 significantly down-regulated IncRNAs with signatures of directional selection might be involved in phenotypic changes associated with domestication. The predicted target genes include 2812 cis-regulated and 1642 transregulated genes. By GO enrichment analysis, we found that nine GO functional categories were significantly overrepresented for cis-regulated target genes, and six were significantly overrepresented for transregulated target genes ($P < 0.01$; Fig. 2, F and G). The enriched GO categories include carbohydrate transport and reserve-related function. Enrichment of these categories suggests that changes in the transport and accumulation of carbohydrates may be important factors in IncRNA-mediated phenotypic evolution during rice domestication (Fig. 2, F and G).

To further explore the molecular functions of the putatively selected IncRNAs, we used transgenic experiments to overexpress wild rice IncRNA alleles in a japonica cultivar (Nipponbare) carrying silenced IncRNA alleles. Because previous studies have suggested that IncRNAs with high expression levels are more likely to act in trans than in cis (19, 20), we selected 10 IncRNAs with high expression in O. rufipogon that show significantly lower expression in japonica rice and molecular signatures of domestication-related selection (table S3). By phenotype screening in T0 progeny with the transgene-negative plants as controls, we found that plants expressing three wild rice IncRNA alleles ($LncRNA_{2308}$, $LncRNA_{1267}$, and $LncRNA_{1631}$; Fig. 3, A to C) showed changes in phenotypes for grain traits, specifically endosperm chalkiness, seed weight, structure of starch granules, and pericarp color (Fig. 3, D and E, and table S4). These results were confirmed by further examination of 30 T2 transgenic plants of three T1 families with single-copy transgenes. All of the transgenic-positive plants showed obviously higher grain chalkiness and lower 1000 grain weight compared to the transgene-negative plants based on two-sided Wilcoxon rank sum test (Fig. 3, F to J, and table S4); these are both phenotypes associated with wild rice. To confirm that the grain phenotypes observed in the transgenic experiments are characteristic of differences between wild and cultivated rice, we phenotyped the 19 accessions of wild and japonica rice (fig. S5). Consistent with the transgenic experiment, japonica rice varieties showed a 75% reduction in grain chalkiness rate, a 116% increase in 1000 grain weight, and a 26% increase in starch content compared to wild rice (fig. S5). These results strongly indicate that these three focal IncRNAs are involved in the regulation of multiple grain traits in rice.

Because IncRNAs can simultaneously act both in trans and in cis to generate differential gene expression patterns among populations, we also compared the expression level of the three focal IncRNAs and the genes immediately flanking these loci using qRT–PCR. All three IncRNAs were distinctly up-regulated in transgene-positive
Fig. 2. IncRNA expression and genome sequence divergence between domesticated and wild rice. (A) Volcano plot of IncRNA expression level in *O. rufipogon* and *O. sativa* ssp. *japonica*. The blue, green, and red dots indicate the following, respectively: nondifferentially expressed IncRNAs, up-regulated IncRNAs, and down-regulated IncRNAs in *japonica* rice as compared with *O. rufipogon*. Black horizontal lines indicate a $-\log_{10}$ adjusted $P$ value of 1.30, corresponding to an adjusted $P$ value of 0.05. (B) Chromosomal distribution of differentially expressed IncRNAs between *O. rufipogon* and *japonica* rice. The red and blue lines represent down-regulated and up-regulated IncRNAs of *japonica* as compared with *O. rufipogon*. The gold and sky blue bars below chromosomes indicate the locations of nine genomic regions that are enriched for differentially expressed IncRNAs (table S3). The gold bars indicate clusters that overlap with domestication sweep regions identified in rice genome scans by Huang et al. (15). (C) Ratio of the nucleotide diversity for upstream regulatory regions (2000 bp) and IncRNA regions between wild rice ($\pi_w$) and cultivated rice ($\pi_j$), comparing IncRNAs with signatures of directional selection (DS) and nondirectionally selected IncRNAs (non-DS). (D to E) $F_{ST}$ and $D_{xy}$ between domesticated and wild rice at directionally selected (DS) and nondirectionally selected (non-DS) IncRNAs. Data in (C) to (E) are presented as means ± SEM, and red lines indicate boundaries for statistical significance in permutation analyses. (F and G) GO enrichment statistics of cis-regulated (F) and transregulated (G) target genes of the differentially expressed IncRNAs between domesticated and wild rice. The x axis and y axis represent GO items and enrichment factor. The size of dots indicates the number of target genes, and the color of dots indicates the range of q value.
Fig. 3. Transgenic analyses of phenotypes for three focal lncRNAs. (A to C) The top of each panel shows the genomic location and structure of annotated genes in the vicinity of three focal lncRNAs. Exons and introns are depicted as rectangles and lines, respectively. Coding regions, untranslated regions, and lncRNAs are shown in black, white and gray, respectively. (D and E) Grain phenotype of wild rice, transgene-negative plants, and transgene-positive plants for LncRNA_2308, LncRNA_1267, and LncRNA_1631. (F to J) LncRNA expression pattern, thousand seed weight, and starch content of wild rice, transgene-negative plants, T2308(+), T1267(+), and T1631(+). (K) Comparative the expression pattern of gene around three focal lncRNAs between transgene-positive and transgene-negative plants for LncRNA_2308, LncRNA-1267, and LncRNA_1631. **P < 0.01 and ***P < 0.001, by two-sided Wilcoxon rank-sum test.
plants compared to transgene-negative plants (Fig. 3, F to H). Measurement of the expression level of genes located near these lncRNAs showed that transgene-positive plants for each of the lncRNAs showed significant up-regulation for a gene located near it (Fig. 3K). Notably, one of the nearby genes, LOC_Os05g03040, encodes an AP2/EREBP family transcription factor and is a known regulatory factor in the rice starch biosynthesis pathway (24). Thus, lncRNA expression variation could potentially affect rice grain starch characteristics through regulatory control of this transcription factor.

