

Identification of long noncoding natural antisense transcripts (lncNATs) correlated with drought stress response in wild rice (*Oryza nivara*)

Yong-Chao Xu

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

Jie Zhang

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

Dong-Yan Zhang

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

Ying-Hui Nan

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

Song Ge

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

Ya-Long Guo (✉ yalong.guo@ibcas.ac.cn)

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences

Research Article

Keywords: drought stress, long noncoding RNA, *O. nivara*, strand-specific RNA-seq, wild rice

Posted Date: March 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-280217/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Wild rice, including *Oryza nivara* and *Oryza rufipogon*, which are considered as the ancestors of Asian cultivated rice (*Oryza sativa* L.), possess high genetic diversity and serve as a crucial resource for breeding novel cultivars of cultivated rice. Although many rice domestication related traits, such as seed shattering and plant architecture, have been intensively studied at the phenotypic and genomic levels, further investigation is needed to understand the molecular basis of phenotypic differences between cultivated and wild rice. Drought stress is one of the most severe abiotic stresses affecting rice growth and production. Adaptation to drought stress involves a cascade of genes and regulatory factors that form complex networks. Long noncoding natural antisense transcripts (lncNATs), a class of long noncoding RNAs (lncRNAs), regulate the corresponding sense transcripts and play an important role in plant growth and development. However, the contribution of lncNATs to drought stress response in wild rice remains largely unknown.

Results

Here, we conducted strand-specific RNA sequencing (ssRNA-seq) analysis of Nipponbare (*O. sativa* ssp. *japonica*) and two *O. nivara* accessions (BJ89 and BJ278) to determine the role of lncNATs in drought stress response in wild rice. A total of 1,246 lncRNAs were identified, including 1,091 coding–noncoding NAT pairs, of which 50 were expressed only in Nipponbare, and 77 were expressed only in BJ89 and/or BJ278. Of the 1,091 coding–noncoding NAT pairs, 240 were differentially expressed between control and drought stress conditions. Among these 240 NAT pairs, 12 were detected only in Nipponbare, and 187 were detected uniquely in *O. nivara*. Furthermore, 10 of the 240 coding–noncoding NAT pairs were correlated with genes previously demonstrated to be involved in stress response; among these, nine pairs were uniquely found in *O. nivara*, and one pair was shared between *O. nivara* and Nipponbare.

Conclusion

We identified lncNATs associated with drought stress response in cultivated rice and *O. nivara*. These results will improve our understanding of the function of lncNATs in drought tolerance and accelerate rice breeding.

Background

Rice is one of the most important crops in the world and a major source of food for billions of people. Asian cultivated rice (*Oryza sativa* L.) was domesticated from wild rice species including *Oryza rufipogon* and *Oryza nivara*, and experienced a bottleneck effect that severely reduced its genetic diversity [1] and decreased its viability in the natural environment [2]. In contrast to cultivated rice, wild rice species

possess higher genetic diversity, which contributes to its greater resistance to biotic and abiotic stresses, and this characteristic of wild rice is crucial for understanding and improving the stress tolerance of cultivated rice.

Drought stress is one of the most severe abiotic stresses affecting crop yield. In the natural environment, plants adapt to drought stress by employing various strategies, such as speeding up their life cycle to avoid drought stress, reducing water loss, and improving water use efficacy [3]. Drought tolerance is a complex trait involving the regulation of a number of physiological and biochemical processes, including stomatal density [4], leaf rolling [5], osmotic adjustment [6], and root system development [7], at different development stages. Plants cope with drought stress by regulating phytohormones (such as abscisic acid [ABA], jasmonic acid [JA], and auxin) or making osmotic adjustments [8]. These mechanisms of drought stress response involve genes belonging to various families including *WRKY*, *MYB*, *NAC*, *ABRE*, *PP2C*, and *SnRK2* [9–13].

Long noncoding RNAs (lncRNAs) constitute a large fraction of the transcriptome that does not encode proteins [14], and play important roles in various biological processes, such as genome stability [15], vernalization [16], telomere maintenance [17], transcriptional activation [18], and other developmental processes [19]. Various transcriptome studies have revealed that pervasive transcription from noncoding transcripts can give rise to functional lncRNAs [20]. lncRNAs are classified as long intergenic noncoding RNAs (lincRNAs) and long noncoding natural antisense transcripts (lncNATs) [21]. lncNATs are transcribed from the opposite strand of sense RNA in the same genomic regions, and may regulate the expression of sense RNA [22–24]. For example, lncNATs affect genes on the opposite strand and increase starch content and grain weight in rice [25]. Furthermore, lncNATs have been studied at the genomic level in many species, including human (*Homo sapiens*) [26], mouse (*Mus musculus*) [27, 28], rice [29–32], *Arabidopsis thaliana* [21, 33], maize (*Zea mays*) [34], *Plasmodium falciparum* [35], and yeast (*Saccharomyces cerevisiae*) [36]. In particular, lncNATs are correlated with the response to various abiotic and biotic stresses [29, 34, 37]. For example, a recent study in *Arabidopsis* revealed that NATs affect plant thermotolerance [38].

Although our understanding of drought stress response in plants has improved substantially, this topic needs to be investigated further, given the ongoing global climate change. Firstly, more genetic variability that contributes to drought tolerance needs to be identified in the plant germplasm. Secondly, since most of the previous studies focused on cultivated rice, the drought stress response in wild rice remains largely unknown and needs special attention.

Here, we performed strand-specific RNA sequencing (ssRNA-seq) analysis of Nipponbare (*O. sativa* ssp. *japonica*; hereafter referred to as Nip) and two *O. nivara* accessions (BJ89 and BJ278). We identified 1,091 coding–noncoding NAT pairs, of which 240 pairs were differentially expressed between control and drought stress conditions, and 187 pairs were specifically found in *O. nivara*. Furthermore, according to the GO enrichment analysis of sense transcripts, 10 coding–noncoding NAT pairs were correlated with drought stress, of which nine pairs were uniquely identified in BJ89 and BJ278. Thus, we identified

numerous IncNATs correlated with drought stress in *O. nivara*. These IncNATs potentially play important roles in the response to drought stress, and will provide new insights into the mechanism of drought tolerance in *O. nivara*, thus facilitating breeding the cultivated rice varieties.

