Linkage analysis, GWAS, transcriptome analysis to identify candidate genes for rice seedlings in response to high temperature stress

Zhaoran Wei[1,2*], Qiaoling Yuan[1*], Hai Lin[1*], Xiaoxia Li[1], Chao Zhang[1], Hongsheng Gao[1], Bin Zhang[1], Huiying He[1], Tianjiao Liu[1], Shandang Shi[1], Zhenyu Gao[3], Lingrang Kong[2], Qian Qian[1,3,#], Lianguang Shang[1,#]

[1]*Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China;*

[2]*State Key Laboratory of Crop Biology, College of Agriculture, Shandong Agricultural University, Tai'an 271018, Shandong, China;*

[3]*State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China;*

* These authors contributed equally to this work.

#Correspondence should be addressed to L.S. (shanglianguang@163.com) and Q.Q. (qianqian188@hotmail.com)

ABSTRACT

Background: With the increase of global temperature, the problem of rice yield decline caused by the rising temperature is becoming more and more prominent, which urgently needs to be solved. Mining heat-resistant genes and applying them to rice breeding is a feasible and effective way to solve the problem.

Result: Three main biomass traits, including shoot length, dry weight, and fresh weight, changed after abnormally high temperature treatment in the seedling stage of a recombinant inbred line and natural groups. Based on a comparison of the results of linkage analysis and genome-wide association analysis, the presence of two sections with lengths of 57 kb and 69 kb in *qDW7* and *qFW6*, respectively, was associated

with the rice response to abnormally high temperatures in the seedling stage. Meanwhile, based on integrated transcriptome analysis, some genes are considered as candidate genes for subsequent research and analysis. Combining with reports of known genes and analysis of homologous genes, it was found that there are 8 genes in candidate genes that need to be focused in subsequent research.

Conclusions: The results indicated several relevant beneficial loci reacted to heat stress in rice seedling stage, which would help researchers to further discover excellent heat-resistant genes that can be applied to rice heat-resistant breeding.

Keywords: Linkage analysis, GWAS, Transcriptome analysis, Rice seedling, High-temperature-mediated growth response.

Background

Rice (*Oryza sativa* L.) is a main food crop worldwide, especially in Asian countries. As an Asian country, China has a population of 1.4 billion, and 2/3 of the people's main food is rice. Fortunately, despite of the large population of eating rice as main food in China, China is the world's largest rice producer, whose rice production ranks first in the world[1]. But, as the global industrialization process accelerates, greenhouse gases, such as carbon dioxide (CO_2) and methane (CH_4), have caused the global temperature to rise continually. The rising high temperature is affecting the rice yield worldwide. Even if some rice varieties from tropical regions have formed ecological characteristics adapted to high temperature environments, their yields will also be affected once the temperature exceeds their appropriate temperature[2]. The ambient temperature exceeding the appropriate temperature during rice growth and development would lead to yield reduction[3]. Some high-temperature sensitive cultivars may have a decrease in yield of more than 10% for every 1℃ increase[4].

Early seedling growth is important for plant morphogenesis, but high-temperature stress during this period is unfavorable for plant growth and development[5]. When rice seedlings experience high temperature above 35℃, the protein type and content of leaves change considerably[6-7], and the shoot length, dry weight, and other traits are negatively affected[8].

In order to further understand the mechanism of rice responding to high-temperature stress, some research groups have used mapping populations to explore the effects of high-temperature stress on rice and obtained heat-tolerant genes or QTLs at different physiological stages[5-6,9-11]. Linkage analysis has been confirmed as a feasible method in mining rice heat-tolerant genes. However, as the number of inbred line populations used in linkage analysis is limited, the quantitative trait locus (QTL) results have a wide range[12]. The detected QTL results are difficult to be used directly in breeding, which requires time-consuming fine mapping[2]. Association analysis using natural populations can effectively circumvent the problem of linkage analysis. So far, there are few reports on using natural populations to mine beneficial loci in response to heat stress during rice vegetative growth.

Genome-wide association analysis (GWAS) has been used to identify causal loci for important agronomic traits, such as grain length[13-14] and the 1000-grain weight[13] in rice. Previously, some key SNPs on chromosome 4 affecting floret fertility were detected under high-temperature conditions using GWAS population[15]. However, the influences of the population genetic structure and allele frequency were easily overlooked in GWAS, which increased the probability of false positives in linkage disequilibrium (LD) mapping results[16]. Combining the results from linkage analysis and correlation analysis has facilitated the attempt to explore quantitative agronomic traits. It has been confirmed that combining these methods can further improve the

efficiency and accuracy of QTL calling[17] and can effectively narrow the intervals of major QTLs in rice[18]. In this study, a natural rice population and the recombinant inbred line (RIL) population derives from the 93-11×PA64s were used to compare the differences in seedling traits with or without heat stress treatment, and then linkage analysis, association analysis and RNA_seq analysis were conducted. QTLs related to heat stress in the rice seedling stage were identified, which provided a basis for later fine mapping and gene editing in heat-resistant molecular-assisted selection breeding.

Results

Phenotypes of RIL and GWAS population

To detect the influence of heat treatment on rice seedlings, the phenotypic traits of RIL and the natural rice population with or without heat treatment were statistically analyzed (Fig. 1). The average and extreme values of the shoot length (SL), dry weight (DW), and fresh weight (FW) of RIL and natural population decreased after heat stress treatment, which confirmed that heat stress had a negative impact on rice growth. The measured values between the RIL control and the treatment groups were different, and the measured values were higher than the average value of the corresponding parents, which suggested the existence of super-parent separation of biomass traits (Table S1). The measured traits of the natural rice population and the RIL were tested to determine whether they followed a normal distribution (Fig. S1 and S2). It showed that shoot length, dry weight, and fresh weight were basically in accordance with a normal distribution in both RIL and GWAS population. According to the traits data, it was determined that the biomass traits of rice seedlings in the experiment are quantitative traits controlled by typical polygenes (Table S1). Biomass traits of the control and treatment groups in RIL were significantly lower than those in

the parental groups. However, the survival rate (SR) of the RIL was higher compared to the parents (Fig. 1, Table S1), which suggested that the RIL had pyramided more heat-tolerant genes from the parental materials during the recombination process. The difference in biomass traits between the control and treatment groups of GWAS population was less than that of RIL, while the survival rate of GWAS population was significant higher than that of RIL (Fig. 1, Table S1). These results inferred that there were more natural heat-tolerant genes and genetic diversity in the natural population. Exploring the heat-resistant QTL of rice seedlings provides a possibility for comparison of control and post-treatment QTL mapping to obtain QTLs for heat-resistant effects.

The correlations among biomass traits of RIL and GWAS population were analyzed separately (Fig. 2). The results showed that the correlations of the biomass traits between the two groups were different. Among them, the strongest correlation was found between dry weight and fresh weight of the GWAS treatment group, and the correlation coefficient was 0.91. In the RIL population, there was also a strong correlation between dry weight and fresh weight of the RIL treatment group, with a correlation coefficient of 0.75. These results suggested that there should be genes in rice seedlings that affect both dry and fresh weight of rice seedlings simultaneously under high temperature stress, and in this way respond to high temperature stress. Previous studies have confirmed that seedling biomass traits can reflect rice seedling vigor[19]. But in this study, although the effect of high temperature stress on rice seedlings can be directly reflected by biomass traits, correlation analysis showed that there is a very weak correlation between survival rate traits that directly reflect the vigor of rice seedlings and the biomass of rice seedlings. This indicates that biomass traits were not the unique manifestation in response to high temperature stress in rice,

and that survival rates and biomass traits were regulated by different genes.

