

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

3
4 Zhaoran Wei^{1,2*}, Qiaoling Yuan^{1*}, Hai Lin^{1*}, Xiaoxia Li¹, Chao Zhang¹, Hongsheng Gao¹, Bin
5 Zhang¹, Huiying He¹, Tianjiao Liu¹, Shandang Shi¹, Zhenyu Gao³, Lingrang Kong², Qian Qian^{1,3,#},
6 Lianguang Shang^{1,#}

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

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

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

14 * These authors contributed equally to this work.

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

17
18 **ABSTRACT**

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

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

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

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

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

38

39 **Background**

40 Rice (*Oryza sativa* L.) is a main food crop worldwide, especially in Asian countries.
41 As an Asian country, China has a population of 1.4 billion, and 2/3 of the people's
42 main food is rice. Fortunately, despite of the large population of eating rice as main
43 food in China, China is the world's largest rice producer, whose rice production ranks
44 first in the world^[1]. But, as the global industrialization process accelerates,
45 greenhouse gases, such as carbon dioxide (CO₂) and methane (CH₄), have caused the
46 global temperature to rise continually. The rising high temperature is affecting the rice
47 yield worldwide. Even if some rice varieties from tropical regions have formed
48 ecological characteristics adapted to high temperature environments, their yields will
49 also be affected once the temperature exceeds their appropriate temperature^[2]. The
50 ambient temperature exceeding the appropriate temperature during rice growth and
51 development would lead to yield reduction^[3]. Some high-temperature sensitive
52 cultivars may have a decrease in yield of more than 10% for every 1 °C increase^[4].

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

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

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

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

85

86 **Results**

87 **Phenotypes of RIL and GWAS population**

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

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

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

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

129

130 **QTL mapping of RIL**

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

149

150 **GWAS of heat-related traits in the seedling stage**

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

153 data from 3k database^[23]. It is generally considered that linkage disequilibrium was
154 between 100 kb and 200 kb in rice^[18]. In this study, the SNP threshold for the
155 significantly associated sites is P value $<10^{-6}$ and the significant SNPs sites within the
156 100 kb interval are considered candidate loci (Table S2). There were a large number
157 of different significant loci in the heat treatment and control groups, which suggested
158 that the high temperature in the seedling stage might affect the growth of the natural
159 population through these loci (Fig 3 and S3). According to the stringent screening
160 conditions, only a small number of reasonable SNPs were associated with the shoot
161 length, so the P value was reduced to $< 10^{-5}$ for subsequent analysis. In this way, a
162 total of 10 intervals for biomass traits were identified in the control groups, while 29
163 intervals for biomass traits were present on all of the chromosomes other than
164 chromosome 3 in the heat treatment group (Table S2). For the dry weight and fresh
165 weight of the treatment groups, the consistent intervals, *qDW_ind8* (3.09Mb-3.29Mb)
166 and *qFW_ind8* (3.09Mb-3.29Mb), were identified on chromosome 8. Therefore, there
167 are important candidate genes controlling each biomass traits under high temperature
168 stress in this interval. It can be inferred that the expression of one or more genes in
169 this interval was in response to high temperature stress by adjusting the dry weight
170 and fresh weight characters of rice seedlings.

171 There are 77 intervals containing 283 SNPs for survival rate after heat treatment (Fig.
172 4a, Table S2). Among them, *qSR_ind9-3* on chromosome 9 contains gene *Ugp1*, a
173 known heat-tolerant gene at the heading stage^[24](Fig. 4c). As *Ugp1* is expressed
174 during the whole growth and development period of rice^[24], it was suggested that this
175 gene could also had an effect on the rice seedling response to heat stress. To further
176 determine the natural variation of *Ugp1*, we performed haplotype analysis using all of
177 the non-synonymous SNPs in the coding region of this gene (Fig. 4d). A total of 10