Results

ssRNA-seq and transcript assembly

To examine IncNAT expression patterns in wild rice under drought stress treatment, one cultivated rice accession (Nip) and two *O. nivara* accessions (BJ89 and BJ278) were grown under control and drought conditions (Fig. 1a–c). Leaves of BJ89 and BJ278 showed lower water loss than those of Nip (Fig. 1d). Furthermore, the two *O. nivara* accessions exhibited a higher survival rate than Nip after 25 days of drought stress treatment (Fig. 1e). These physiological data suggest that *O. nivara* accessions BJ89 and BJ278 are more drought tolerant than Nip at the seedling stage.

Next, we conducted ssRNA-seq analysis of these three accessions treated with or without 25% polyethylene glycol (PEG-4000; w/v) for 10 days. A total of 18 strand-specific cDNA libraries were constructed from leaf tissues, with three replicates per accession in both control and drought stress treatments. In total, 445.5 million paired-end reads (2×125 bp) were generated by ssRNA-seq using Illumina HiSeq 2500, of which 373.6 million reads (83.8%) mapped perfectly on to the Nip reference genome (Table S1). Pearson correlation coefficients of the three biological replicates of each accession were greater than 0.9, indicating the high reproducibility of our ssRNA-seq data (Figure S1). Among the 373.6 million paired-end reads, we identified 62,201 transcripts, including 17,583 novel transcripts (8,905 known gene loci and 2,704 new gene loci) with hisat2 [39] and stringtie [40] (Fig. 2a). We also conducted principal component analysis (PCA) of gene expression data. The results showed that PC2 clearly distinguished between the control and drought treated samples, while PC3 separated the different accessions (Figure S2).

Identification of lncRNAs and NAT pairs

To identify lncRNAs, novel transcripts larger than 200 nt were mapped against the Rfam 13.0 database to exclude micro RNAs (miRNAs), ribosomal RNAs (rRNAs), and other small noncoding RNAs [41]. Then, any transcripts with a coding potential, according to Coding Potential Calculator (CPC) [42] and Pfam with HAMMER scan [43], were filtered out (Fig. 2a). Finally, a total of 1,246 lncRNAs were identified, including 940 in Nip, 959 in BJ89, and 974 in BJ278 (reads in the same accession under both control and drought stress conditions were combined to identify lncRNAs) (Fig. 2b and Table S2). Among the 1,246 lncRNAs, 692 (55.5%) were common to all three accessions, 306 (24.6%) were uniquely found in at least one of the two *O. nivara* accessions, and 111 (8.9%) were present in both *O. nivara* accessions (Fig. 2b). The expression profiles of lncRNAs were more different between the three accessions than between the drought and control conditions of the same accession (Fig. 2c).

Of the 1,246 lncRNAs, a total of 394 lncRNAs were differentially expressed between control and drought stress treatments (118 in Nip, 227 in BJ89, and 174 in BJ278). Among these 394 lncRNAs, 23 were common to all three accessions; 34 were identified as being shared between one of the *O. nivara* accessions and Nip; 45 were shared between the two *O. nivara* accessions (BJ89 and BJ278); and 139, 92, and 61 were uniquely found in BJ89, BJ278, and Nip, respectively (Fig. 3a, b).

Based on their location relative to the gene coding regions, 675 of the 1,246 lncRNAs were long intergenic noncoding RNAs (lncRNAs; lncRNAs located in intergenic regions), and 571 lncRNAs were long noncoding natural antisense transcripts (lncNATs; lncRNAs overlapped with coding genes on the opposite DNA strand) (Table S2). The lncRNAs contained fewer exons than lncNATs; 72.7% lncRNAs contained only one exon (Figure S3). By contrast, mRNAs contained more exons than both lncRNAs and lncNATs (Figure S3). In addition, mRNAs showed higher expression variation than lncRNAs and lncNATs under either control or drought stress condition at the genome level (Figure S4).

It has been shown that lncNATs can regulate the expression of sense transcripts [44, 45], and each strand of a NAT pair could potentially represent a protein-coding gene. Therefore, in addition to identifying NAT pairs from lncRNAs, we scanned NAT pairs across the whole genome of the three accessions, based on the annotation of the Nip reference genome. All transcripts annotated in the Nip reference genome were integrated and assembled with the ssRNA-seq data generated in this study. A total of 8,529 NAT pairs with overlapping regions greater than 25 nt were identified according to a previous study [31]. According to the coding capacity of the sense–antisense pair [27], 86.88% (7,410) of NAT pairs were coding–coding pairs (both transcripts with protein-coding capacity), 0.33% (28) were noncoding–noncoding pairs (both transcripts represented lncRNAs), and 12.79% (1,091) were coding–noncoding pairs (one strand showed protein-coding capacity, while the other strand represented an lncRNA) (Table S3), and each transcript (transcripts with protein-coding capacity or lncRNAs) could flank a few different transcripts. Depending on the direction and location of the sense and antisense transcripts, 61.33% of the 8,529 NAT pairs were enclosed (one transcript fully embedded in the other), 23.38% were divergent (head-to-head, 5'-end overlap), and 15.29% were convergent (tail-to-tail, 3'-end overlap) (Fig. 3d).

Of the 8,529 NAT pairs, 5,866 were detected through ssRNA-seq analysis of the three accessions, of which 4,783, 4,813, and 4,596 were detected in Nip, BJ89, and BJ278, respectively (Fig. 3c, Figure S5, Table S3). Of the 5,866 NAT pairs, 62.2% (3,651) were shared among all three accessions (2,792 coding–coding pairs, 13 noncoding–noncoding pairs, and 846 coding–noncoding pairs) (Figure S5, Table S3); 8.8% (517) were expressed only in Nip (462 coding–coding pairs, 5 noncoding–noncoding pairs, and 50 coding–noncoding pairs) (Figure S5, Table S3); and 18.5% (1,083) were uniquely expressed in *O. nivara* accessions (998 coding–coding pairs, 8 noncoding–noncoding pairs, and 77 coding–noncoding pairs) (Figure S5, Table S3).

Genes involved in the response to drought stress

To determine the differences in gene expression patterns between control and drought treatments, we identified differentially expressed genes (DEGs) using the following criteria: fold change (FC) ≥ 2.0 and

false discovery rate (FDR) ≤ 0.01 . A total of 3,934, 5,880, and 5,036 DEGs were identified between control and drought treatments in Nip, BJ89, and BJ278, respectively (Fig. 3a, Table S4).

