

**Directional selection at gene expression level contributes to the speciation of  
Asian rice cultivars**

1 Lihong Xie<sup>1#</sup>, Kehan Yu<sup>1#</sup>, Dongjing Chen<sup>1\*</sup>

2 1. *AmazingX Academy, Baoli Zhongchen Tower, Foshan 528300, Guangdong,*

3 *People's Republic of China*

4 <sup>#</sup> these authors contribute equally to this work.

5 \* corresponding author: [sysucdj@amazingx.org](mailto:sysucdj@amazingx.org)

6

7

8

## 9 **Abstract**

10 Differences in expression levels play important roles in phenotypic variation across  
11 species, especially those closely related species with limited genomic differences.  
12 Therefore, studying gene evolution at expression level is important for illustrating  
13 phenotypic differentiation between species, such as the two Asian rice cultivars,  
14 *Oryza sativa* L. ssp. *indica* and *Oryza sativa* L. ssp. *japonica*. In this study, we  
15 evaluated the gene expression variation at inter-subspecies and intra-subspecies level  
16 using transcriptome data from seedlings of three *indica* and *japonica* rice and defined  
17 four groups of genes under different natural selections. We found a substantial of  
18 genes (about 79%) that are under stabilizing selection at the expression level in both  
19 subspecies, while about 16% of genes are under directional selection. Genes under  
20 directional selection have higher expression level and lower expression variation than  
21 those under stabilizing selection, which suggest a potential explanation to subspecies  
22 adaptation to different environments and interspecific phenotypic differences.  
23 Subsequent functional enrichment analysis of genes under directional selection shows  
24 that *indica* rice have experienced the adaptation to environmental stresses, and also  
25 show differences in biosynthesis and metabolism pathways. Our study provide an  
26 avenue of investigating *indica-japonica* differentiation through gene expression  
27 variation, which may guide to rice breeding and yield improvement.

28

29 Key words: *japonica* rice; *indica* rice; gene expression evolution; speciation,  
30 adaptation

31

## 32 **Introduction**

33 Rice grown in Asia (*Oryza sativa* L.) is a very important food crop for China and the  
34 world (Ruan Bosheng, 2008), and more than half of the world's population uses it as a  
35 staple food (Sasaki & Burr, 2000). It is mainly divided into two different subspecies,  
36 namely the subspecies *Oryza sativa* L. ssp. *indica* (*indica*) and the subspecies *Oryza*  
37 *sativa* L. ssp. *japonica* (*japonica*). It is speculated that *japonica* and *indica* rice were

38 domesticated 9,000 years ago (Purugganan & Fuller, 2009), but debate about their  
39 origins still exists. The first model, the single-source model, indicated that the two  
40 main subspecies of Asian rice, *indica* and *japonica*, were domesticated from wild rice  
41 (*O. rufipogon.*) and then differentiated (Ting, Y., 1957) (Lu et al. al., 2002; Wang et al.,  
42 2008). In contrast, the second model, the multiple independent domestication model,  
43 proposes that the two main rice types are domesticated separately (Oka & Morishima,  
44 1982). *Indica* rice evolved from wild rice. Because, *japonica* rice is the continuous  
45 evolution and artificial selection of *indica* rice in the process of people's continuous  
46 introduction to high latitudes and high altitude areas (Lu Baorong et al., 2009).  
47 Therefore, *indica* rice is the basic type, and *japonica* rice is the variant type (Molina  
48 et al., 2011). The latter has been supported by many researchers after the observation  
49 of the strong genetic differentiation between *indica* and *japonica* and the development  
50 of several systems of rice domestication (Wang, X. et al., 1984). The third origin  
51 model claims that *indica* and *japonica* were independently domesticated or at least  
52 domesticated twice and then differentiated from wild rice (*O. rufipogon.*) (Garris et al.,  
53 2005; Second, 1982).

54

55 In most cases, *indica* rice ecotypes are mainly distributed in low-latitude or  
56 high-altitude tropical and subtropical rice growing areas, while *japonica* rice ecotypes  
57 are mainly distributed in high-latitude temperate areas (Lu et al., 2009). Due to  
58 long-term adaptation to different ecological environments, *japonica* rice and *indica*  
59 rice have differences in morphological characteristics, agronomic traits, and genes. In  
60 terms of morphological characteristics, the stems of *indica* rice are thicker, and the  
61 plant height is generally more than 1 meter. The tillering ability is stronger, the leaf  
62 color is lighter, the grains are slender, easy to fall, and the rice yield rate is low.  
63 However, *japonica* rice generally has a thin stem and a plant height of 75-95 cm.  
64 Traditional *japonica* rice varieties have a lower tillering ability than *indica* rice. The  
65 leaves are darker, the grains are short and round, and they are not easy to shatter, and  
66 the rice yield rate is higher. (*Flower Encyclopedia*; Xu Zhengjin et al., 2003). In terms  
67 of agronomic characteristics, *indica* rice with a short growth period is more resistant