178 SNPs in the coding sequence (CDS) region were detected, and four major haplotypes
179 were identified based on the 10 SNPs. Hap.1 is the main haplotype in nature
180 population, and the varieties belonging to Hap.1 had higher survival rate after high
181 temperature stress treatment in rice seedling stage. It can be found that the expression
182 of *Ugp1* does have an effect on the response of rice seedlings to high temperature
183 stress, and the varieties belonging to Hap.1 had higher survival rate after high
184 temperature stress treatment in rice seedling stage. It can be found that the expression
185 of *Ugp1* does affect the response of rice seedlings to high temperature stress. Among
186 255 materials, a total of 191 varieties belong to the Hap.1 haplotype, which proves
187 that Hap.1 has a higher utilization rate in the population. Compared with Hap.1, the
188 survival rate of Hap.2 after the change of chr9_21920470 was low, but the biomass
189 increased significantly. For Hap.3, all sites except chr9_21920470 changed. Although
190 the survival rate of Hap.3 is also lower than that of Hap.1, the decline rate of survival
191 rate is lower than that of Hap.2. Therefore, we judge that the chr9_21920470 locus
192 has a stronger effect on the survival rate of rice seedlings than other loci. In Hap.4,
193 where chr9_21920592 changed, compared with Hap.1, shoot length is lower. From
194 this, we judge that this site may be related to plant growth.

195

196 **Compare results from linkage analysis and GWAS**

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

203 (5.62 Mb-5.69 Mb). According to MSU V7.0^[25] annotated information, there are 13
204 candidate genes in this interval, including 12 genes involved in expression (Table S3).
205 The same analysis was performed on the interval of 200 kb (17.82 Mb-18.02 Mb) in
206 *qDW7* and *qDW_ind7* on chromosome 7, we reduced the candidate interval on
207 chromosome 7 to 57 kb (17.90 Mb-17.95 Mb), and identified 6 genes (Table S3). To
208 further investigate candidate genes, we performed homology analysis on the obtained
209 genes. The results of homology analysis showed that the homolog gene *LECRK-VII.2*
210 of *LOC_Os06g10790* is involved in the response to high temperature and drought in
211 *Arabidopsis*^[26]. The homologous gene *UGT73B3*^[27] of *LOC_Os06g10860* and the
212 homologous gene *UGT85A2*^[28] of *LOC_Os07g30330* are involved in the stress
213 response of *Arabidopsis*. The homologous gene *PHT3*^[29] of *LOC_Os06g10810* is
214 related to the redox reaction of *Arabidopsis thaliana*. These homologous genes on the
215 one hand confirm the accuracy of the experimental results, on the other hand, further
216 narrow the range of candidate genes in this study.

217

218 **Transcriptome analysis by RNA sequencing**

219 In order to study the transcriptome response of *indica* rice to high temperature in the
220 seedling stage, independent of the above genomic analysis, two heat-tolerant varieties
221 and two heat-sensitive varieties were selected for ribonucleic acid sequencing
222 (RNA_seq) analysis. After comparing the gene expression between the heat-resistant
223 groups and the heat-sensitive groups, it was determined that a total of 529 genes were
224 differentially expressed in the four groups (Fig. 6a), and it was considered that there
225 were key genes responding to high temperature stress in rice seedling. Kyoto
226 encyclopedia of genes and genomes (KEGG) analysis was performed on 529
227 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,

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

259

260 **Discussion**

261 **The combination of linkage, association analysis and transcriptome provides** 262 **effective information for target genes mining**

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

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

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

286 **Identification of candidate genes**

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

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

303

304 **Exploration of genes related to high temperature in rice seedling**

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

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

318 Experiments have shown that various biomass traits at the seedling stage can show
319 different responses in different varieties of seedlings to high temperature. The three
320 traits, SL, FW and DW, changed significantly in the RIL and natural rice population
321 after heat treatment (Fig. 1). Utilizing the different characteristics of the RIL
322 population and the natural population, linkage analysis and association analysis were
323 performed and obtained multiple candidate intervals. Researchers have used linkage
324 analysis or association analysis to identify heat-tolerant genes in rice
325 previously^[9-12,36-38]. In the study of other agronomic traits, it has also been confirmed
326 that the combined application of linkage and association analysis can obtain more
327 accurate candidate intervals^[18]. In this study, we narrowed the candidate intervals to
328 69 kb on chromosome 6 (5.62 Mb-5.69 Mb) and the interval 57 kb on chromosome 7
329 (17.90 Mb-17.95 Mb) based on the overlapping intervals and the rate of linkage
330 disequilibrium decay of SNPs in the linkage analysis and association analysis. 19
331 candidate genes were identified in the two fragments according to the gene annotation