To identify genes within biological processes related to drought stress, GO enrichment analysis was performed on all DEGs ($FC \geq 2.0$ and $FDR \leq 0.01$) and highly differentially expressed genes (HDEGs) ($FC \geq 4.0$ and $FDR \leq 0.01$) identified in each accession. A total of 57 GO terms were enriched in the three accessions, and most terms were related to primary metabolic pathways essential for plant growth and development, such as ‘biosynthetic process’, ‘cellular biosynthetic process’, and ‘primary metabolic process’ (Figure S6). In addition, different GO terms were enriched in the three accessions in response to drought; for example, the ‘response to water’ GO term was uniquely enriched in BJ278 (Figure S6).

Based on the HDEGs ($FC \geq 4.0$ and $FDR \leq 0.01$), 63 GO terms in the biological process category were enriched, including 10 terms related to stress, such as ‘response to jasmonic acid stimulus’, ‘oxidation reduction’, and ‘gibberellin metabolic process’ [46, 47] (Fig. 4a, Table S5). Among these ten stress related terms, three (‘response to chemical stimulus’, ‘response to stimulus’, and ‘response to stress’) were detected in both BJ89 and Nip; three terms (‘jasmonic acid mediated signaling pathway’, ‘response to jasmonic acid stimulus’, and ‘response to biotic stimulus’) were uniquely enriched in BJ89; and four terms (‘response to abiotic stimulus’, ‘oxidation reduction’, and ‘response to water and gibberellin metabolic process’) were only enriched in Nip. In BJ278, only one GO term (‘carbohydrate metabolic process’) was enriched. These results suggest that Nip, BJ89, and BJ278 employ different mechanisms to respond to drought stress (Fig. 4a, Figure S6).

A total of 134 HDEGs were enriched in these 10 stress related GO terms. Among these 134 genes, 48 were found only in *O. nivara* (either one or both accessions); 12 were found only in Nip; 35 were shared between Nip and one of the two *O. nivara* accessions; and 39 were found in all three accessions but showed different expression levels ($FC \geq 2.0$ and $FDR \leq 0.01$) either between BJ89 and Nip or between BJ278 and Nip under the drought stress condition (Table S5).

A recent study reported the role of protein kinases in abiotic stress [48]. Therefore, we focused on drought stress-responsive protein kinases in our ssRNA-seq data. We found 51 genes encoding protein kinases (24 belonging to ABA co-receptor clade A type 2C protein phosphatases [PP2Cs]), which were induced in response to drought stress and were differentially expressed between drought stress and control conditions in Nip, BJ89, and BJ278 or between *O. nivara* (BJ89 or BJ278) and cultivated rice (Nip) under the drought condition (Figure S7). Among the 24 PP2C genes, 14 were downregulated either in *O. nivara* accessions under drought stress compared with their control counterparts or in *O. nivara* accessions compared with Nip under the drought stress condition (Figure S8).

NAT pairs responsive to drought stress

Antisense transcription could silence or concordantly regulate the sense transcripts [27]. To detect NAT pairs responsive to drought stress, both sense and antisense transcripts of each NAT pair showing differential expression ($FC \geq 2.0$ and $FDR \leq 0.01$) between control and drought stress conditions were

identified as differentially expressed NAT pairs. A total of 369 differentially expressed NAT pairs were identified, of which 240 were coding–noncoding pairs (193 in BJ89, 96 in BJ278, and 53 in Nip). Additionally, among these 240 differentially expressed NAT pairs, 23 were common in all accessions; 18 were shared between Nip and BJ89 or BJ278; 12 were present only in Nip; and 187 were found only in *O. nivara* accessions (Figure S9, Table S6).

According to the effect of antisense transcripts on sense transcripts, we classified the NAT pairs into two categories, as described previously [21]: concordant (sense and antisense transcripts expressed coordinately) and discordant (sense and antisense transcripts showing opposite expression patterns). A total of 24 discordant NAT pairs were identified (one shared among all accessions; two shared between BJ89 and BJ278; one shared between Nip and BJ89; 8, 9, and 12 uniquely found in Nip, BJ89, and BJ278, respectively) (Figure S10), including one NAT pair discordant in Nip but concordant in BJ89, and one pair discordant in BJ89 and BJ278 but concordant in Nip (Fig. 4b and Table S6). Additionally, 102 concordant NAT pairs were upregulated (30 in Nip, 61 in BJ89, and 54 in BJ278) (Fig. 4b). Among the upregulated concordant NAT pairs, 10 were shared among all accessions, 7 were shared between Nip and BJ278, 13 were found only in Nip, and 72 were found only in *O. nivara* (Figure S11 and Table S6).

A total of 245 concordant NAT pairs were downregulated (45 in Nip, 205 in BJ89, and 82 in BJ278), of which 17 were shared among all three accessions, 14 were shared between Nip and BJ89 or BJ278, 14 were found only in Nip, and 200 were found only in *O. nivara* (Fig. 4b, Figure S12, and Table S6). Among the 10 coding–noncoding NAT pairs related to GO enrichment terms for the drought stress response, nine were uniquely found in *O. nivara*, and one was common to all accessions (Table S7, Figure S13). Furthermore, among the nine coding–noncoding NAT pairs uniquely found in *O. nivara*, six were correlated with response to stress, three were correlated with the jasmonic acid stimulus pathway, and one was correlated with oxidation reduction (Table S7).

Discussion

Antisense transcripts are present in various organisms and play important roles in regulating gene expression. For example, the gene encoding the famous transcriptional repressor FLOWERING LOCUS C (FLC), which delays flowering time in Arabidopsis, is repressed at warm temperatures by *COLD INDUCED LONG ANTISENSE INTRAGENIC RNA (COOLAIR)*, an antisense RNA, via histone demethylation [45, 49, 50]. Antisense lncRNAs can also upregulate gene expression; for example, an lncRNA transcribed from promoter region of the *Pcdha* gene leads to DNA demethylation of the CTCF binding sites and the activation of sense promoters [51]. Natural cis-antisense transcripts also define the function of short interfering RNAs (siRNAs) and affect their biogenesis; for example *P5CDH* and *SRO5* regulate salt tolerance by generating two types of siRNAs in Arabidopsis [37]. In addition, NATs also contribute to heterochromatin formation and DNA methylation, and suppress gene expression in tumorous cells [52]. However, the function of conserved lncRNAs can vary across different species. For example, in human and mouse, the conserved lncRNAs exhibit different subcellular localization patterns [53].

O. nivara, one of the wild progenitor species of cultivated rice, inhabits swampy areas with a seasonally dry climate. *O. nivara* possesses greater genetic diversity than cultivated rice and thus is a valuable resource for breeding rice cultivars with desirable traits, such as stress resistance. In this study, we identified 187 coding–noncoding NAT pairs in the leaves of *O. nivara* accessions, of which 10 were correlated with drought stress. These findings are expected to facilitate further analysis of drought stress tolerance in wild rice. However, to identify more drought resistance genes and NAT pairs, a more comprehensive analysis of different tissues and developmental stages of wild rice should be conducted.