QTL mapping of RIL

Using the high density map of RIL in 2013[20], QTL mapping of the RIL with or without heat stress treatment was carried out. In total, 20 QTLs, distributed on the chromosomes 2, 3, 4, 6, 7, 9,11, and 12, were detected (Table 1). The detection of multiple QTLs for the same trait indicates that the biomass trait is indeed a quantitative trait. There are several known genes related to plant heat tolerance in the above QTL intervals, including *GS8*[21] and *abl1*[22]. All of these known genes confirmed the accuracy of the linkage analysis. The results showed that *qDW11*, a locus for dry weight trait in the treatment group, had the highest logarithm of odds (LOD) value (6.72). Meanwhile, we found *qDW11* and another loci *qFW11.1* for fresh weight in the treatment group are two identical intervals, which is also verified by the correlation analysis results (Fig. 2). In the analysis of dry and fresh weights of the control group, the same set of QTLs (*qCDW11* and *qCFW11*) were also identified on chromosome 11. However, the same interval for dry and fresh weight detected in the control group was inconsistent with that in the treatment group, indicating that the QTL identified in the treatment group responded to the influence of high temperature. Therefore, we speculate that *qDW11* and *qFW11.1* contain genes that respond to high temperature stress, which are not expressed under normal growth conditions, and we will focus on this candidate interval later.

GWAS of heat-related traits in the seedling stage

A genome-wide association analysis of 255 natural rice lines was performed using the existing genome-wide coverage of 14,779,691 single-nucleotide polymorphism (SNP)

data from 3k database[23]. It is generally considered that linkage disequilibrium was between 100 kb and 200 kb in rice[18]. In this study, the SNP threshold for the significantly associated sites is P value $<10^{-6}$ and the significant SNPs sites within the 100 kb interval are considered candidate loci (Table S2). There were a large number of different significant loci in the heat treatment and control groups, which suggested that the high temperature in the seedling stage might affect the growth of the natural population through these loci (Fig 3 and S3). According to the stringent screening conditions, only a small number of reasonable SNPs were associated with the shoot length, so the P value was reduced to $< 10^{-5}$ for subsequent analysis. In this way, a total of 10 intervals for biomass traits were identified in the control groups, while 29 intervals for biomass traits were present on all of the chromosomes other than chromosome 3 in the heat treatment group (Table S2). For the dry weight and fresh weight of the treatment groups, the consistent intervals, *qDW_ind8* (3.09Mb-3.29Mb) and *qFW_ind8* (3.09Mb-3.29Mb), were identified on chromosome 8. Therefore, there are important candidate genes controlling each biomass traits under high temperature stress in this interval. It can be inferred that the expression of one or more genes in this interval was in response to high temperature stress by adjusting the dry weight and fresh weight characters of rice seedlings.

There are 77 intervals containing 283 SNPs for survival rate after heat treatment (Fig. 4a, Table S2). Among them, *qSR_ind9-3* on chromosome 9 contains gene *Ugp1*, a known heat-tolerant gene at the heading stage[24](Fig. 4c). As *Ugp1* is expressed during the whole growth and development period of rice[24], it was suggested that this gene could also had an effect on the rice seedling response to heat stress. To further determine the natural variation of *Ugp1*, we performed haplotype analysis using all of the non-synonymous SNPs in the coding region of this gene (Fig. 4d). A total of 10

SNPs in the coding sequence (CDS) region were detected, and four major haplotypes were identified based on the 10 SNPs. Hap.1 is the main haplotype in nature population, and the varieties belonging to Hap.1 had higher survival rate after high temperature stress treatment in rice seedling stage. It can be found that the expression of *Ugp1* does have an effect on the response of rice seedlings to high temperature stress, and the varieties belonging to Hap.1 had higher survival rate after high temperature stress treatment in rice seedling stage. It can be found that the expression of *Ugp1* does affect the response of rice seedlings to high temperature stress. Among 255 materials, a total of 191 varieties belong to the Hap.1 haplotype, which proves that Hap.1 has a higher utilization rate in the population. Compared with Hap.1, the survival rate of Hap.2 after the change of chr9_21920470 was low, but the biomass increased significantly. For Hap.3, all sites except chr9_21920470 changed. Although the survival rate of Hap.3 is also lower than that of Hap.1, the decline rate of survival rate is lower than that of Hap.2. Therefore, we judge that the chr9_21920470 locus has a stronger effect on the survival rate of rice seedlings than other loci. In Hap.4, where chr9_21920592 changed, compared with Hap.1, shoot length is lower. From this, we judge that this site may be related to plant growth.

Compare results from linkage analysis and GWAS

We compared the QTLs obtained by linkage analysis with the candidate intervals identified by GWAS. Two groups of candidate regions identified on chromosome 7 and chromosome 6 were co-located by linkage analysis and association analysis. There was a 200 kb (5.59 Mb-5.70 Mb) coincidence interval in *qFW6* and *qFW_ind6* for the fresh weight trait in the treatment group. Analyzing the LD attenuation distance of this interval (Fig. 5), we further reduced the candidate interval to 69 kb

(5.62 Mb-5.69 Mb). According to MSU V7.0[25] annotated information, there are 13 candidate genes in this interval, including 12 genes involved in expression (Table S3). The same analysis was performed on the interval of 200 kb (17.82 Mb-18.02 Mb) in *qDW7* and *qDW_ind7* on chromosome 7, we reduced the candidate interval on chromosome 7 to 57 kb (17.90 Mb-17.95 Mb), and identified 6 genes (Table S3). To further investigate candidate genes, we performed homology analysis on the obtained genes. The results of homology analysis showed that the homolog gene *LECRK-VII.2* of *LOC_Os06g10790* is involved in the response to high temperature and drought in *Arabidopsis*[26]. The homologous gene *UGT73B3*[27] of *LOC_Os06g10860* and the homologous gene *UGT85A2*[28] of *LOC_Os07g30330* are involved in the stress response of *Arabidopsis*. The homologous gene *PHT3*[29] of *LOC_Os06g10810* is related to the redox reaction of *Arabidopsis thaliana*. These homologous genes on the one hand confirm the accuracy of the experimental results, on the other hand, further narrow the range of candidate genes in this study.

Transcriptome analysis by RNA sequencing

In order to study the transcriptome response of *indica* rice to high temperature in the seedling stage, independent of the above genomic analysis, two heat-tolerant varieties and two heat-sensitive varieties were selected for ribonucleic acid sequencing (RNA_seq) analysis. After comparing the gene expression between the heat-resistant groups and the heat-sensitive groups, it was determined that a total of 529 genes were differentially expressed in the four groups (Fig. 6a), and it was considered that there were key genes responding to high temperature stress in rice seedling. Kyoto encyclopedia of genes and genomes (KEGG) analysis was performed on 529 differential expressed genes (DEGs). The results showed that 75 DEGs were