**lncRNA regulation of gene expression contributes to rice grain quality traits**

To further explore the downstream regulatory effects of the three focal lncRNAs, we carried out transcriptome analysis for the 18 complementary DNA (cDNA) libraries derived from the panicles of the overexpression lines of LncRNA_2308, LncRNA_1267, LncRNA_1631, and from wild-type (WT) plants (table S5). Our analysis identified 351, 320, and 336 genes with significantly altered expression in the three overexpression lines compared with those in WT, respectively. Among the genes showing expression differences in the lncRNA transgene-positive plants, we found significant expression effects in the amylose synthesis gene Waxy (LOC_Os06g06560), which encodes granule-bound starch synthase and is a major determinant of rice grain starch composition (table S5). Waxy expression is down-regulated in lncRNA overexpression lines. This lncRNA-mediated regulation may have contributed to the altered starch content that is observed in transgenic-positive plants and is characteristic of differences between wild and cultivated rice (Fig. 3, D and E, and table S4).

Besides the altered expression of the starch biosynthesis pathway, the transcriptome of LncRNA_2308 transgene-positive plants also showed significantly increased expression of the gene LOC_Os04g10350, which encodes the 2OG-Fe(II) oxygenase domain that constitutes the conserved domain of plant anthocyanin synthetase (table S5). This altered gene expression is consistent with the greater pericarp pigmentation observed in the transgene-positive plants (Fig. 3D) and may play a role in grain pigmentation differences between wild and cultivated rice. Note that the constructs carrying lncRNA sequences in these overexpression lines were randomly inserted in the genomes. From this and the observed phenotypes, we speculate that the lncRNAs may regulate their downstream genes in trans.

We next sequenced the three focal lncRNAs and their promoters in 239 cultivated and wild rice accessions. On the basis of the nucleotide polymorphisms, we could divide the sequences of the cultivated and wild accessions into five distinct groups for LncRNA_2308 and LncRNA_126 and four groups for LncRNA_1631 (Fig. 4, A to C). Most of the japonica varieties formed a monophyletic clade for each lncRNA gene (cluster 2 for LncRNA_2308, cluster 4 for LncRNA_1267, and cluster 4 for LncRNA_1631; Fig. 4, A to C). The japonica clusters showed significantly lower lncRNA expression levels and a lower grain chalkiness rates than other clusters (Fig. 4, D to I). Furthermore, we compared the diversity between cultivated and wild rice and performed tests for deviations from neutrality. The diversity test showed that, on average, the japonica cluster sequences exhibited much lower nucleotide diversity (\( \pi = 0.0002 \) for LncRNA_2308, \( \pi = 0.0005 \) for LncRNA_1267, and \( \pi = 0.0011 \) for LncRNA_1631) than other cultivated rice groups and wild rice (Fig. 4J). These sequences are also characterized by significantly negative Tajima’s D values at all three loci (Fig. 4K), consistent with deviations from neutral equilibrium. Together, these results demonstrate that the downstream expression effects that can be attributed to these lncRNAs have had a significant impact on multiple rice grain traits and that these effects were likely selected on during japonica domestication.

**DISCUSSION**

An important challenge in understanding the regulatory organization of the genome has been to identify noncoding regions that are functional and to characterize their roles in determining phenotypes and patterns of morphological evolution (25). In the context of domestication, the potential importance of noncoding regions for the emergence of domestication-related traits has been suggested by genome scans; studies in rice (9), soybean (26), and maize (3) have identified numerous putative targets of selection in regions devoid of protein-coding genes. Among the elements found in noncoding regions, the lncRNAs have come to be recognized as pervasive in plant genomes and important regulators of gene expression (19, 21).

In plants, they have been shown to be key regulators of flowering (27), male sterility (28), nutrient metabolism (29), fruit development (26), and fruit ripening (30). However, previous studies have so far been confined to several individual transcriptomes, which do not allow for population comparisons, or to datasets compiled from public expressed sequence tag databases, which generally lack transcript direction information and annotations for monoexonic transcripts and intragenic lncRNAs. In the present study, the breadth and depth of our data (19 accesses and >260 Gb of sequence, respectively) have enabled sensitive detection and robust quantification of lncRNA transcription, providing validation of previously identified lncRNAs and new information on previously undescribed loci.