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

334

335 **Identification of *Ugp1* favorable haplotypes**

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

347

348

349 **Conclusions**

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

357

358 **METHODS**

359 **Plant materials**

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

363

364 **Seedling culture and high-temperature stress treatment**

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

375

376 **Statistical analysis of data**

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

379

380 **Linkage analysis**

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

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

389

390 **Association analysis**

391 Association analysis of genotypes and phenotypes was performed with the efficient
392 mixed-model association eXpedited (EMMAX)^[42]. The genotype data were referenced
393 with sequencing information released for the 3k rice database^[23]. SNP site settings
394 with a deletion rate greater than 40% and a minimum allele frequency (MAF) of less
395 than 5% were filtered out. When the site had a significant level, $P < 10^{-5}$, it was
396 considered to be associated with the trait and was used for subsequent analysis. The
397 relevant drawings in the subsequent analysis were completed by the corresponding
398 drawing package of R x64 3.4.2. The method for determining and naming GWAS
399 candidate intervals refers to the method proposed by Zhao^[43].

400

401 **Transcriptome analysis**

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

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

417

418 **Supplementary information**

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

421 Table S2 Candidate QTLs identified by GWAS.

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

424 Table S4 KEGG pathway analysis of DEGs.

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

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

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

431

432 **Abbreviations**

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

443

444 **Authors' contributions**

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

450

451

452 **Funding**

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

459 (Grant No. 2020M672903).

460

461 **Availability of data and materials**

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

464

465 **Ethics approval and consent to participate**

466 Not applicable.

467

468 **Consent for publication**

469 Not applicable.

470

471 **Competing interests**

472 The authors declare that they have no competing interests.

473

474 **REFERENCES**

475 1. Lei Z, Jianguo L, Yingjin H, et al. Mapping quantitative trait loci for heat tolerance
476 at anthesis in rice using chromosomal segment substitution lines[J]. *Breeding*
477 *Science*, 2016, 66 (3):358-366.

478 2. Driedonks N, Rieu I, Vriezen W H. Breeding for plant heat tolerance at vegetative
479 and reproductive stages[J]. *Plant Reproduction*, 2016, 29 (1-2):67-79.

480 3. Akman Z. Comparison of high temperature tolerance in maize, rice and sorghum
481 seeds by plant growth regulators[J]. *Journal of Animal and Veterinary Advances*,
482 2009, 8 (2):358-361.

483 4. Peng S, Huang J, Sheehy J E, et al. Rice yields decline with higher night
484 temperature from global warming[J]. *Proceedings of the National Academy of*

- 485 Sciences, 2004, 101 (27):9971-9975.
- 486 5. Wei H, Liu J, Wang Y, et al. A Dominant Major Locus in Chromosome 9 of Rice
487 (*Oryza sativa* L.) Confers Tolerance to 48 °C High Temperature at Seedling
488 Stage[J]. *Journal of Heredity*, 2013, 104 (2):287-294.
- 489 6. Beáta Barnabás, Katalin Jäger, Attila Fehér. The effect of drought and heat stress on
490 reproductive processes in cereals[J]. *Plant Cell and Environment*, 2008, 31
491 (1):11-38.
- 492 7. Han F, Chen H, Li X J, et al. A comparative proteomic analysis of rice seedlings
493 under various high-temperature stresses[J]. *Biochimica et Biophysica Acta*, 2009,
494 1794 (11):1625-1634.
- 495 8. Wahid A, Close T J. Expression of dehydrins under heat stress and their
496 relationship with water relations of sugarcane leaves[J]. *Biologia Plantarum*
497 (Prague), 2007, 51 (1):104-109.
- 498 9. Jagadish S V K, Cairns J, Lafitte R, et al. Genetic Analysis of Heat Tolerance at
499 Anthesis in Rice[J]. *Crop Science*, 2010, 50 (5):1633.
- 500 10. Lei D, Tan L, Liu F, et al. Identification of heat-sensitive QTL derived from
501 common wild rice (*Oryza rufipogon* Griff.) [J]. *Plant Science*, 2013,
502 201-202:121-127.
- 503 11. Li X M, Chao D Y, Wu Y, et al. Natural alleles of a proteasome α 2 subunit gene
504 contribute to thermotolerance and adaptation of African rice[J]. *Nature Genetics*,
505 2015, 47 (7):827-833.
- 506 12. Salvi S, Tuberosa R. To clone or not to clone plant QTLs: present and future
507 challenges[J]. *Trends in Plant Science*, 2005, 10 (6):0-304.
- 508 13. Wang C, Chen S, Yu S. Functional markers developed from multiple loci in GS3
509 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