228 significantly enriched in pathways such as metabolic pathways, pyruvate metabolism,
229 starch and sucrose metabolism (Table S4). *OsCML4*[30], which is enriched in the
230 mitogen-activated protein kinase (MAPK) signaling pathway-plant pathway, has been
231 shown to improve the rice tolerance by removing reactive oxygen species and
232 inducing other stress related genes in a form independent of ABA Survival rate. This
233 indicates that there are genes in DEGs genes that can respond to high temperature
234 stress for subsequent research. Further, we identified *LOC_Os01g09450*,
235 *LOC_Os03g59040* and *LOC_Os12g42980* in the GWAS candidate intervals
236 *qSR_ind1-1, qSR_ind3-9*, and *qSR_ind12-7*, respectively. The three genes were
237 located in the KEGG pathway in Plant hormone signal transduction, Metabolic
238 pathways, and Cysteine and methionine metabolism pathways, and the expression of
239 *LOC_Os03g59040* was confirmed to be related to the stomatal conductance of rice[31].
240 We conclude that *LOC_Os01g09450*, *LOC_Os03g59040* and *LOC_Os12g42980*
241 could be used as candidate genes for rice seedlings in response to high temperature
242 stress. Based on the gene ontology (GO) enrichment analysis, most genes were
243 enriched in the oxidation reduction, oxidoreductase activity, transmembrane transport
244 and others (Fig. 6b). We focused on the 44 genes enriched in oxidation reduction and
245 oxidation activity activities. Among the 44 genes identified, there are 10
246 down-regulated genes and 22 up-regulated genes (Fig. 6c, d). After a comparison with
247 these genes and the intervals obtained in GWAS, it was found that *LOC_Os02g12890*
248 was located on the *qSR_ind2-1* for survival rate. It was suggested that
249 *LOC_Os02g12890* could be the candidate gene which may respond to high
250 temperature stress in the rice seedling stage and directly affect the heat tolerance of
251 rice seedlings.
252 In summary, a total of 23 candidate genes were co-localized by two or more methods,

including 19 genes co-located by linkage analysis and GWAS and genes in 4 related pathways by GWAS and transcriptome analysis. 8 genes, such as *LOC_Os06g10790*, *LOC_Os06g10860*, *LOC_Os07g30330*, *LOC_Os06g10810*, *LOC_Os01g09450*, *LOC_Os03g59040*, *LOC_Os12g42980*, *LOC_Os02g12890*, etc. have been reported to be related to the plant response to heat stress due to homology analysis or the focus has been reported in the follow-up research on the plant response to heat stress.

Discussion

The combination of linkage, association analysis and transcriptome provides effective information for target genes mining

Abnormally high temperatures can affect the growth and development of rice. At present, temperatures in some areas have reached the critical value of the optimal temperature for rice growth, and rising temperatures would lead to a decline in rice production[32]. Therefore, exploring the genetic mechanism of heat resistance in rice is the key to cultivation of heat-resistant rice to adapt to the global warming environment. Linkage analysis based on RIL and association analysis based on a natural population are two complementary approaches to revealing the candidate genetic variation leading to traits of interest. GWAS analysis can evaluate the diversified effect of many alleles with higher mapping resolution, which can make up for the inferiority of QTL analysis that the number of natural allelic diversity is limited. In our study, a total of 116 loci were detected by GWAS, and only 20 loci were detected by linkage analysis. In the present study, two colocalized regions, a 200 kb interval in *qFW6* and *qFW_ind6,* and another 200 kb interval in *qDW7* and *qDW_ind7*, were detected by both linkage analysis and association analysis. These two reliable intervals should contain important candidate genes contributing to heat resistance in rice seedling stage. Further functional verification of these genes will reveal their underlying genetic and molecular mechanisms.

The transcriptome analysis is an effective approach for mining genes associated with

given trait. In our study, we sequenced the transcriptome of two heat-tolerant varieties and two heat-sensitive varieties, compared the DEGs with GWAS, and obtained four candidate genes. These results suggest that the combination of linkage, association analysis and transcriptome is a powerful method for mining target genes responsible for rice heat resistance.

Identification of candidate genes

To further narrow the scope of candidate genes, transcriptome analysis and gene homology analysis were applied to this study. Among the 19 genes identified by both linkage and association analysis, the homologous genes of *LOC_Os06g10860*, *LOC_Os06g10790*, *LOC_Os07g30330* and *LOC_Os06g10810* have been confirmed to be involved in life activities related to high temperature stress in *Arabidopsis*. These genes will be used as reliable candidate genes in subsequent research. Among the four candidate genes, *LOC_Os06g10790* was determined to participate in high temperature and drought responses in homology analysis. In subsequent studies, we will focus on this candidate gene.

At the same time, in the transcriptome analysis, *LOC_Os02g12890*, *LOC_Os01g09450*, *LOC_Os03g59040* and *LOC_Os12g42980* was found in the extreme group among the differential genes. These genes are located in the candidate interval which was identified by GWAS. The transcriptome analysis and GWAS are mutually verified and supplemented, indicating that *LOC_Os02g12890*, *LOC_Os01g09450, LOC_Os03g59040* and *LOC_Os12g42980* can also be used as a candidate gene in subsequent analysis.

Exploration of genes related to high temperature in rice seedling

Research on the high temperature stress of rice was concentrated only on the late stage of rice growth and development, as an abnormally high temperature during

reproductive growth has a direct impact on rice yield[6,33-34]. However, previous experiments have also proved that heat treatment affects rice seedling biomass, and high temperatures above 40℃ or long-term heat treatment affect the survival rate of rice[3-4,6-7,11,35]. In particular, Li et al. completed the cloning of the heat-resistant gene *TT1* in 2015[11], which greatly facilitated the breeding of heat-resistant and heat-resistant molecules in the later period. It is worth mentioning that the *TT1* gene was identified in wild African rice, and this gene was not identified in this experiment. The wild rice material analysis was not used in this experiment, so it is speculated that this is the reason that the gene was not found in this experiment. Nevertheless, the discovery of *TT1* did further confirms the feasibility and necessity of studying candidate genes that respond to high temperature stress at the seedling stage.

Experiments have shown that various biomass traits at the seedling stage can show different responses in different varieties of seedlings to high temperature. The three traits, SL, FW and DW, changed significantly in the RIL and natural rice population after heat treatment (Fig. 1). Utilizing the different characteristics of the RIL population and the natural population, linkage analysis and association analysis were performed and obtained multiple candidate intervals. Researchers have used linkage analysis or association analysis to identify heat-tolerant genes in rice previously[9-12,36-38]. In the study of other agronomic traits, it has also been confirmed that the combined application of linkage and association analysis can obtain more accurate candidate intervals[18]. In this study, we narrowed the candidate intervals to 69 kb on chromosome 6 (5.62 Mb-5.69 Mb) and the interval 57 kb on chromosome 7 (17.90 Mb-17.95 Mb) based on the overlapping intervals and the rate of linkage disequilibrium decay of SNPs in the linkage analysis and association analysis. 19 candidate genes were identified in the two fragments according to the gene annotation

(Table S3), which provided reliable genetic evidence for fine mapping and more reliable sequencing.

Identification of *Ugp1* favorable haplotypes

In order to obtain the ideal high-temperature tolerance of rice, we also determined the favorable haplotype of the known gene *Ugp1*[24] identified in the GWAS results. By determining the survival rate of high-temperature stress seedlings among different haplotypes, it was confirmed that the haplotype change had an impact on the survival rate of rice. The results of haplotype analysis indicated that the major haplotype consisting of non-synonymous SNPs in the coding sequence (CDS) within a single locus, represented the haplotypes of most varieties. At the same time, the haplotype analysis of *Ugp1*[24] confirmed that haplotype analysis can identify key sites and favorable alleles of the target genes, providing a research direction for the identification of subsequent genes. This finding also facilitated the selection of optimal haplotypes for subsequent breeding.