One key finding of our study is that 11.5% of lncRNAs in early developing rice panicles show significant differential expression between the wild progenitor and crop: the large majority of these are down-regulated (Fig. 2A). This level of differential expression is higher than that has been observed in previous studies, which have compared whole transcriptome differences between wild species and their domesticated or weedy relatives. For example, using whole-genome transcriptomic surveys in maize and its progenitor teosinte, Swanson-Wagner et al. (31) found that about 3.3% of 18,242 profiled genes showed significant expression differences in the domesticate. In another study that considered the entire transcriptome, Lai et al. (32) found that 5% of the total genes showed significantly differential expression between wild and weedy sunflowers (Helianthus annuus). The higher proportion of differentially expressed loci observed in the present study suggests that lncRNAs may play a disproportionately important role in regulatory change. lncRNAs are known to play a major role in modulating and fine-tuning patterns of gene expression during development. For example, in our study, three lncRNAs were found to affect seed starch content and accumulation by regulating Waxy. The high proportion of differentially expressed genes in the present study may also reflect our specific focus on young panicle tissue, which represents a developmental stage where patterns of gene expression would be especially likely to play an important role in the generation of domesticated grain traits.

Beyond documenting differential expression, an additional key advance in the present study is our finding that lncRNAs were likely targets of artificial selection during japonica rice domestication. Several lines of evidence suggest that at least 6% of the lncRNAs that show differential expression (208 loci) bear molecular signatures of selection. These include the following: (i) significant differences in
expression between wild and cultivated rice coupled with significant within-group homogeneity (fig. S3, F and G), (ii) differential loss of nucleotide diversity in lncRNAs between wild and cultivated rice (Fig. 2C), (iii) differential elevation of $F_{ST}$ and $D_{xy}$ for lncRNAs between the two groups (Fig. 2, D and E), and (iv) a significantly elevated proportion of fixed SNPs in lncRNA regions of the domesticate. Complementing these molecular tests of selection, our transgenic experiments and population analysis in three focal lncRNAs indicate phenotype change and significant selection signatures at these loci (Figs. 3 and 4). Together, these findings strongly suggest that rice domestication was characterized by selection for down-regulation of lncRNAs and that the resulting transcriptomic changes played an important role in the emergence of domestication traits in developing rice grains.

While the present study is the first to establish a causal relationship between lncRNA expression variation and crop domestication traits (Figs. 3 and 4), previous work has strongly suggested a potential role for lncRNAs in the domestication process. In cotton, co-expression network construction helped to identify several functional lncRNA candidates involved in fiber initiation and elongation (5, 8). Analyses in both rice and maize have identified correlations between lncRNA loci and SNPs associated with domestication-related traits (5). In addition, genome scans for domestication-related selection in rice have implicated genomic regions enriched for lncRNAs; notably, these putative sweep regions include six of the nine lncRNA clusters identified in the present study (Fig. 2B). Results of the present study can also help to explain the common phenomenon in quantitative trait locus (QTL) mapping and genome-wide association studies analysis of detecting significant QTLs in noncoding genomic regions. Together, these findings suggest that, in cereal crops and other crop species, lncRNA expression variation is likely to play an important role in the regulatory changes associated with crop domestication.

Fig. 4. Natural variation and diversity analysis for LncRNA_2308, LncRNA_1267, and LncRNA_1631. (A to C) Unrooted neighbor-joining phylogenetic trees of the three lncRNAs. Accessions are color-coded by variety group or species as indicated in the legend and showed in table S1. (D to F) Relative expressions of three lncRNAs in the different clusters. The clusters are showed in (A) to (C). (G to I) Grain chalkiness rates of cultivated and wild accessions in the different clusters. The clusters are showed in (A) to (C). Numbers of plants used in each study are indicated in the brackets. (J and K) Nucleotide diversity and neutral test in three lncRNAs and promoter regions (2 kb).
For the particular case of rice, our study suggests that a significant proportion of expression divergence between wild and domesticated *japonica* rice is likely adaptive and that IncRNA-mediated regulatory changes have played an important role in reshaping the rice transcriptome during the domestication process. The apparent pervasiveness of selection on IncRNAs across the *japonica* rice genome suggests that this class of regulatory factor has been an important target of selection during domestication. Changes at individual IncRNA loci are likely to contribute only slightly to phenotypic change; however, considered in the aggregate, these changes can contribute substantially to phenotypic evolution (Fig. 5). Selection on IncRNA expression can thus be thought of as a way of modulating expression of downstream genes and pathways without markedly altering protein function. The IncRNAs with high expression diversities (Fig. 1E) can be considered as the transregulators fine tuning the expression levels of these genes (Fig. 5). Therefore, the high diversities of IncRNAs provide a rich selectable candidate resource in addition to the protein-coding genes. Together, selection on IncRNAs and on protein-coding genes can provide complementary mechanisms for generating phenotypic diversity and optimal fitness. In light of the present findings, future breeding efforts could potentially target altered IncRNA expression as a means by which to achieve desired agronomic traits.