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

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

539 24. Chen R, Zhao X, Shao Z, et al. Rice UDP-Glucose Pyrophosphorylase1 Is
540 Essential for Pollen Callose Deposition and Its Cosuppression Results in a New
541 Type of Thermosensitive Genic Male Sterility[J]. *THE PLANT CELL ONLINE*,
542 2007, 19 (3):847-861.

543 25. Kawahara Y, Bastide M D L, Hamilton J P, et al. Improvement of the *Oryza*
544 *sativa* Nipponbare reference genome using next generation sequence and optical
545 map data[J]. *Rice*, 2013, 6 (1):4.

546 26. Rizhsky, L. When Defense Pathways Collide. The Response of Arabidopsis to a
547 Combination of Drought and Heat Stress[J]. *PLANT PHYSIOLOGY*, 2004,
548 134(4):1683-1696.

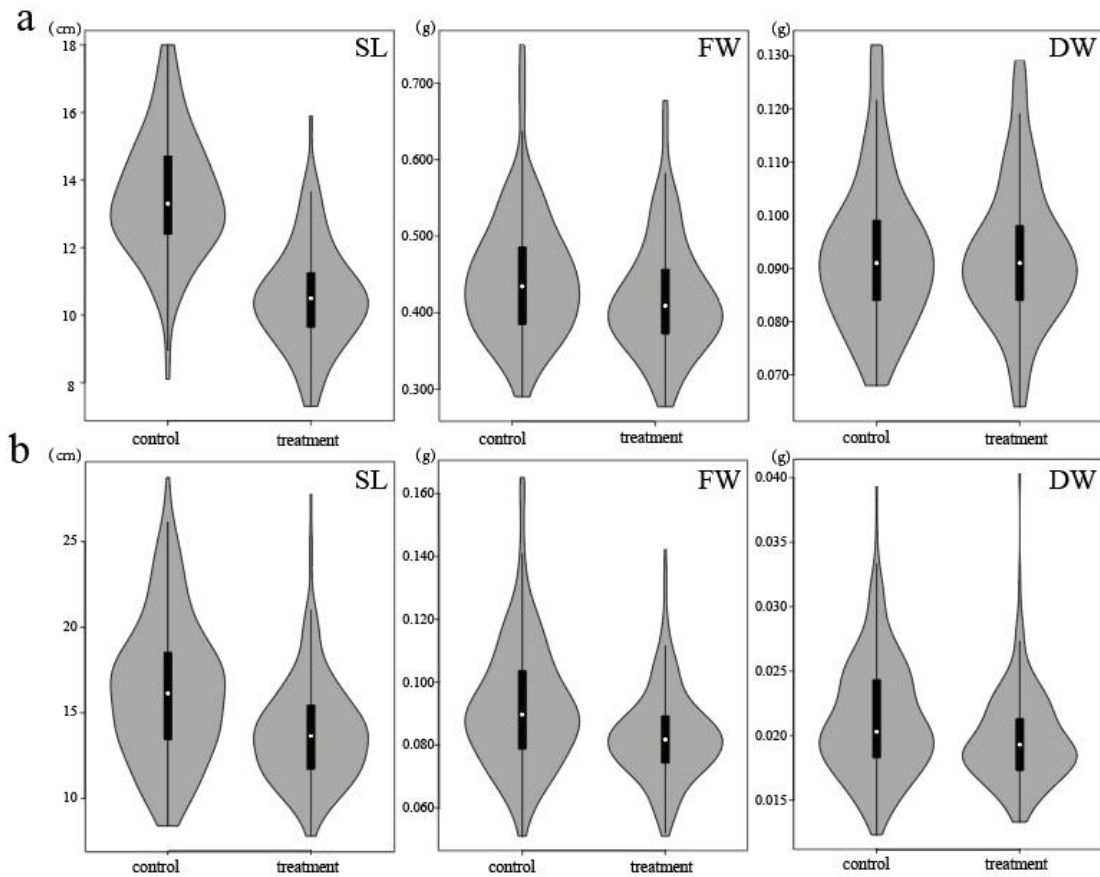
549 27. Lee T A, Bailey-Serres J. Integrative analysis from the epigenome to transcriptome
550 uncovers patterns of dominant nuclear regulation during transient stress[J]. *The*
551 *Plant Cell*, 2019, 31(11): 2573-2595.