Conclusions

In this study, three methods of linkage analysis, association analysis, and transcriptome analysis were used to complement and verify each other. New genes for rice seedlings responding to abnormally high temperatures were explored, and the genes identified by two or more methods were selected as candidate genes. Finally, based on homology analysis and known gene information, it was determined that eight candidate genes could be focused on in subsequent studies. These results provide a reference for rice heat-resistant gene cloning and molecular breeding.

METHODS

Plant materials

The rice variety 93-11, PA64s and their RIL were used for linkage analysis[20]. A natural population of 255 Asian cultivated rice varieties with high genetic diversity from the 3000 rice project[23] was used for association analysis.

Seedling culture and high-temperature stress treatment

Intact seeds of the RIL and natural population were selected and immersed in deionized water for one night, and then transferred to a 32-well tray filled with nutrient solution. The nutrient solution was the conventional nutrient solution configuration of the International Rice Research Institute[39]. The pH value of the nutrient solution was adjusted between 5.5 and 6, and the nutrient solution was replaced once every three days. On the 15th day, the seedlings of the RIL and natural population required to be heat-treated were placed in an incubator at 45 ℃ for 52 hours[10]. On the 20th day, the biomass traits, including plant height, dry weight, fresh weight, and seedling survival rate of RIL and natural population with or without heat stress treatment, were measured and recorded.

Statistical analysis of data

The phenotypic data with or without seedling heat stress treatment were statistically analyzed by R ×64 3. 4. 2.

Linkage analysis

Linkage analysis was performed according to a genetic map constructed in previous

study[20] by combining genotypic analysis and phenotypic identification of 124 RILs. The QTL of RIL was analyzed by composite interval mapping (CIM)[40] using Windows QTL Cartographer V2.5 software (http://statgen. ncsu. edu/qtlcarl/WQTLCart. htm, 2008). The threshold of logarithm of odds was set as 2.5 for the presence of QTLs and physical locations of the obtained QTLs were determined according to a high-density SNP-labeled recombination bin map[20]. The detected QTLs were named by reference to the method proposed by McCouch[41].

Association analysis

Association analysis of genotypes and phenotypes was performed with the efficient mixed-model association eXpedited (EMMAx)[42]. The genotype data were referenced with sequencing information released for the 3k rice database[23]. SNP site settings with a deletion rate greater than 40% and a minimum allele frequency (MAF) of less than 5% were filtered out. When the site had a significant level, $P < 10^{-5}$, it was considered to be associated with the trait and was used for subsequent analysis. The relevant drawings in the subsequent analysis were completed by the corresponding drawing package of R x64 3. 4. 2. The method for determining and naming GWAS candidate intervals refers to the method proposed by Zhao[43].

Transcriptome analysis

Two varieties were selected according to the highest and lowest survival rates of the natural population after heat treatment, respectively, and a total of four varieties were used for transcriptome analysis. The RNA was extracted using TRIzol reagent according to the manufacturer's guidelines, and two biological replicates were measured for each variety. In order to ensure that the RNA sequencing libraries had

high-quality RNA, agarose gel electrophoresis, and spectrophotometry were used to check RNA concentration and purity. TOPhat2 software[44] was used to align the cleanup data to the reference genome[25] and gene expression was quantified by fragment per kilobase million (FPKM) using the Cufflinks[44] default parameters. The genes with adjusted P values less than e^{-5} and log fold change (FC) absolute value higher than 0.58 were considered as differentially expressed genes (DEGs)[23]. Finally, Gene Ontology (GO) enrichment analysis was performed for DEGs using an online tool agriGO[45] and the GO terms were assigned to the GO categories biological process (BP), molecular function (MF), and cellular component (CC). The threshold of FDR<0.05 was selected to identify significant enriched GO terms.

Supplementary information

Table S1 Traits statics during the seedling stage of RIL, together with their parents and natural population.

Table S2 Candidate QTLs identified by GWAS.

Table S3 Annotation information of candidate genes identified in GWAS and linkage analysis.

Table S4 KEGG pathway analysis of DEGs.

Fig. S1 Normal distribution maps of biomass traits and survival rate in the RIL in control and treatmet groups.

Fig. S2 Normal distribution maps of biomass traits and survival rate in the natural population in control and treatment groups.

Fig. S3 QQplots for Genome-wide association analysis in the treatment and control groups of the natural population.

Abbreviations

BP: Biological process; CC: Cellular component; CDS: Coding sequence; CIM: Composite interval mapping; DEGs: Differential expressed genes; DW: Dry weight; EMMAX: Efficient mixed model association eXpedited; FC: fold change; FPKM: Fragments per kilobase million; FW: Fresh weight; GO: Gene ontology; GWAS: Genome-wide association analysis; KEGG: Kyoto encyclopedia of genes and genomes; LD: Linkage disequilibrium; LOD: Logarithm of odds; MAF: Minimum allele frequency; MAPK: Mitogen-activated protein kinase; MF: Molecular function; QTL: Quantitative trait locus; RNA_seq: Ribonucleic acid sequencing; RILs: Recombinant inbred lines; SL: Shoot length; SNP: Single nucleotide polymorphism; SR: Survival rate.

Authors' contributions

L.S., L.K. and Q.Q. designed the research, Z.W., Q.Y. and H.L. performed most of experiments and analyzed the data. B.Z., H.H., S.S., H.G., helped with characterizing the phenotypes; X.L., C.Z., T.L., helped with the transcriptomic analysis; L.S., Z.W. and Z.G. conceived the experiment and wrote the manuscript. All of the authors read and approved the final manuscript.

Funding

This work was supported by grants from the Agricultural Science and Technology Innovation Program, Shenzhen Science and Technology Program (Grant No. KQTD2016113010482651), Projects Subsidized by Special Funds for Science Technology Innovation and Industrial Development of Shenzhen Dapeng New District （Grant No. RC201901-05), Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515110557), China Postdoctoral Science Foundation

(Grant No. 2020M672903).

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

REFERENCES

1. Lei Z, Jianguo L, Yingjin H, et al. Mapping quantitative trait loci for heat tolerance at anthesis in rice using chromosomal segment substitution lines[J]. Breeding Science, 2016, 66 (3):358-366.
2. Driedonks N, Rieu I, Vriezen W H. Breeding for plant heat tolerance at vegetative and reproductive stages[J]. Plant Reproduction, 2016, 29 (1-2):67-79.
3. Akman Z. Comparison of high temperature tolerance in maize, rice and sorghum seeds by plant growth regulators[J]. Journal of Animal and Veterinary Advances, 2009, 8 (2):358-361.
4. Peng S, Huang J, Sheehy J E, et al. Rice yields decline with higher night temperature from global warming[J]. Proceedings of the National Academy of

Sciences, 2004, 101 (27):9971-9975.

5. Wei H, Liu J, Wang Y, et al. A Dominant Major Locus in Chromosome 9 of Rice (Oryza sativa L.) Confers Tolerance to 48 ℃ High Temperature at Seedling Stage[J]. Journal of Heredity, 2013, 104 (2):287-294.

6. Beáta Barnabás, Katalin Jäger, Attila Fehér. The effect of drought and heat stress on reproductive processes in cereals[J]. Plant Cell and Environment, 2008, 31 (1):11-38.

7. Han F, Chen H, Li X J, et al. A comparative proteomic analysis of rice seedlings under various high-temperature stresses[J]. Biochimica et Biophysica Acta, 2009, 1794 (11):1625-1634.

8. Wahid A, Close T J. Expression of dehydrins under heat stress and their relationship with water relations of sugarcane leaves[J]. Biologia Plantarum (Prague), 2007, 51 (1):104-109.