Furthermore, we identified NAT pairs based solely on ssRNA-seq data. These NAT pairs need to be validated through experimentation. Additionally, the functions of a vast majority of NAT pairs remain unknown. Therefore, further studies are needed to understand the molecular mechanism of action of these coding–noncoding NAT pairs and their regulatory network under drought stress conditions.

Conclusions

In this study, we performed ssRNA-seq analysis of cultivated rice Nipponbare and two *O. nivara* accessions, and systematically identified numerous genes and NAT pairs responsive to drought stress in all three accessions. Overall, the results of this study will enhance our understanding of the evolution of NATs and drought tolerance in rice, thus facilitating rice breeding.

Methods

Plant material and growth conditions

Oryza nivara accessions BJ89 and BJ278 were collected from Cambodia and Laos, respectively. All experiments were conducted in an environmentally-controlled growth chamber (Percival Scientific, Inc.). Seeds were incubated at 42°C for at least 5 days to break dormancy. Seeds were then placed on a wet filter paper in a culture dish at 30°C for 3 days. The most uniformly germinated seeds were sown in a bottom-less 96-well plate placed in a dish containing Kimura B culture solution [54], and incubated in the growth chamber at 30°C day/25°C night temperature under a 12-h light/12-h dark photoperiod. The culture solution was renewed every 3 days. To conduct drought stress and control treatments, 12-day-old seedlings were transferred to a Kimura B culture solution containing 25% PEG-4000 (w/v) or no PEG-4000, respectively, and incubated for 10 days. Three biological replications were performed for each accession in each treatment.

Strand-specific cDNA library preparation and ssRNA-seq

Total RNA was extracted from leaves of five plants that had been subjected to 10-day-drought or control treatments using Promega Total RNA Isolation System (Z3100). Then, mRNA was fragmented in the fragmentation buffer (Ambion). The first strand of cDNA was reversed transcribed with SuperScript II (Invitrogen) using random primer (TAKARA). Next, dTTP was replaced by dUTP for the synthesis of the second strand. The repaired double-stranded cDNA was ligated with Illumina TruSeq adaptor and

digested using USER enzyme (NEB). Finally, 350–450 bp fragments were recovered and purified, and 18 cDNA libraries (3 accessions × 2 treatments × 3 replicates) were sequenced on the Illumina HiSeq 2500 platform.

Measurement of water loss and survival rate

A water loss experiment was conducted to investigate the variation in the rate of water loss rate from leaves among Nip, BJ89 and BJ278. Five topmost fully expanded leaves were collected from 25-day-old seedlings of each accession and weighted every 30 min at room temperature. Three replications were performed for each accession. The relative water loss rate was calculated using the following equation:

$$\text{Relative water loss rate} = (\text{FW} - \text{CW}) / \text{FW}$$

where FW and CW represent the fresh and current weight of leaves, respectively.

To examine the differences in survival rate among the three accessions, 12-day-old seedlings of each accession were treated with Kimura B culture solution containing 25% PEG-4000 (w/v) for 4 weeks and then transferred to Kimura B culture solution without PEG for 1 week. The survival rate of each accession was calculated as the ratio of alive plants to all plants.

Transcript assembly and lncRNA identification

Paired-end (2 × 125 bp) strand-specific reads of each accession were combined and mapped to the Nip reference genome sequence (IRGSP-1.0) using the hisat2 (2.1.0) software [39]. Stringtie was used to assemble transcripts for regions with read coverage greater than 5X [55]. Then, transcripts of the three accessions were integrated using Cuffmerge and Cuffcompare utilities in the Cufflinks package [56]. Gene annotations of the Nip reference genome (<http://rapdb.dna.affrc.go.jp>) were integrated, and gene expression levels (estimated as FPKM [Fragments Per Kilobase of transcript per Million mapped reads] values) and DEGs were determined using Cuffdiff in the Cufflinks package [57]. Transcripts longer than 200 nt were mapped against the Rfam 13.0 database to exclude miRNAs, rRNAs, and other small noncoding RNAs [41]. Then, Coding Potential Calculator (CPC) [42] and Pfam [43] were used to filter out potential coding transcripts. Gene function annotations were available from the RAP-DB (<http://rapdb.dna.affrc.go.jp/index.html>). GO enrichment analysis was conducted using agriGO [58].

Statistical analysis

All statistical analyses were performed using R (<http://www.r-project.org/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The ssRNA-seq datasets generated during the current study are available in the China National Center for Bioinformation (<https://bigd.big.ac.cn/databases>) under the accession number: CRA003736.

Ethics approval and consent to participate

The study did not require ethical approval or consent, because no endangered or protected plant species were involved.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (31925004 to Y.-L.G.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27010305 and XDA08020103 to Y.-L.G.), The Innovative Academy of Seed Design, Chinese Academy of Sciences (Y.-L.G.), and China Postdoctoral Science Foundation (2020M680749 to Y.-C.X.).

Authors' contributions

Y.-L.G. conceived the study; Y.-C.X., J.Z., S.G., and Y.-L.G. analyzed the data; J.Z. and Y.-H.N. performed phenotypic measurements; D.-Y.Z. collected the leaf samples and extracted RNA for RNA-seq; S.G. provided the *O. nivara* accessions; Y.-C.X. and Y.-L.G. wrote the manuscript, with contributions from all authors.

Acknowledgements

We thank members of the Guo lab for providing helpful suggestions for this study.

References

1. Doebley JF, Gaut BS, Smith BD: The molecular genetics of crop domestication. *Cell* 2006, 127(7):1309-1321.
2. Meyer RS, Purugganan MD: Evolution of crop species: genetics of domestication and diversification. *Nature reviews Genetics* 2013, 14(12):840-852.
3. Hu H, Xiong L: Genetic engineering and breeding of drought-resistant crops. *Annual review of plant biology* 2014, 65:715-741.