**Materials and Methods**

Sample and tissue collection

We obtained seeds of 638 *O. sativa* accessions (33, 34) and 220 accessions of *O. rufipogon* representing a wide geographical distribution (17, 35) and grew them in the greenhouse of the Chinese Academy of Agricultural Sciences (CAAS) in Beijing, China (table S1). In a preliminary morphological and molecular marker survey, we were able to evaluate morphological and genetic variations of *japonica* and *O. rufipogon* and their relationship. Abundant research based on different genetic markers—including SSRs, transposons, neutral genes, or whole-genome SNPs—indicates that the *O. rufipogon* populations in China were the recent ancestor of domesticated *O. sativa* ssp. *japonica* (15). Because of their close and unambiguous phylogenetic relationship and lower levels of influence of contemporary crop-to-wild gene flow compared to *O. sativa* ssp. *indica*, the species pair *O. rufipogon* and *O. sativa* ssp. *japonica* were a superior research system for comparative genome analyses (17, 36). Selection of specific accessions for inclusion in analyses was based on three specific criteria: (i) typical morphology of wild rice (fig. S1). Wild rice typically displays the following traits: long awns, severe seed shattering, proanthocyanidin-pigmented pericarps (giving the seed a reddish-brown color), consistently small seeds, open panicles with few secondary branches bearing relatively few grains, and a nonerect to prostrate/creeping growth habit. (ii) Representativeness of genetic diversity based on phylogenetic tree (Fig. 1, A and B) and (iii) representativeness of geographical distribution. Genetic distance was calculated using the log shared allele distance, and phylogenetic reconstruction was based on the neighbor-joining method implemented in PowerMarker. On this basis, we chose 19 accessions, including 10 *japonica* and 9 *O. rufipogon* from China, which were phylogenetically closest to *japonica* (Fig. 1B), and planted them in growth chambers at CAAS (12-hour day length; table S1). We collected the early panicles when the inflorescence became approximately 40 mm long. During this developmental stage, the reproductive organs, including the stamen and pistil, have become differentiated pollen grains and embryo sacs (19). Komiy et al. (36) found that a number of IncRNAs might play an important role in the development of premeiotic germ cells.

**Isolation, quantification, and quality determination for RNA**

We collected three early developing panicles from different plants of each accession to extract the RNA and then balanced the concentrations to equalize the RNA contributions of each extraction. In the experiment, the total RNA was extracted from all samples based on the standard protocol of the TRIzol RNA Extraction Kit (Invitrogen, USA). RNA degradation and contamination monitoring were performed using aliquots run on 1% agarose gels. RNA purity was assessed using the NanoPhotometer spectrophotometer (IMPLEN, USA).

---

**Fig. 5. A model for the evolution of IncRNAs during rice domestication.** Selection reduces genetic diversity at IncRNAs, as illustrated by the loss of sequence variants in the domesticated species. Multiple lncRNAs interact to determine the expression level of a given target gene. During domestication, some IncRNAs are down-regulated. The resulting decrease in mRNA levels (gray lines) contributes to domestication associated phenotypes.
RNA concentration was measured by the Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, USA). In addition, for RNA integrity assessment, we used the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, USA).

Library preparation for IncRNA sequencing
The total RNA of each sample was 9 μg, which was the raw material for RNA sample preparation. First, ribosomal RNA was removed by the Epicentre Ribo-zero RNA Removal Kit (Epicentre, USA), and ribosomal RNA (rRNA) free residue was cleaned up by ethanol precipitation. The generation of the sequencing library was performed using the rRNA-depleted RNA by the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, USA) and following the manufacturer’s recommendations. Both poly(A\(^+)\) and poly(A\(^-\)) IncRNAs were identified on the basis of the NEBNext Ultra Directional RNA Library Prep Kit. As a last step, the products were purified, and the quality of the sequencing libraries was assessed by the Agilent Bioanalyzer 2100 system (Agilent Technologies, USA).

Clustering and sequencing
Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, USA) according to the manufacturer’s instructions. After cluster generation, the libraries were then sequenced on an Illumina HiSeq 2500 platform, and 125-bp paired-end reads were generated. The sequencing data were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive, and the numbers are SRR9263865 to SRR9263883.

Data filtering
Clean data were obtained by removing the reads containing adapter, poly-N, and low quality reads from raw data. The volume of clean data for each sample was shown in fig. S2C. We calculated Q20, Q30, and GC content of the clean data. The *japonica* genome assembly (RGAP 7.0) was used as the reference genome throughout this study. The index of the reference genome was built using Bowtie v2.2.6, and paired-end clean reads were aligned to the reference genome using TopHat v2.1.0. The resulting alignment files from Bowtie v2.2.6 and TopHat v2.1.0 were merged. Then, SAMTools v1.2.5 and Picard v1.39 (https://github.com/broadinstitute/picard) were used to filter reads that mapped to multiple positions. Statistics for these results can be found in fig S2.

Identification pipeline and characters of IncRNA
We used two kinds of software to confirm the splicing to reduce the rate of false positives (20). The mapped reads were assembled by both Scripture Beta2 and Cufflinks v2.1.1 with a reference-based approach. Although both methods use spliced reads to determine exon connectivity, Scripture uses a statistical segmentation model to identify the expressed segments, whereas Cufflinks v2.1.1 uses a probabilistic model to simultaneously assemble and quantify the expression level in a given locus. Scripture Beta2 was run with default parameters, and Cufflinks v2.1.1 was run with “-min-frags-per-transfrag = 0” and “--library-type,” while other parameters were set as default. Each individual’s sequencing data was mapped separately, and each genome mapping result yielded that individual’s transcript dataset for subsequent analyses. The union of all individual transcript datasets was the object of the analysis. A total of 559,652 rice transcripts were reconstructed from all of the RNA-seq data using the combined results of Cufflinks v2.1.1 and Scripture Beta2 using Cuffcompare (a program of Cufflinks v2.1.1).