9. Jagadish S V K, Cairns J, Lafitte R, et al. Genetic Analysis of Heat Tolerance at Anthesis in Rice[J]. Crop Science, 2010, 50 (5):1633.

10. Lei D, Tan L, Liu F, et al. Identification of heat-sensitive QTL derived from common wild rice (Oryza rufipogon Griff.)[J]. Plant Science, 2013, 201-202:121-127.

11. Li X M, Chao D Y, Wu Y, et al. Natural alleles of a proteasome α 2 subunit gene contribute to thermotolerance and adaptation of African rice[J]. Nature Genetics, 2015, 47 (7):827-833.

12. Salvi S, Tuberosa R. To clone or not to clone plant QTLs: present and future challenges[J]. Trends in Plant Science, 2005, 10 (6):0-304.

13. Wang C, Chen S, Yu S. Functional markers developed from multiple loci in GS3 for fine marker-assisted selection of grain length in rice[J]. Theoretical and Applied

510 Genetics, 2011, 122 (5):905-913.
511 14. Si L, Chen J, Huang X, et al. OsSPL13 controls grain size in cultivated rice[J].
512 Nature Genetics, 2016.
513 15. Lafarge T, Bueno C, Frouin J, et al. Genome-wide association analysis for heat
514 tolerance at flowering detected a large set of genes involved in adaptation to
515 thermal and other stresses. [J]. Plos One, 2017, 12 (2):e0171254.
516 16. Yu J, Buckler E S. Genetic association mapping and genome organization of
517 maize[J]. Current Opinion in Biotechnology, 2006, 17 (2):155-160.
518 17. Lu Y, Zhang S, Shah T, et al. Joint linkage-linkage disequilibrium mapping is a
519 powerful approach to detecting quantitative trait loci underlying drought tolerance
520 in maize[J]. Proc Natl Acad Sci U S A, 2010, 107 (45):19585-19590.
521 18. Daware A V, Rishi S, Singh A K, et al. Regional Association Analysis of
522 MetaQTLs Delineates Candidate Grain Size Genes in Rice[J]. Frontiers in Plant
523 Science, 2017, 8:807-.
524 19. Mahender A, Anandan A, Pradhan S K. Early seedling vigour, an imperative trait
525 for direct-seeded rice: an overview on physio-morphological parameters and
526 molecular markers. [J]. Planta, 2015, 241 (5):1027-50.
527 20. Gao Z Y, Zhao S C, He W M, et al. Dissecting yield-associated loci in super
528 hybrid rice by resequencing recombinant inbred lines and improving parental
529 genome sequences[J]. Proceedings of the National Academy of Sciences, 2013, 110
530 (35):14492-14497.
531 21. Ashraf E K, Yong-Mei B, Kosala R, et al. The Rice R2R3-MYB Transcription
532 Factor OsMYB55 Is Involved in the Tolerance to High Temperature and Modulates
533 Amino Acid Metabolism[J]. PLoS ONE, 2012, 7 (12):e52030-.
534 22. Jin X F, Xiong A S, Peng R H, et al. OsAREB1, an ABRE-binding protein

responding to ABA and glucose, has multiple functions in Arabidopsis[J]. Bmb Reports, 2010, 43 (1):34.

23. Wang W, Mauleon R, Hu Z, et al. Genomic variation in 3, 010 diverse accessions of Asian cultivated rice[J]. Nature, 2018, 557 (7703).

24. Chen R, Zhao X, Shao Z, et al. Rice UDP-Glucose Pyrophosphorylase1 Is Essential for Pollen Callose Deposition and Its Cosuppression Results in a New Type of Thermosensitive Genic Male Sterility[J]. THE PLANT CELL ONLINE, 2007, 19 (3):847-861.

25. Kawahara Y, Bastide M D L, Hamilton J P, et al. Improvement of theOryza sativaNipponbare reference genome using next generation sequence and optical map data[J]. Rice, 2013, 6 (1):4.

26. Rizhsky, L. When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress[J]. PLANT PHYSIOLOGY, 2004, 134(4):1683-1696.

27. Lee T A, Bailey-Serres J. Integrative analysis from the epigenome to translatome uncovers patterns of dominant nuclear regulation during transient stress[J]. The Plant Cell, 2019, 31(11): 2573-2595.

28. Rehman H M, Nawaz M A, Shah Z H, et al. Comparative genomic and transcriptomic analyses of Family-1 UDP glycosyltransferase in three Brassica species and Arabidopsis indicates stress-responsive regulation[J]. Scientific reports, 2018, 8(1): 1-18.

29. Jia F, Wan X, Zhu W, et al. Overexpression of mitochondrial phosphate transporter 3 severely hampers plant development through regulating mitochondrial function in Arabidopsis[J]. PloS one, 2015, 10(6).

559 30. Yin X M, Huang L F, Zhang X, et al. OsCML4 improves drought tolerance
560 through scavenging of reactive oxygen species in rice[J]. Journal of Plant Biology,
561 2015, 58(1):68-73.
562 31. Manavalan L P, Xi C, Joseph C, et al. RNAi-mediated disruption of squalene
563 synthase improves drought tolerance and yield in rice[J]. Journal of Experimental
564 Botany(1):1.
565 32. Wassmann R, † S. V. K. Jagadish, Sumfleth K, et al. Regional Vulnerability of
566 Climate Change Impacts on Asian Rice Production and Scope for Adaptation[J].
567 Advances in Agronomy, 2009, 102 (09):91-133.
568 33. Ainsworth E A, Ort D R. How Do We Improve Crop Production in a Warming
569 World?[J]. PLANT PHYSIOLOGY, 2010, 154 (2):526-530.
570 34. Jagadish K S V, Craufurd P, Shi W, et al. A phenotypic marker for quantifying heat
571 stress impact during microsporogenesis in rice (Oryza sativa L.)[J]. Functional
572 Plant Biology, 2014, 41 (1):48.
573 35. Lyman N B, Jagadish K S V, Nalley L L, et al. Neglecting Rice Milling Yield and
574 Quality Underestimates Economic Losses from High-Temperature Stress[J]. PLOS
575 ONE, 2013, 8.
576 36. Jha U C, Bohra A, Singh N P, et al. Heat stress in crop plants: its nature, impacts
577 and integrated breeding strategies to improve heat tolerance[J]. Plant Breeding,
578 2014, 133 (6):679-701.
579 37. Liu J, Zhang C, Wei C, et al. The RING Finger Ubiquitin E3 Ligase OsHTAS
580 Enhances Heat Tolerance by Promoting H2O2-Induced Stomatal Closure in Rice.
581 [J]. Plant Physiology, 2015, 170 (1):429.

582 38. Ye C, Tenorio F A, Argayoso M A, et al. Identifying and confirming quantitative
583 trait loci associated with heat tolerance at flowering stage in different rice
584 populations[J]. BMC Genetics, 2015, 16 (1):41.
585 39. Yoshida S, Forno D A, Cock J H. Laboratory manual for physiological studies of
586 rice. [J]. 1971.
587 40. Zeng Z B. Precision mapping of quantitative trait loci. [J]. Genetics, 1994, 136
588 (4):1457-68.
589 41. Mccouch S R, CGSNL (Committee on Gene Symbolization, Nomenclature and
590 Linkage, Rice Genetics Cooperative). Gene Nomenclature System for Rice[J]. Rice,
591 2008, 1 (1):72-84.
592 42. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for
593 association studies[J]. Nature Genetics, 2012, 44 (7):821-8.
594 43. Zhao J, Yang W, Zhang S, et al. Genome-wide association study and candidate
595 gene analysis of rice cadmium accumulation in grain in a diverse rice collection[J].
596 Rice, 2018, 11(1).
597 44. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression
598 analysis of RNA_seq experiments with TopHat and Cufflinks[J]. NATURE
599 PROTOCOLS, 2012, 7 (3):562-578.
600 45. Du Z, Zhou X, Ling Y, et al. agriGO: a GO analysis toolkit for the agricultural
601 community[J]. Nucleic Acids Research, 2010, 38 (Web Server):W64-W70.