4. Ishimaru K, Shirota K, Higa M, Kawamitsu Y: Identification of quantitative trait loci for adaxial and abaxial stomatal frequencies in *Oryza sativa*. *Plant Physiol Bioch* 2001, 39(2):173-177.
5. Yue B, Xue W, Xiong L, Yu X, Luo L, Cui K, Jin D, Xing Y, Zhang Q: Genetic basis of drought resistance at reproductive stage in rice: separation of drought tolerance from drought avoidance. *Genetics* 2006, 172(2):1213-1228.
6. Peleg Z, Fahima T, Krugman T, Abbo S, Yakir D, Korol AB, Saranga Y: Genomic dissection of drought resistance in durum wheat x wild emmer wheat recombinant inbreed line population. *Plant, cell & environment* 2009, 32(7):758-779.
7. Barbez E, Dunser K, Gaidora A, Lendl T, Busch W: Auxin steers root cell expansion via apoplastic pH regulation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 2017, 114(24):E4884-E4893.
8. Gong ZZ, Xiong LM, Shi HZ, Yang SH, Herrera-Estrella LR, Xu GH, Chao DY, Li JR, Wang PY, Qin F et al: Plant abiotic stress response and nutrient use efficiency. *Sci China Life Sci* 2020, 63(5):635-674.
9. Lin Z, Li Y, Zhang Z, Liu X, Hsu CC, Du Y, Sang T, Zhu C, Wang Y, Satheesh V et al: A RAF-SnRK2 kinase cascade mediates early osmotic stress signaling in higher plants. *Nat Commun* 2020, 11(1):613.
10. Miao J, Li X, Li X, Tan W, You A, Wu S, Tao Y, Chen C, Wang J, Zhang D et al: OsPP2C09, a negative regulatory factor in abscisic acid signalling, plays an essential role in balancing plant growth and drought tolerance in rice. *New Phytol* 2020, 227(5):1417-1433.
11. Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, Li J, Tran LS, Qin F: A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat Commun* 2015, 6:8326.
12. Tao Z, Kou Y, Liu H, Li X, Xiao J, Wang S: OsWRKY45 alleles play different roles in abscisic acid signalling and salt stress tolerance but similar roles in drought and cold tolerance in rice. *J Exp Bot* 2011, 62(14):4863-4874.
13. Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K: AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *The Plant journal : for cell and molecular biology* 2010, 61(4):672-685.
14. Marchese FP, Raimondi I, Huarte M: The multidimensional mechanisms of long noncoding RNA function. *Genome biology* 2017, 18(1):206.
15. Hu WL, Jin L, Xu A, Wang YF, Thorne RF, Zhang XD, Wu M: GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nature cell biology* 2018, 20(4):492-502.
16. Heo JB, Sung S: Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 2011, 331(6013):76-79.
17. Feretzaki M, Pospisilova M, Valador Fernandes R, Lunardi T, Krejci L, Lingner J: RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops. *Nature* 2020, 587(7833):303-308.
18. Grossi E, Raimondi I, Goni E, Gonzalez J, Marchese FP, Chapaprieta V, Martin-Subero JI, Guo S, Huarte M: A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter

- regions. *Nat Commun* 2020, 11(1):936.
19. Coudert AE, Pibouin L, Vi-Fane B, Thomas BL, Macdougall M, Choudhury A, Robert B, Sharpe PT, Berdal A, Lezot F: Expression and regulation of the Msx1 natural antisense transcript during development. *Nucleic acids research* 2005, 33(16):5208-5218.
20. Palazzo AF, Koonin EV: Functional long non-coding RNAs evolve from junk transcripts. *Cell* 2020, 183(5):1151-1161.
21. Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, Chua NH: Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome research* 2014, 24(3):444-453.
22. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA *et al*: A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011, 472(7341):120-124.
23. Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D: A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 2012, 149(4):819-831.
24. Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, Suzuki K, Kawai J, Carninci P, Ohtomo Y, Murakami K *et al*: Antisense transcripts with rice full-length cDNAs. *Genome biology* 2003, 5(1):R5.
25. Zheng XM, Chen J, Pang HB, Liu S, Gao Q, Wang JR, Qiao WH, Wang H, Liu J, Olsen KM *et al*: Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. *Science advances* 2019, 5(12):eaax3619.
26. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW: The antisense transcriptomes of human cells. *Science* 2008, 322(5909):1855-1857.
27. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J *et al*: Antisense transcription in the mammalian transcriptome. *Science* 2005, 309(5740):1564-1566.
28. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C *et al*: The transcriptional landscape of the mammalian genome. *Science* 2005, 309(5740):1559-1563.
29. Oono Y, Yazawa T, Kanamori H, Sasaki H, Mori S, Matsumoto T: Genome-wide analysis of rice cis-natural antisense transcription under cadmium exposure using strand-specific RNA-Seq. *BMC genomics* 2017, 18(1):761.
30. Zhou X, Sunkar R, Jin H, Zhu JK, Zhang W: Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. *Genome research* 2009, 19(1):70-78.
31. Lu T, Zhu C, Lu G, Guo Y, Zhou Y, Zhang Z, Zhao Y, Li W, Lu Y, Tang W *et al*: Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice. *BMC genomics* 2012, 13:721.
32. Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, Chen YQ: Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual

- reproduction of rice. *Genome biology* 2014, 15(12):512.
33. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH: Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *The Plant cell* 2012, 24(11):4333-4345.
34. Xu J, Wang Q, Freeling M, Zhang X, Xu Y, Mao Y, Tang X, Wu F, Lan H, Cao M *et al*: Natural antisense transcripts are significantly involved in regulation of drought stress in maize. *Nucleic acids research* 2017, 45(9):5126-5141.
35. Siegel TN, Hon CC, Zhang Q, Lopez-Rubio JJ, Scheidig-Benatar C, Martins RM, Sismeiro O, Coppee JY, Scherf A: Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC genomics* 2014, 15:150.
36. Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson DA, Friedman N, Regev A: Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome biology* 2010, 11(8):R87.
37. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 2005, 123(7):1279-1291.
38. Li Y, Li X, Yang J, He Y: Natural antisense transcripts of MIR398 genes suppress microR398 processing and attenuate plant thermotolerance. *Nat Commun* 2020, 11(1):5351.
39. Kim D, Landmead B, Salzberg SL: HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015, 12(4):357-U121.
40. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015, 33(3):290-+.
41. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI: Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids research* 2018, 46(D1):D335-D342.
42. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G: CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research* 2007, 35(Web Server issue):W345-349.
43. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* 2016, 44(D1):D279-285.
44. Sun Q, Csorba T, Skourtis-Stathaki K, Proudfoot NJ, Dean C: R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* 2013, 340(6132):619-621.
45. Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C: Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. *Molecular cell* 2014, 54(1):156-165.
46. Shanker AK, Maheswari M, Yadav SK, Desai S, Bhanu D, Attal NB, Venkateswarlu B: Drought stress responses in crops. *Functional & integrative genomics* 2014, 14(1):11-22.