We then filtered the transcripts based on the following three criteria. First, we filtered out the low expression transcripts characterized by multiple exons with an FPKM of ≤0.5 and transcripts characterized by a single exon with an FPKM of ≤2. Then, we removed the transcripts with lengths smaller than 200 nucleotides or those overlapping with known genes. Last, all transcripts were compared with mRNA using class_code of Cuffcompare. Three types of results were considered as candidate IncRNA transcripts (fig. S2G) (i, transcripts with a trans-frag falling entirely within a reference intron; u, unknown, intergenic transcripts; x, the transcripts overlapping with reference exons on the opposite strand). After filtering, 9296 candidate IncRNA transcripts were obtained.

We then evaluated the coding potential of the remaining transcripts to identify previously unidentified IncRNAs. We used two methods to distinguish the coding and noncoding transcripts. First, CPC (coding potential calculator, version 0.9-r2; https://github.com/biocoder/cpc) was used to assess the extent and quality of the open reading frame in a transcript and search the sequences in the known protein sequence database to clarify the coding and noncoding transcripts based on the NCBI eukaryotes’ protein database (e value cutoff: 1 × 10\(^{-30}\)). All transcripts with CPC scores >0 were discarded. On the basis of Kong et al. (37), the false-negative rate was less than 10% for CPC analysis. Second, we translated each transcript in all three possible frames using PfamScan v1.3 to identify the occurrence of any of the known protein family domains documented in the Pfam database (release 27; used both Pfam A and Pfam B). The default parameters used for Pfam searches were -E 0.001 --domE 0.001. Any transcript with a Pfam hit was excluded in following steps. A detailed flow chart of IncRNA identification is shown in fig S2A.

Last, we characterized the basic genomic features of the obtained IncRNAs and compared these features with the available IncRNAs features of rice (8), *Arabidopsis* (20), and human (21) or available mRNA features of rice (fig. S2, I to K). Similar to the findings in *Arabidopsis* (20), only around half of IncRNAs were spliced (42.8%), while more than 98% of human IncRNAs were spliced (21). Rice IncRNAs had fewer exons than mRNAs (1.86 for lncRNA and 4.48 for mRNA, on average, respectively), but their exon lengths (median length of 955 nucleotides; fig S3) were shorter than those of mRNA (median length of 1580 nucleotides; fig S2). Full-length rice IncRNA transcripts were longer than the IncRNA transcripts of *Arabidopsis* (median length of 285 nucleotides) (33) and human IncRNA transcripts (median length of 592 nucleotides) (21). Rice IncRNAs were much more A/U-rich than the coding sequences and the 5’ untranslated regions (5’UTRs) of protein-coding genes but were less A/U-rich than 3’UTRs (fig S2K). This feature was also observed in the IncRNAs of *Arabidopsis*, rice, and animals in previous studies (20, 38, 39), implying that this feature might be related to the functions of IncRNAs. At the same time, we compare the expression profile of core developmental genes between *japonica* and *O. rufipogon*. On the basis of rice plant developmental (19), we selected three genes that function in floral organ differentiation and would be expected to be expressed in early developing panicles: *SPW1* (LOC_Os06g49840), *OsMADS2* (LOC_Os01g66030), and *OsMADS45* (LOC_Os08g41950). There were no significant different expression pattern for *SPW1* (P = 0.23; fig. S3B), *OsMADS2* (P = 0.31; fig. S3C), and *OsMADS45* (P = 0.82; fig. S3D) between *japonica* and...
Identification of differentially expressed IncRNAs loci

FPKM was calculated on the basis of the lengths of the fragments and the number of reads mapped to this fragment. IncRNA FPKMs were computed by summing the FPKMs of exon in each IncRNA. Cufflinks v2.1.1 was used to calculate FPKMs of lncRNAs in each sample. In our analysis, we conducted Pearson correlation coefficients and a principal variance component analysis of expression variation to evaluate the expression difference between japonica and O. rufipogon populations using R scripts. Hence, we considered all samples in japonica or O. rufipogon as biological replicates. Cuffdiff v2.1.1 provides statistical routines for determining differential expression in the transcripts of IncRNA expression data using a model based on the negative binomial distribution. To control the FDR, an adjusted P value was calculated for each IncRNA locus, and the IncRNAs that remained statistically significant with FDR (adjusted P < 0.05) were assigned as differentially expressed IncRNAs. IncRNAs with higher read counts were more reliable, and for further analysis, we kept the subset of IncRNAs with at least 20 mapped reads in the 10 japonica and 9 wild rice accessions. We also selected the transcripts with highest expression for further analysis.

Testing for directional selection of IncRNAs

A rank-based method was further used to identify the genes that were likely under directional selection following the protocol of Guo et al. (38) and Blekhman et al. (39). This method identifies genes under directional selection as those showing low within-species expression variation and high between-species differences in expression; genes showing the opposite expression were inferred to be evolving neutrally (40). Initially, the variance between individuals was estimated by analysis of variance (ANOVA) for each gene (R 3.4.3), and a distribution map of the ranked between-individual variances was generated (fig. S3, F and G). From these analyses, we concluded that the bottom 20% genes based on the distribution map and the significantly differentially expressed genes (FDR < 0.05) were IncRNAs that have evolved under directional selection.