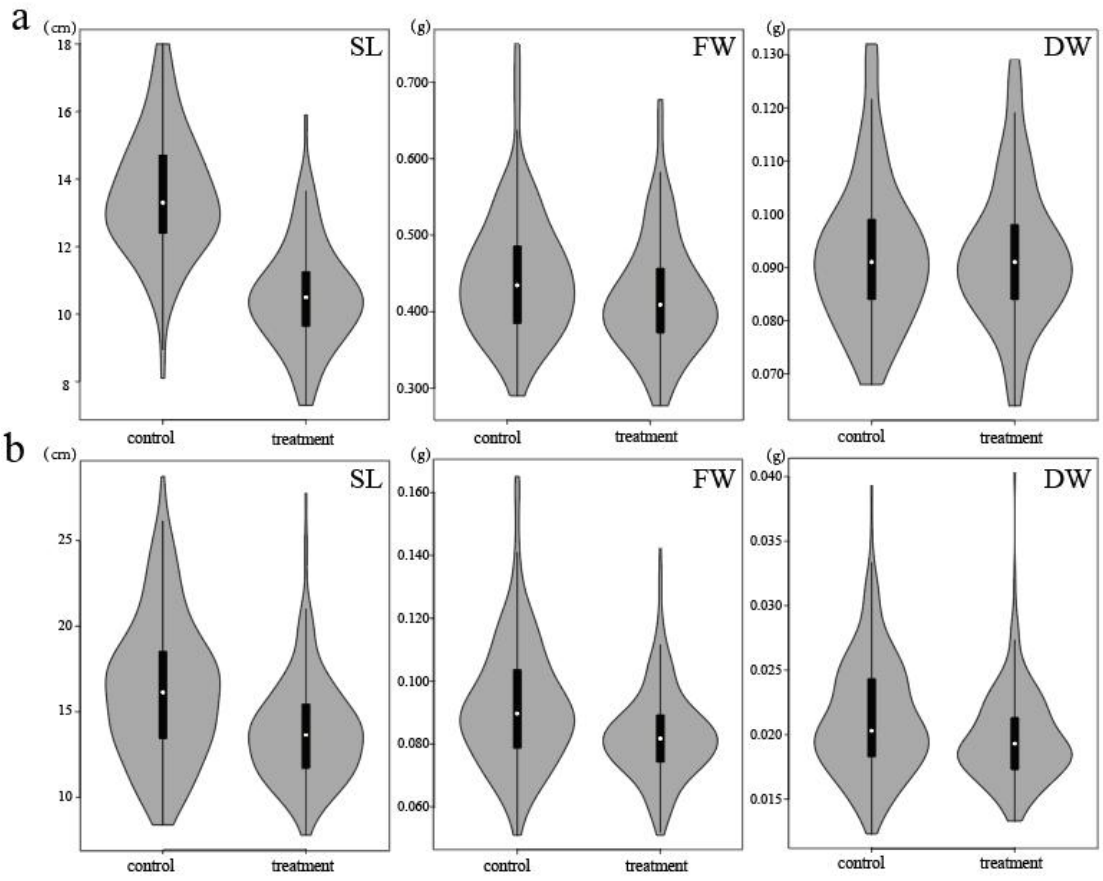


Fig. 1 The phenotypic traits of RIL and the natural population

a, Phenotypic traits of the RIL used for linkage analysis. b, Phenotypic traits of the natural rice population used for association analysis. CSL, CFW, and CDW indicate shoot length, fresh weight, and dry weight of the seedlings in control groups, respectively. SL, FW, DW indicate shoot length, fresh weight, and dry weight of the seedlings in treatment groups, respectively.

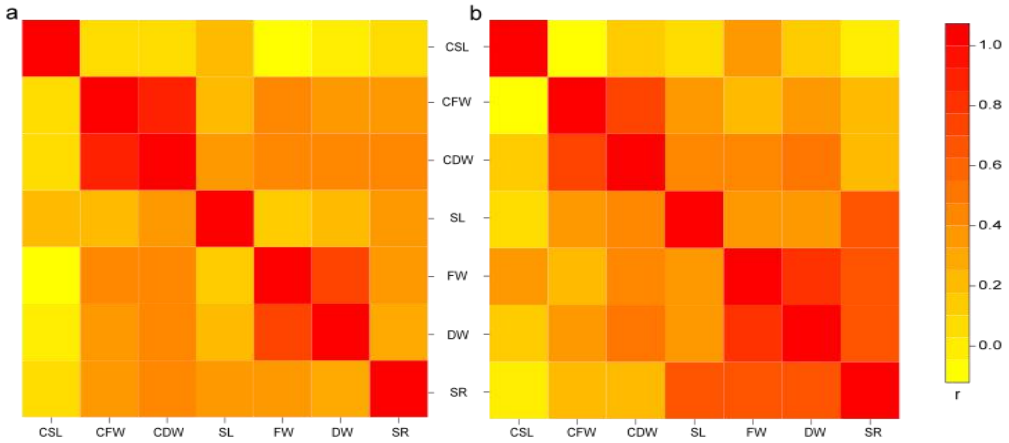


Fig. 2 Correlation analysis among biomass traits of RIL and the natural population.

a, Recombinant inbred lines. b, Natural rice population. The color indicates the correlation coefficient.