47. Colebrook EH, Thomas SG, Phillips AL, Hedden P: The role of gibberellin signalling in plant responses to abiotic stress. *The Journal of experimental biology* 2014, 217(Pt 1):67-75.
48. Chen X, Ding Y, Yang Y, Song C, Wang B, Yang S, Guo Y, Gong Z: Protein kinases in plant responses to drought, salt, and cold stress. *Journal of integrative plant biology* 2021.
49. Liu F, Quesada V, Crevillen P, Baurle I, Swiezewski S, Dean C: The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Molecular cell* 2007, 28(3):398-407.
50. Liu F, Marquardt S, Lister C, Swiezewski S, Dean C: Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science* 2010, 327(5961):94-97.
51. Canzio D, Nwakeze CL, Horta A, Rajkumar SM, Coffey EL, Duffy EE, Duffie R, Monahan K, O'Keeffe S, Simon MD *et al*: Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin alpha promoter choice. *Cell* 2019, 177(3):639-653 e615.
52. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H: Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 2008, 451(7175):202-206.
53. Guo CJ, Ma XK, Xing YH, Zheng CC, Xu YF, Shan L, Zhang J, Wang S, Wang Y, Carmichael GG *et al*: Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* 2020, 181(3):621-636 e622.
54. Ma JF, Goto S, Tamai K, Ichii M: Role of root hairs and lateral roots in silicon uptake by rice. *Plant physiology* 2001, 127(4):1773-1780.
55. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015, 33(3):290-295.
56. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010, 28(5):511-515.
57. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L: Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 2013, 31(1):46-53.
58. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z: agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic acids research* 2017.

Figures

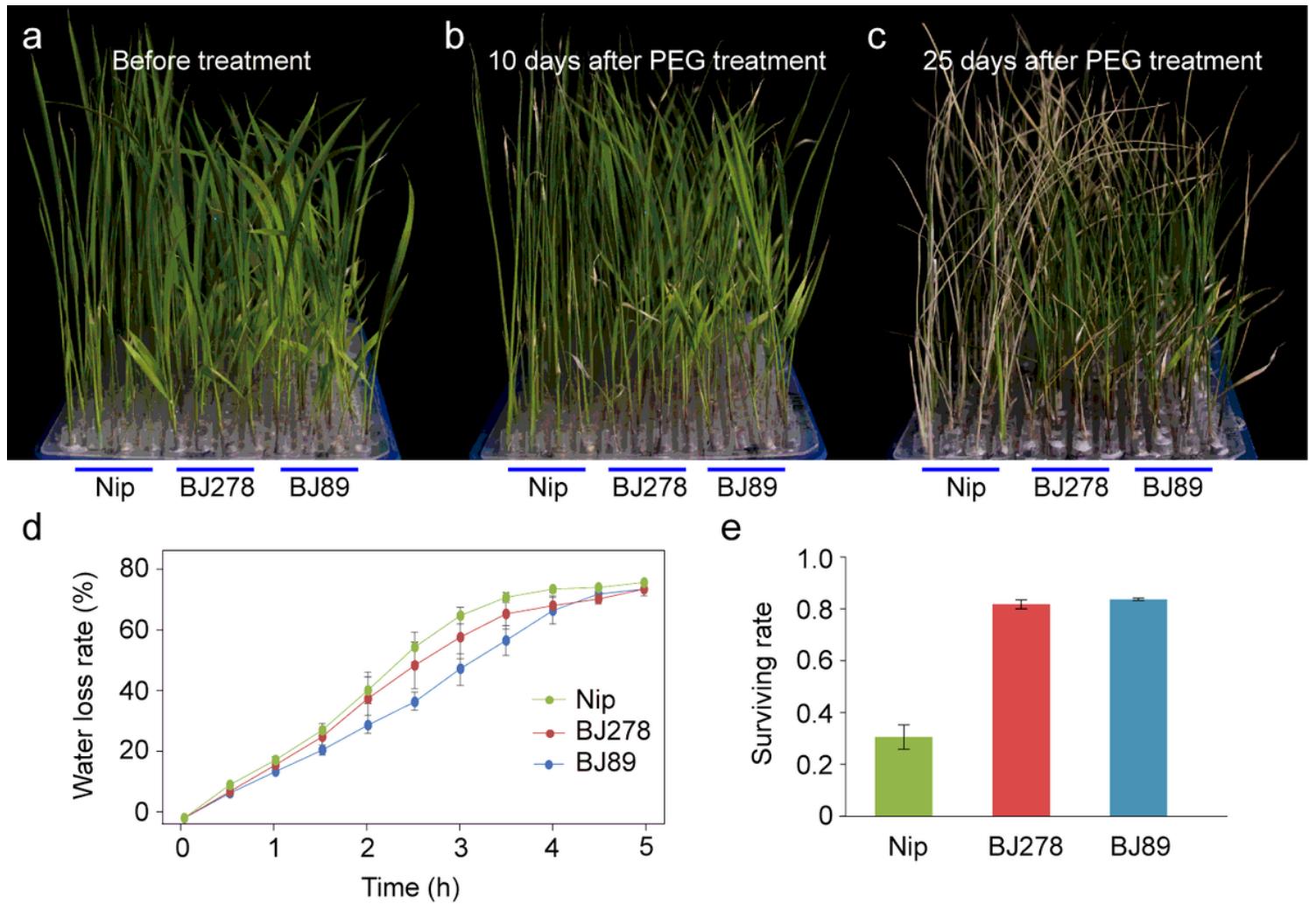


Figure 1

Phenotypic analysis of cultivated rice (*Nipponbare*; Nip) and *O. nivara* accessions (BJ89 and BJ278) before and after treatment with polyethylene glycol (PEG) to induce drought stress. (a) Photographs of 12-day-old seedlings captured before the PEG treatment. (b, c) Photographs of 12-day-old seedlings captured 10 days (b) and 25 days (c) after the PEG treatment. (d) Water loss rate of leaves. (e) Survival rate of seedlings after the PEG treatment.