Adding to expression level measurements, we further compared patterns of nucleotide diversity between candidate selected and nonselected IncRNAs based on genome resequencing data. We used our published resequencing data for 177 japonica and 65 O. rufipogon in China to calculate the polymorphism and divergence across all candidate IncRNAs to test for molecular signatures of directional selection. A set of 242 individuals was sequenced at >20× coverage. Before we carried out further analyses, we examined the quality of our sequencing data with the FastQC software. The reads derived from PCR/optical duplicates were removed using SAMtools rmdup. Second, mpileup files were generated by SAMtools mpileup with the parameters “-q 1 -C 50 -S -D -m 2 -f 0.002.” Third, SNPs detection was performed using the GATK HaploypeCaller, which calls SNPs via local de novo assembly of haplotypes in an active region. We set a high SNP confidence score with parameters -stand_call_conf 30 -stand_emit_conf 40, and ~42 M SNPs were obtained. Then, we removed called variants if any of the following filter criteria were detected: (i) >2 alleles (retaining biallelic sites only), (ii) <4× coverage in any sample, (iii) Phred-scaled probability score of less than 25.0, (iv) the loci exhibiting the highest levels of heterozygosity according to Hardy-Weinberg equilibrium, (v) loci with a maximum missing rate more than 0.2, and (vi) the SNPs with expanded form is MAF (minor allele frequency) < 0.01. A total of ~2.7 M SNPs were considered valid for the study. Ratios of nucleotide diversity between O. rufipogon (πw) and japonica rice (πi), FST and SNP frequencies between the two groups, were calculated for each IncRNA region using Vcftools v0.1.13, and Dxy was calculated on the basis of the PopGenome of R Package. The upstream gene regulatory region (the 2000-bp region immediately flanking the transcription start site) and IncRNA region were selected to calculate the nucleotide diversity (πi), FST, and Dxy values for each IncRNA in the analyses. We used the Wilcoxon signed rank method to determine whether the diversity and differentiation levels for candidate selected and nonselected IncRNAs were significantly different from each other. Three measures (πw, πi, FST, and Dxy) were assessed individually and compared between candidate selected and nonselected IncRNAs. If the three values for candidate selected IncRNAs were significantly larger than the values of nonselected IncRNAs, we considered the candidate selected IncRNAs to be showing a signature of selection. Statistical significance was also assessed using a permutation analysis. A set of 208 IncRNAs was randomly selected from the pool of 3155 nondirectionally selected IncRNAs each time, with the resampling procedure repeated 1000 times. The interval of permutation analysis results was between the red lines in graphs in Fig. 2 (C to E). We can see the values of πw/πi, FST, and Dxy for candidate selected IncRNAs were above the significance threshold (the red line). These analyses demonstrated that there was a significant sequence difference between the directionally selected IncRNAs and the other IncRNAs.

Coalescent simulations for derived alleles based on two models

Because there were two major factors, selection and demographic bottlenecks, which could affect genetic variation in the japonica genome compared to its wild progenitor, we conducted coalescent simulations using SLim3 software to distinguish between the two possibilities. The bottleneck model (s = 0) assumes that a single ancestral population of size Na at time t generation (10,000 years) ago experienced an instantaneous size shift to a bottlenecked population of size Nb based on previous estimates. The duration of the bottleneck was d, and the bottleneck population expanded instantaneously to the present population of size Np, (125,000) at time t - d generation ago. The severity of the bottleneck (K) was the ratio of Na and d. We simulated the average number of SNP patterns for 208 IncRNAs (average length is 500 bp). We obtained Na = 156,000 based on \( \theta_{\text{wild}} = 4 \times N_a \times \mu \times r \). The parameters of bottleneck plus selection model for each simulation were drawn from previous probability distributions, with \( s (0.01, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20) \). For each locus, we started with π and calculated \( \pi_{\text{w}}/\pi_i \). The simulation was accepted if the value of \( \pi_{\text{w}}/\pi_i \) fell within 20% of the observed data for wild rice. The AIC was calculated by multiplying likelihoods per locus, and we then assessed the fit of bottleneck or bottleneck plus selection during japonica domestication in 3363 putatively selected IncRNAs. AIC was calculated by two times of the number of parameters minus two times of the ln (likelihood value). The code is below:

```python
initialize() {
    initializeMutationRate(1e-8);
    initializeMutationType("m1", 0.5, "f", 0.0); // non-coding
    initializeMutationType("m2", 0.1, "g", -0.03, 0.2); // deleterious
    initializeMutationType("m3", 0.8, "e", 0.1); // beneficial
    initialize() {
    initializeMutationRate(1e-8);
    initializeMutationType("m1", 0.5, "f", 0.0); // non-coding
    initializeMutationType("m2", 0.1, "g", -0.03, 0.2); // deleterious
    initializeMutationType("m3", 0.8, "e", 0.1); // beneficial
```
Total RNA of all samples was extracted using TRIzol (Invitrogen, USA), according to the manufacturer’s standard protocol. The first-strand cDNA was synthesized from isolated total RNA by RT using the ImProm-II Reverse Transcription Kit (Promega, USA) according to the manufacturer’s protocol with random primers. qRT-PCR was performed using the SYBR Premix Ex TaqTM Kit (Takara, Japan) in an ABI PRISM 7900HT (Applied Biosystems, USA), according to the manufacturer’s instructions. PCR was independently carried out three times for each RNA sample, and we do two independent biological replicates (two independent RNA extractions) for each accession. The rice ubiquitin 1 (LOC_0103g13170) was used as the internal control. The primers for qRT-PCR were listed in table S2.