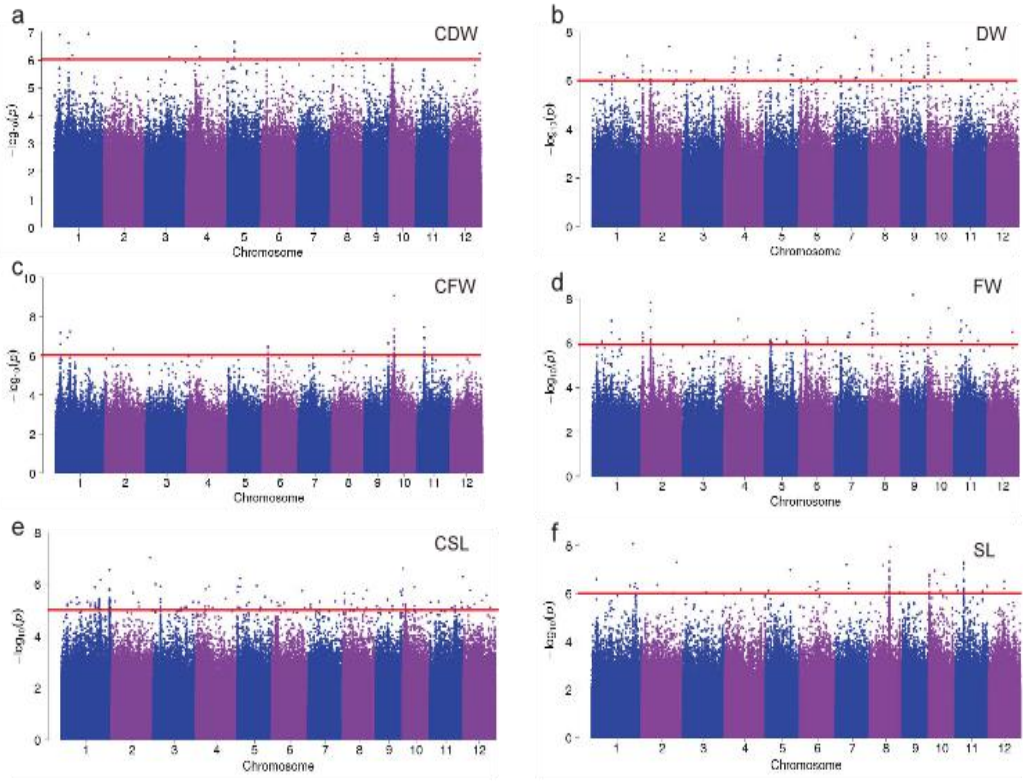
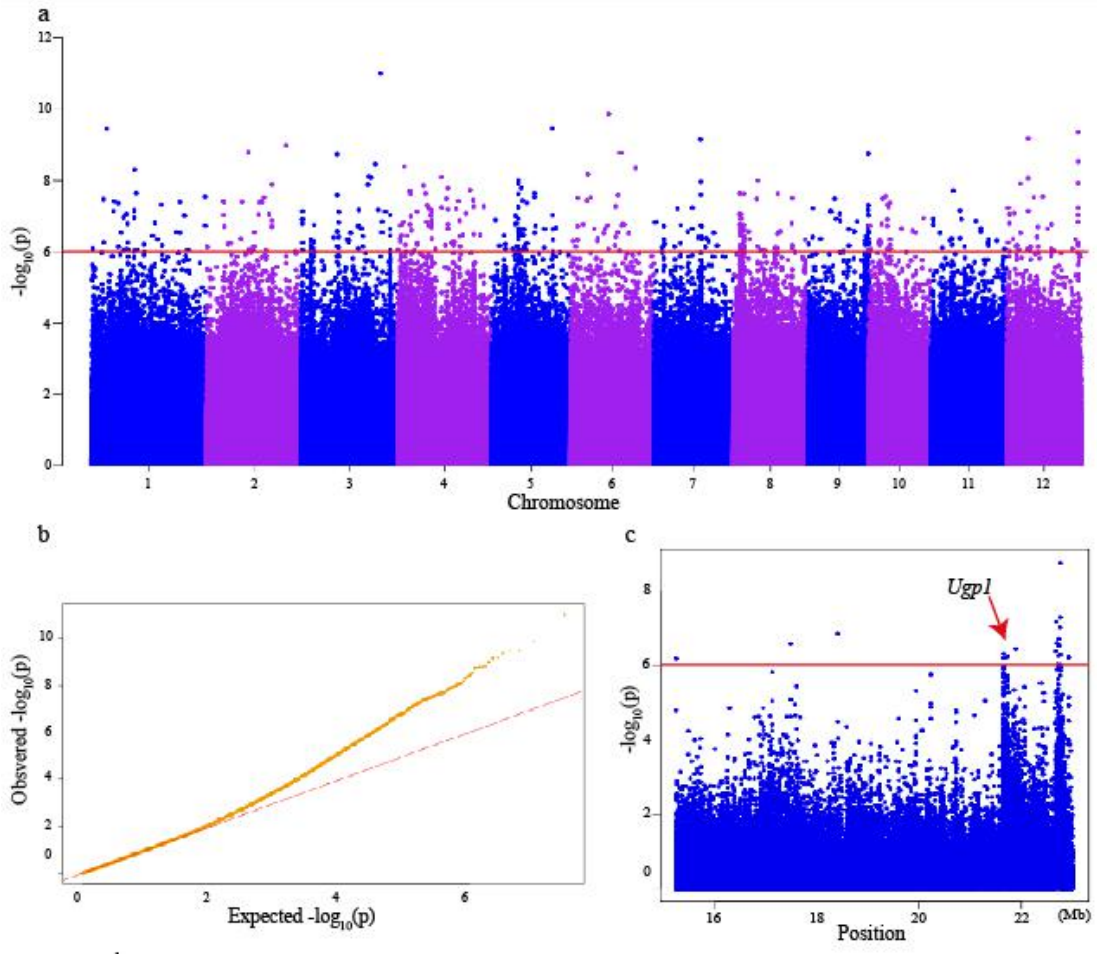


Fig. 3 Genome-wide association analysis of biomass traits under heat stress. Manhattan plots of a, Dry weight of control groups. b, Dry weight of treatment groups. c, Fresh weight of control groups. d, Fresh weight of treatment groups. e, Shoot length of control groups. f, Shoot length of treatment groups. CSL, CFW, and CDW indicate shoot length, fresh weight, and dry weight of the seedlings in control groups, respectively. SL, FW, and DW indicate the respective shoot length, fresh weight, and dry weight of the seedling in treatment groups, respectively.



d

Hap	21920756	21919424	21921215	21920513	21920592	21920470	21917982	21919625	21919315	21921937	Num.	SR	SL	GW	DW
Hap.1	C	C	C	C	C	T	A	G	G	A	191	0.94	13.78	0.08	0.02
Hap.2	C	C	C	C	C	C	A	G	G	A	17	0.84	15.45	0.09	0.02
Hap.3	T	T	T	T	T	T	C	A	T	G	10	0.91	13.05	0.09	0.02
Hap.4	C	C	C	C	T	T	A	G	G	A	6	0.92	12.63	0.08	0.02

Fig. 4 Correlation analysis of the survival rate of the natural rice population and analysis of known gene haplotypes.

a, Manhattan plot of survival rate (SR). b, QQplot of survival rate. c, Scatter plot of *Ugp1*. The arrow indicates that the site is the approximate site of *Ugp1*. d, Gene structures (left) and biomass traits of different haplotypes (right) of *Ugp1*. Red colored numbers indicate the key SNPs among major haplotypes.