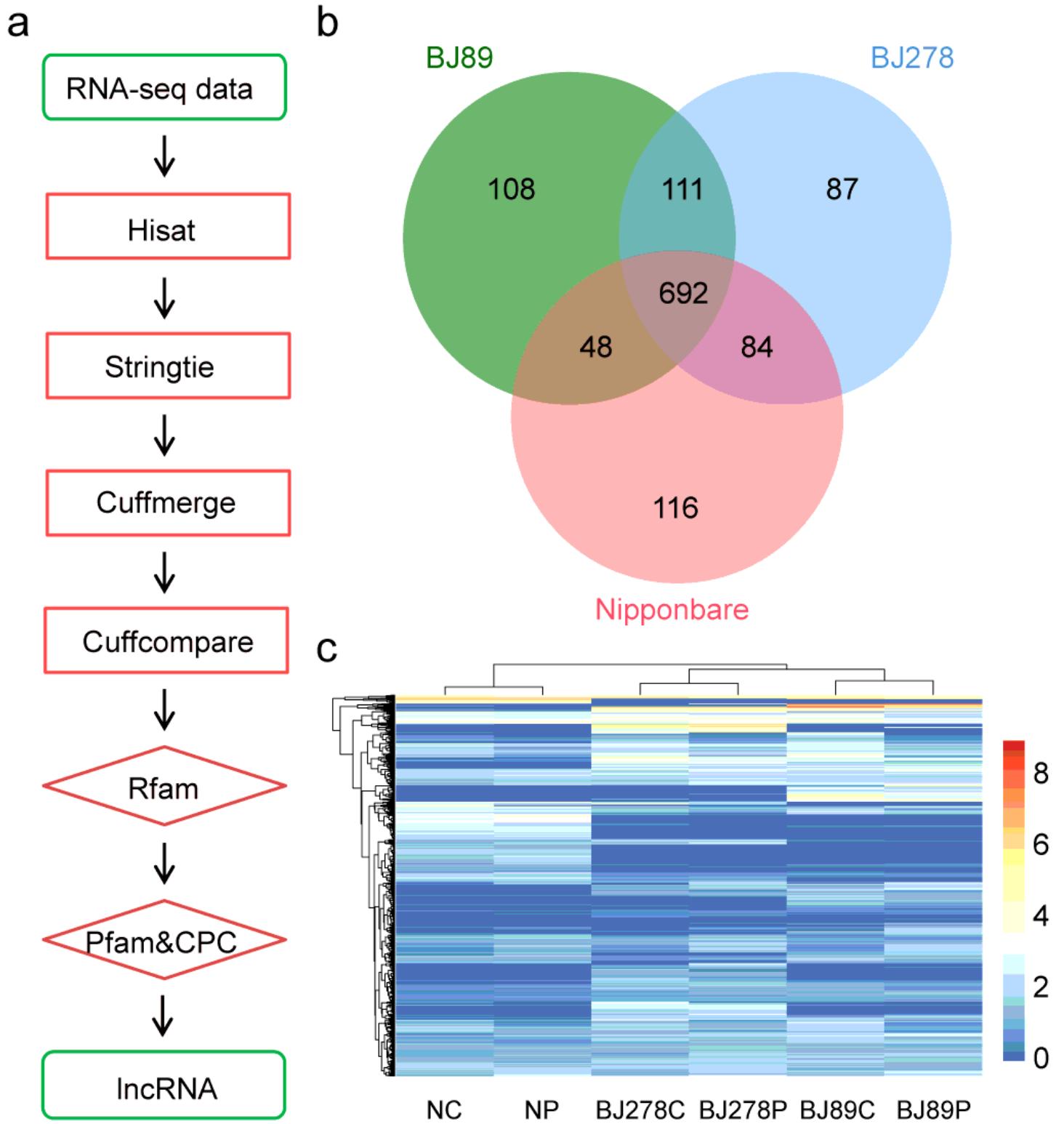


Figure 2

Identification and expression profiles of long noncoding RNAs (lncRNAs) in Nip, BJ89, and BJ278. (a) Flow chart showing the procedure used to identify lncRNAs. (b) Number of lncRNAs in Nip, BJ89, and BJ278. (c) Expression profiles of lncRNAs in Nip, BJ89, and BJ278 under control and drought stress conditions. NC and NP indicate Nip samples under control and drought stress conditions, respectively;

278C and 278P represent BJ278 samples under control and drought stress conditions, respectively; 89C and 89P represent BJ89 samples under control and drought stress conditions, respectively.