552 28. Rehman H M, Nawaz M A, Shah Z H, et al. Comparative genomic and
553 transcriptomic analyses of Family-1 UDP glycosyltransferase in three Brassica
554 species and Arabidopsis indicates stress-responsive regulation[J]. *Scientific reports*,
555 2018, 8(1): 1-18.

556 29. Jia F, Wan X, Zhu W, et al. Overexpression of mitochondrial phosphate transporter
557 3 severely hampers plant development through regulating mitochondrial function in
558 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.

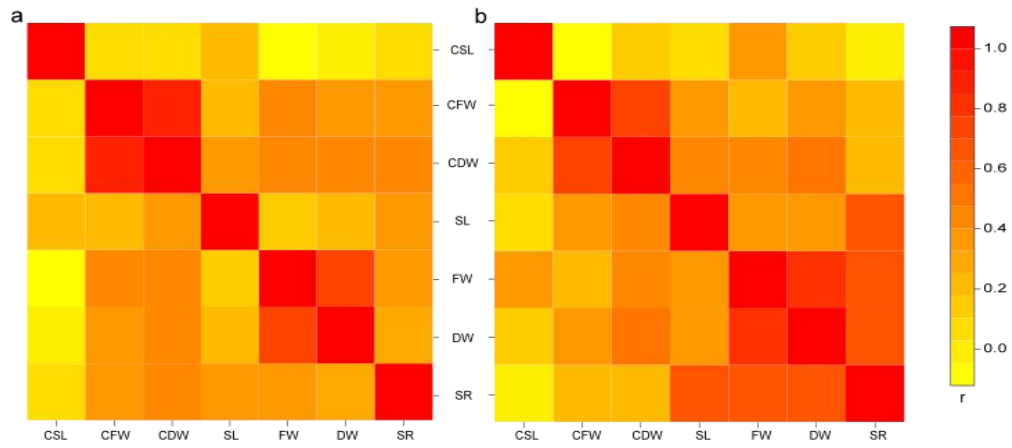
- 582 38. Ye C, Tenorio F A, Argayoso M A, et al. Identifying and confirming quantitative
583 trait loci associated with heat tolerance at flowering stage in different rice
584 populations[J]. BMC Genetics, 2015, 16 (1):41.
- 585 39. Yoshida S, Forno D A, Cock J H. Laboratory manual for physiological 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.