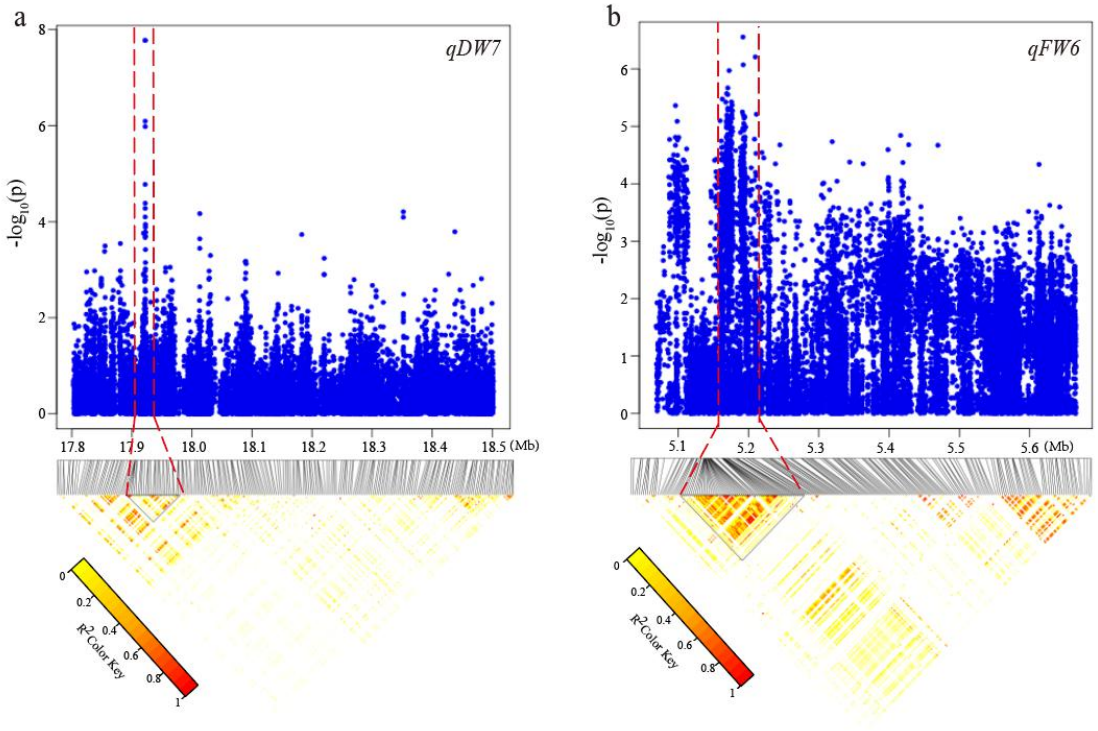
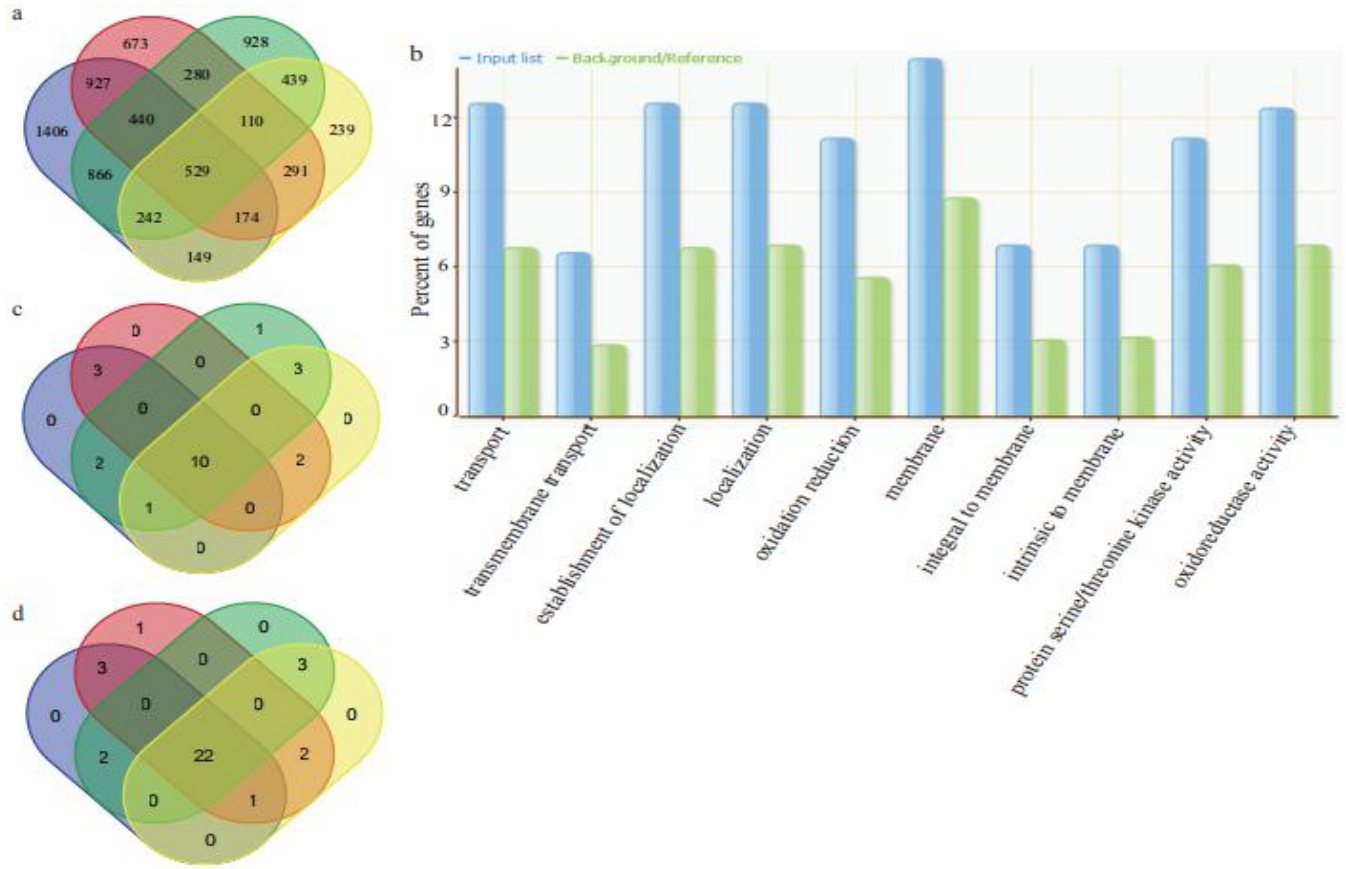


Fig. 5 Manhattan plot (top) and LD heatmap (bottom) surrounding the peak of the candidate interval.

Interval of linkage analysis and genome-wide association analysis in a, *qDW7*. b, *qFW6*. Dashed lines indicate the candidate region of the peak.



641

642 **Fig. 6 Genome-wide transcriptome analysis of heat stress between heat-tolerant and heat-sensitive rice varieties.**

643 a, Venn diagrams showing the differential genes in response to heat stress. b, GO enrichment analysis of differential genes.c, genes up-regulated

644 by redox-related pathways. d. Redox-related pathways down-regulate genes. HSR.1 and HSR.2 are varieties with extremely high survival rate;

645 LSR.1 and LSR.2 are varieties with extremely low survival rate.

Table 1. QTLs for various traits during seedling stage

Trait	QTL	Chromosome	LOD	Genetic Distance (cM)	Known Gene
CDW	*qCDW2.1*	2	2.88	98.2-101.3	
	qCDW2.2	2	3.65	105.8-107.8	
	qCDW2.3	2	3.30	142.5-145.3	
	qCDW11	11	4.01	8.1-9.4	
CFW	*qCFW6.1*	6	3.09	58.3-60.1	
	qCFW6.2	6	2.92	64.2-65.7	
	qCFW11	11	3.59	7.8-9.4	
CSL	*qCSL4.1*	4	2.57	74.4-78.3	
	qCSL4.2	4	3.81	84.2-85.8	
	qCSL9	9	5.15	27.3-29.1	
DW	*qDW3*	3	3.82	11.0-14.3	*GS8*[21]
	qDW7.1	7	3.68	28.0-29.0	
	qDW11	11	6.72	60.5-63.5	
FW	*qFW6*	6	2.77	19.1-22.0	*abl1*[22]
	qFW11.1	11	5.01	60.5-64.0	
	qFW11.2	11	2.53	66.1-70.4	
SL	*qSL12.1*	12	4.80	9.7-11.5	
	qSL12.2	12	2.82	21.9-25.9	
SR	*qSR3.1*	3	2.96	122.3-125.1	
	qSR3.2	9	3.54	2.0-7.5	

CSL, CFW, and CDW indicate shoot length, fresh weight, and dry weight of the seedlings in control groups, respectively. SL, FW, and DW indicate shoot length, fresh weight, and dry weight of the seedling in treatment groups, respectively. SR indicates the survival rate.