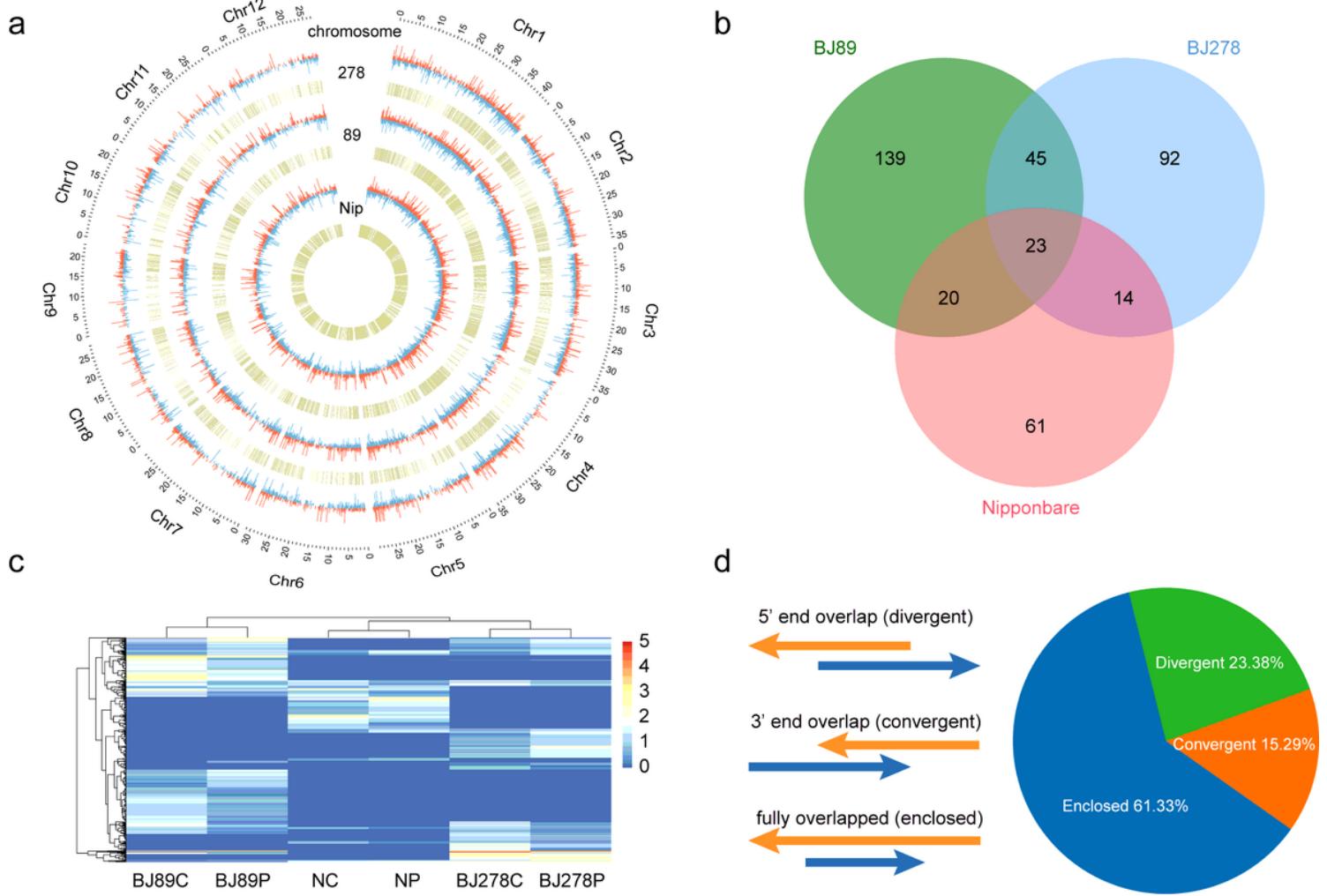


Figure 3

Genome wide identification of natural antisense transcript (NAT) pairs in Nip, BJ89, and BJ278. (a, b) Expression patterns (a) and Venn diagrams (b) of differentially expressed lncRNAs in the three accessions. (c) Genome-wide distribution of NAT pairs and differentially expressed genes (DEGs) under drought stress conditions. Red and blue lines represent up- and downregulated genes, respectively, under drought stress. Yellow lines indicate the genomic positions of NAT pairs. (d) Classification of NAT pairs according to their direction of transcription and the overlap region between sense and antisense transcripts. Orange arrows indicate sense transcripts, and blue arrows indicate antisense transcripts.

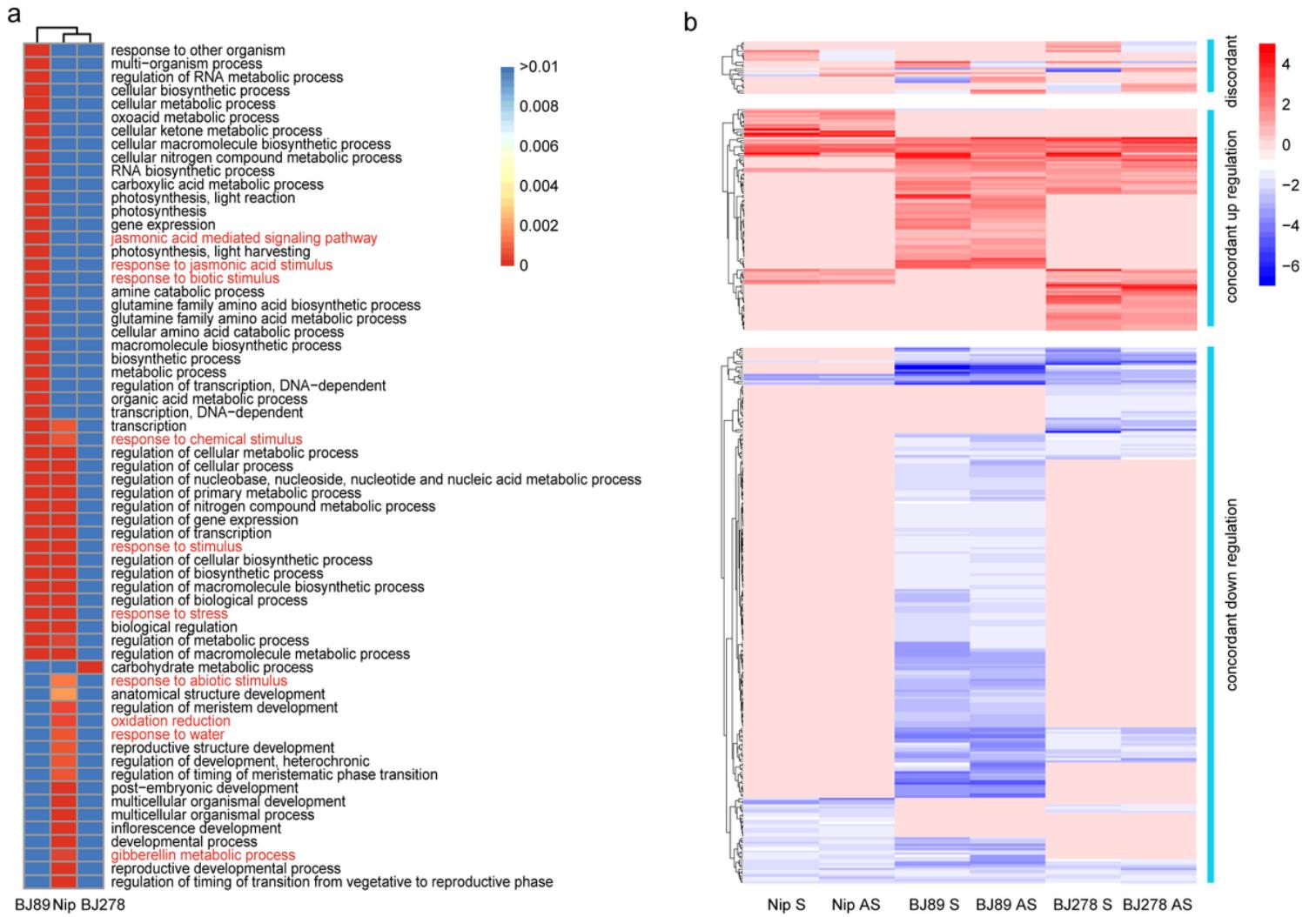


Figure 4

Analysis of drought stress-responsive genes and NAT pairs. (a) Gene ontology (GO) enrichment analysis of DEGs identified in Nip, BJ89, and BJ278 under drought stress conditions. Colors of the heatmap indicate P-values of GO enrichment results. GO terms related to drought stress are highlighted in red. (b) Discordant and concordant NAT pairs under drought stress. Pink rectangles indicate NAT pairs showing no difference in expression between drought stress and control conditions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile4.xlsx
- Additionalfile5.xlsx

- Additionalfile6.xlsx
- Additionalfile7.xlsx