Vector construction and plant transformation
For preparation of the overexpression construct, we selected the top 10 lncRNAs (Table S5) with the most significant difference in expression level between O. rufipogon and japonica rice. We designed four pairs of oligonucleotide primers, which cover the genomic region ~1 kb in length surrounding each lncRNA to identify the nearly full-length cDNA sequences of these lncRNAs (Table S3). The allele of each lncRNA was selected from wild rice accessions with the highest expression level of this lncRNA and was fused with an ubiquitin promoter to make the transformation constructs. The PCR product of these lncRNAs was amplified from the panicles of O. rufipogon accessions with the highest expression level and then cloned into PstI site of the binary vector pCAMBIA1390, which contained the maize ubiquitin gene promoter using the In-Fusion Advantage PCR Cloning Kits (Clontech, USA). Constructs were individually transformed into the Agrobacterium tumefaciens strain EHA105 and then introduced into Nipponbare via agrobacteria-mediated transformation. Primers used for construct preparation were listed in table S2. At least 40 transgenic events were produced for each construct. The chalkiness and percentage of chalky rice were determined according to GB/T 17891-1999.

RNA-seq on the transformed lines
Total RNA was prepared from early panicles of transgene-positive plants and transgene-negative plants for three focal lncRNAs using the RNAeasy Plant Mini Kit (Qiagen, MD, USA). The integrity of RNA was checked using the Bioanalyzer 2100 algorithm (Agilent Technologies, Tokyo, Japan). Six cDNA libraries were built using the TruSeq RNA Sample Preparation Kit (Illumina, Tokyo, Japan). Paired end reads (100 bp) were got for transcriptome sequencing on an Illumina HiSeq 2000 platform (Takara, Tokyo, Japan). Raw sequences obtained from the Illumina platform were analyzed using publicly available tools and trimmed low-quality bases and the adapters from both ends of the sequences using a customized program. The sequences were mapped to the RGAP 7.0 reference genome sequence using Bowtie v2.2.6 and TopHat v2.1.0. Reference-based assembly of the reads was performed using Cufflinks v2.1.1 and Cuffmerge. The expression level of each transcript was expressed as FPKM value. Cuffdiff was used to detect differentially expressed genes using at least two replicates, with a correlation coefficient of more than 0.90 in each library.

Population genetic analysis in cultivated and wild rice for LncRNA_2308, LncRNA_1267, and LncRNA_1631
To determine whether LncRNA_2308, LncRNA_1267, and LncRNA_1631 were targets of selection, we amplified and sequenced the 726-bp LncRNA_2308, 286-bp LncRNA_1267, 309-bp LncRNA_1631 region,
and upstream gene regulatory region (the 2000-bp region immediately flanking the transcription start site) for a core collection of 239 cultivated and wild rice accessions, including 90 O. sativa ssp. indica, 74 O. sativa ssp. japonica, and 75 O. rufipogon (Table S1). Sequences were aligned by ClustalX. The genealogical trees of three IncRNA regions were constructed by MEGA7 based on neighbor-joining methods. For each taxon of three IncRNA regions, the expected heterozygosity per nucleotide site and Tajima’s D were calculated by DnaSP 6. Allele-frequency spectrum analyses were performed by popstrata (Table S2). Primers used in this study.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/12/eaax3619/DC1

**Fig. S1.** The morphology of wild rice in this study.

**Fig. S2.** The computational pipeline and characteristics of IncRNAs in genome.

**Fig. S3.** The IncRNA expression patterns.

**Fig. S4.** The plot of AIC for each coalescent simulation model.

**Fig. S5.** The grain phenotype of O. sativa ssp. japonica and O. rufipogon.

**Table S1.** Information on the 858 cultivated and wild rice accessions used in this study.

**Table S2.** Primers used in this study.

**Table S3.** Summary information on IncRNAs identified in this study.

**Table S4.** Grain chalkiness rate, 1000 grain weight, and the IncRNA expression level in transgenic plants of T1 and T2 plants harboring a construct of T2308 (IncRNA_2308), T2167 (IncRNA_1267), and T1631 (IncRNA_1631) with the maize ubiquitin gene promoter.

**Table S5.** Transcriptome comparison between the transgene-negative and transgene-positive plants.

*View request a protocol for this paper from Bio-protocol.*

**REFERENCES AND NOTES**

1. J. F. Doebley, B. S. Gaut, B. D. Smith, The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).

2. K. M. Olsen, J. F. Wendel, A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* **64**, 67–70 (2013).

3. M. B. Hufford, X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, R. A. Cartwright, R. J. Elshire, J. C. Glabizzi, K. G. Guille, S. M. Kaeppler, J. Lai, P. L. Moore, M. L. Shannon, C. Song, N. M. Springer, R. A. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M. D. McMullen, D. Ware, E. S. Buckler, S. Yang, J. Ross-Ibarra, Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **44**, 808–811 (2012).