602
 603
 604
 605
 606
 607
 608

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

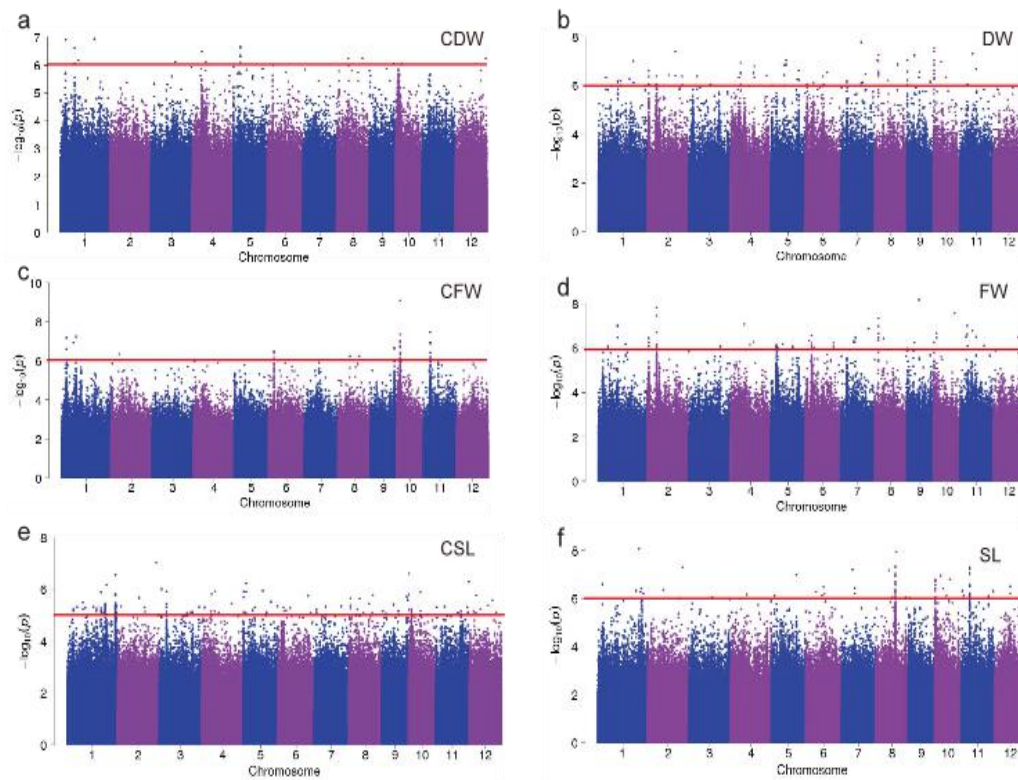
a, Phenotypic traits of the RIL used for linkage analysis. b, Phenotypic traits of the natural rice population used for association analysis. CSL, CFW, and CDW indicate shoot length, fresh weight, and dry weight of the seedlings in control groups, respectively. SL, FW, DW indicate shoot length, fresh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



609

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

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



615

616 **Fig. 3 Genome-wide association analysis of biomass traits under heat stress.**

617 Manhattan plots of a, Dry weight of control groups. b, Dry weight of treatment groups.

618 c, Fresh weight of control groups. d, Fresh weight of treatment groups. e, Shoot length

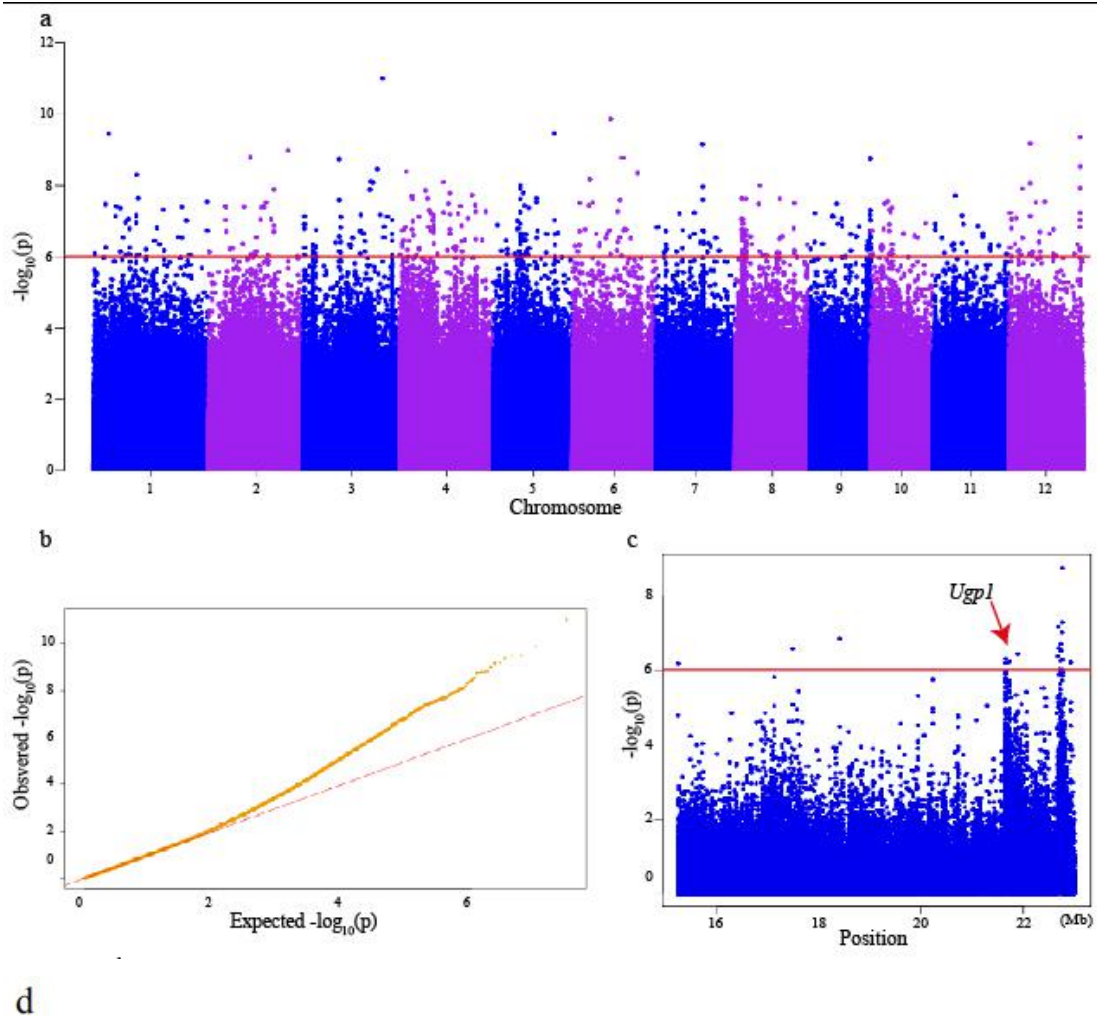
619 of control groups. f, Shoot length of treatment groups. CSL, CFW, and CDW indicate