4. R. Kumar, A. Mustafiz, K. K. Sahoo, V. Sharma, S. Samanta, S. K. Sopory, A. Pareek, S. L. Singlea-Pareek, Functional screening of cDNA library from a salt tolerant rice genotype Pokkali identifies mannose-1-phosphate guanylyl transferase gene (OsMPT1) as a key member of salinity stress response. *Plant J.* **81**, 1245–1259 (2021).

5. H. Wang, Q.-W. Niu, H.-W. Wu, J. Liu, J. Ye, N. Y. Xu, N.-H. Chua, Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved IncRNAs associated with agriculture traits.* Plant J.* **84**, 404–416 (2015).

6. Z. Zhou, Y. Shen, M. R. Khan, A. Li, LncReg: A reference resource for IncRNA-associated regulatory networks. *Database* **2015**, 2015–2015.

7. Y. Long, X. Wang, D. T. Youmans, T. R. Chich, How do IncRNAs regulate translation? *Sci. Adv.* **3**, eaa2110 (2017).

8. Y.-C. Zhang, J.-Y. Liao, Z.-Y. Li, J. Yu, P.-Z. Zhang, Q.-F. Li, I.-H. Qu, W.-S. Shu, Y.-Q. Chen, Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* **15**, 512 (2014).

9. Y. Wang, X. Luo, F. Sun, J. Hu, X. Zha, W. Su, J. Yang, Overexpressing IncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* **9**, 3518 (2018).

10. A. Nceucela, M. Soroush, M. Warnfors, A. Liechti, T. Daish, U. Zeller, J. C. Baker, F. Grützner, H. Kaesemann, The evolution of IncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
32. Z. Lai, N. C. Kane, Y. Zou, L. H. Rieseberg, Natural variation in gene expression between wild and weedy populations of Helianthus annuus. Genetics 179, 1881–1890 (2008).
33. C.-H. Wang, X.-M. Zheng, Q. Xu, X.-P. Yuan, L. Huang, H.-F. Zhou, X.-H. Wei, S. Ge, Genetic diversity and classification of Oryza sativa with emphasis on Chinese rice germplasm. Heredity 112, 489–496 (2014).
34. X.-M. Zheng, T. Gong, H.-L. Ou, D. Xue, W. Qiao, J. Wang, S. Liu, Q. Yang, K. M. Olsen, Genome-wide association study of rice grain width variation. Genome 61, 233–240 (2017).
35. X. Wei, W.-H. Qiao, Y.-T. Chen, R.-S. Wang, L.-R. Cao, W.-X. Zhang, N.-N. Yuan, Z.-C. Li, H.-L. Zeng, Q.-W. Yang, Domestication and geographic origin of Oryza sativa in China: Insights from multilocus analysis of nucleotide variation of O. sativa and O. rufipogon. Mol. Ecol. 21, 5073–5087 (2012).
36. R. Komiya, H. Ohyanagi, M. Niihama, T. Watanabe, M. Nakano, N. Kurata, K. Nonomura, Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. Plant J. 78, 385–397 (2014).
37. L. Kong, Y. Zhang, Z.-Q. Ye, X.-Q. Liu, S.-Q. Zhao, L. Wei, G. Gao, CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35, W345–W349 (2007).
38. J. Guo, R. Liu, L. Huang, X.-M. Zheng, P.-L. Liu, Y.-S. Du, Z. Cai, L. Zhou, X.-H. Wei, F.-M. Zhang, S. Ge, Widespread and adaptive alterations in genome-wide gene expression associated with ecological divergence of two Oryza species. Mol. Biol. Evol. 33, 62–78 (2016).
39. R. Blekhman, A. Oshlack, A. E. Chabot, G. K. Smyth, Y. Gilad, Gene regulation in primates evolves under tissue-specific selection pressures. PLOS Genet. 4, e1000271 (2008).
40. I. G. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene regulation. Nat. Rev. Genet. 13, 505–516 (2012).

Acknowledgments: We thank J. Wan, S. Huang, C. Zou, and Y. Wu at Chinese Academy of Agriculture Sciences for critical reading and comments on the manuscript. We acknowledge TopEdit LLC for the linguistic editing and proofreading during the preparation of this manuscript. Funding: This work was supported by the National Key Research and Development Program of China 2016YFD0100301 (to X.M.Z.), the National Natural Science Foundation of China 31670211 (to X.M.Z.), and the Agricultural Science and Technology Innovation Program of Chinese Academy of Agriculture Sciences. Author contributions: X.M.Z. and Q.W.Y. designed research. X.M.Z., J.C., H.B.P., S.L., Q.G., J.R.W., W.H.Q., and H.W. performed experiments, analyzed data, and wrote the manuscript draft. J.L., K.M.O., and Q.W.Y. conceived and supervised the project and wrote the paper. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 19 March 2019
Accepted 30 October 2019
Published 18 December 2019
10.1126/sciadv.aax3619

Citation: X. M. Zheng, J. Chen, H. B. Pang, S. Liu, Q. Gao, J. R. Wang, W. H. Qiao, H. Wang, J. Liu, K. M. Olsen, Q. W. Yang, Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. Sci. Adv. 5, eaax3619 (2019).
Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication
X. M. Zheng, J. Chen, H. B. Pang, S. Liu, Q. Gao, J. R. Wang, W. H. Qiao, H. Wang, J. Liu, K. M. Olsen and Q. W. Yang

Sci Adv 5 (12), eaax3619.
DOI: 10.1126/sciadv.aax3619