620 shoot length, fresh weight, and dry weight of the seedlings in control groups,

621 respectively. SL, FW, and DW indicate the respective shoot length, fresh weight, and

622 dry weight of the seedling in treatment groups, respectively.

623



624

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

625

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

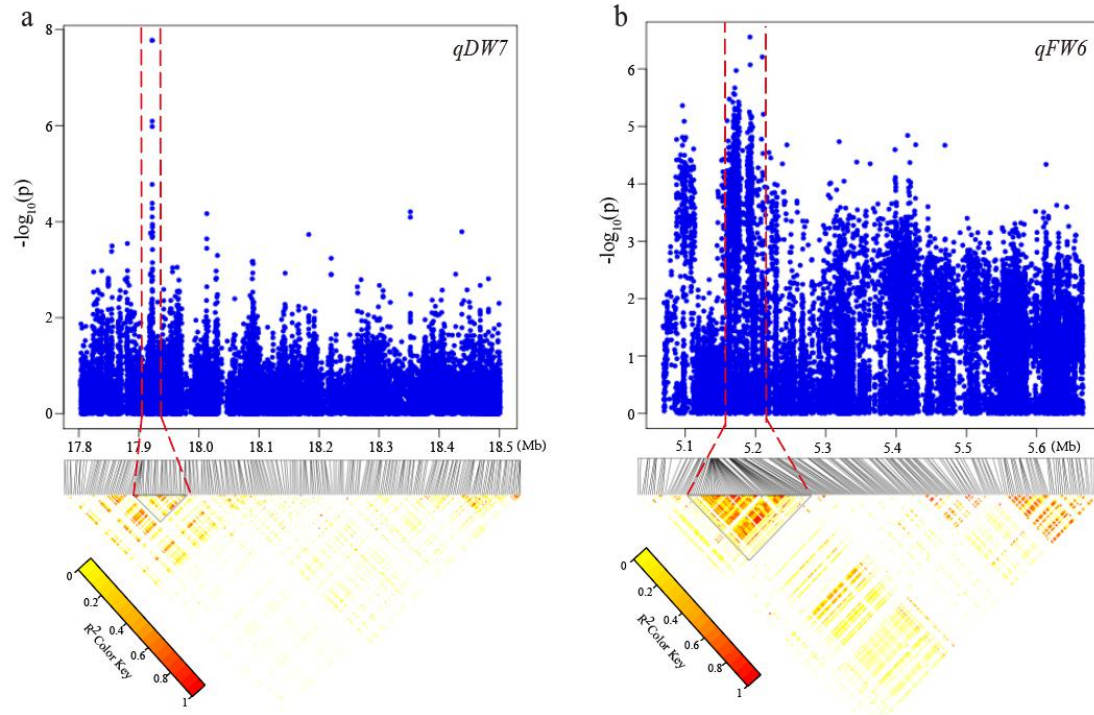
628 a, Manhattan plot of survival rate (SR). b, QQplot of survival rate. c, Scatter plot of
 629 *Ugp1*. The arrow indicates that the site is the approximate site of *Ugp1*. d, Gene
 630 structures (left) and biomass traits of different haplotypes (right) of *Ugp1*. Red
 631 colored numbers indicate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

632

633

634

635

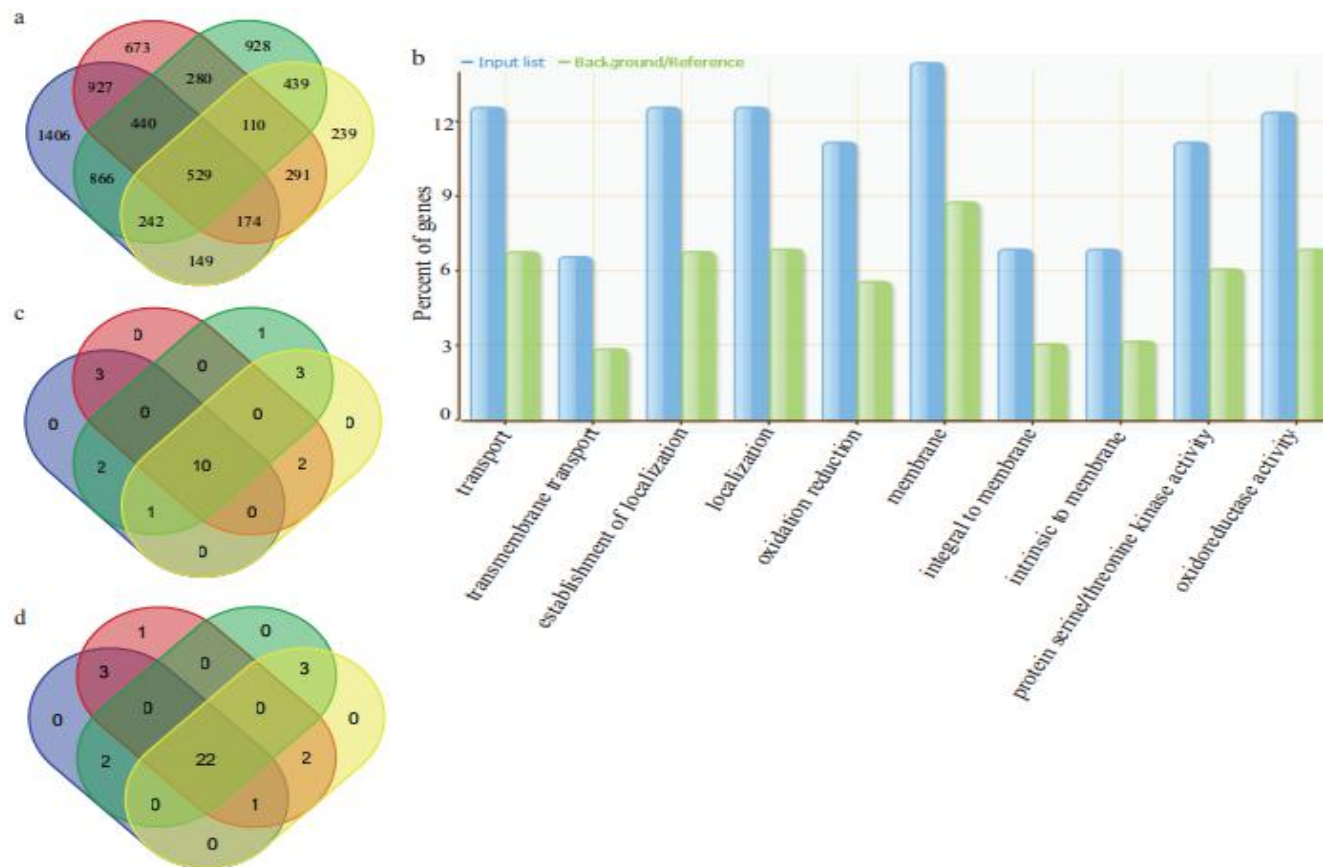


636

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

639 Interval of linkage analysis and genome-wide association analysis in a, *qDW7*. b,

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

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