68 to humidity, heat, and strong light, but not cold tolerant. After being hulled into *indica*  
69 rice, the transparency of the rice grains is low. Because it contains about 20% amylose,  
70 *indica* rice is drier and looser when cooked. However, *japonica* rice, which has a long  
71 growth period, generally only matures once a year, is more cold-tolerant and tolerant  
72 to low light, but not tolerant to high temperatures. After being hulled into *japonica*  
73 rice, the rice grains have high transparency(Douding.com). Because it contains less  
74 amylose, less than 15%, it is medium-viscosity (Report on Rice Factory Seedling  
75 Breeding-Douding.com), and its cooked food characteristics are between glutinous  
76 rice and *indica* rice (Jiang Jian et al., 2001; Miao Xiangwei & Wang Dexin, 2009; Xu  
77 Hai et al., 2007).

78

79 We want to study how the advantages of gene expression evolution between *japonica*  
80 rice and *indica* rice are reflected in the phenotype. *Japonica* rice and *indica* rice are  
81 two subspecies with similar evolutionary distances and similar genomes. Because the  
82 differences between species caused by DNA sequences are relatively small, the  
83 differences caused by gene expression regulation are very important for the  
84 interpretation of phenotypic differences.

85

86 Previous studies on phenotypic differences in species focused on the evolution of  
87 gene lineage to study species differentiation (Onishi et al., 2007; Ting et al., 2000),  
88 but a small number of study focused on the study of transcriptomes. The  
89 transcriptome is a collection of all transcription products in a cell under a specific  
90 developmental stage or physiological condition of a species. Understanding the  
91 transcriptome is of great significance for explaining the functional elements of the  
92 genome and the phenotypic traits controlled by its line (Wang et al., 2009). Therefore,  
93 we want to study the influence of the transcriptome on the differences in gene  
94 expression and the changes in traits. This will provide great research significance for  
95 species differentiation.

96

97 In order to measure the extent and specific traits of the differences between Asian

98 cultivated rice subspecies caused by gene expression levels, we divided genes into  
99 four categories according to the differences in gene expression between subspecies  
100 and within subspecies: (1) genes subjected to directional selection: genes with large  
101 differences in expression between species and small differences in expression within  
102 species are the genes that continue to strengthen the selection pressure for a certain  
103 trait among populations. These genes are also the main genes that cause differences in  
104 traits between species (Mitchell-Olds et al., 2007; Rieseberg et al., 2002). (2) Genes  
105 subjected to stabilizing selection: genes with small differences between species and  
106 within species are the continuous selection pressure for interspecies traits in the  
107 population. This natural selection will not lead to population differentiation  
108 (Charlesworth et al., 1982). (3) Genetic drift: genes with large differences between  
109 species and within species are a type of gene affected by random factors in the  
110 population (Melo & Marroig, 2015; Nielsen, 2005). (4) Complex scenario: genes with  
111 small inter-species differences and large intra-species differences. After the  
112 identification of natural selection, we analyze for gene enrichment of the genes  
113 subject to directional selection and understand its specific traits and functions  
114 (Subramanian et al., 2005).

115

116 This article studies the following scientific issues. First, whether the genes subject to  
117 directional selection in expression are the genes that cause the phenotypic differences  
118 between *japonica* rice and *indica* rice. Second, whether the functions of these genes  
119 are related to phenotypic differences.

120

121 The purpose and significance of this research are mainly to increase yield and  
122 breeding. It is to study that this species is closely related to the food production of all  
123 human beings, and to analyze the reasons for the differences in rice phenotypes so  
124 that people can better control the traits to achieve the ideal yield and quality of rice.

125

126

127 **Materials and methods**

128 **The acquisition of rice transcriptome data and the measurement of gene**  
129 **expression levels**

130

131 We obtained the seedling transcriptome data of three different *indica* and *japonica*  
132 rice lines (*Indica*: Khao Dawk Mali 105, Guangluai 4, and Rathuwee; *Japonica*:  
133 Taipei 309, Heukgyeong, and Nipponbare) from Wen et al. (Wen et al., 2016).  
134 Subsequently, we used hisat2 (version: 2.2.0, default parameters) to map the read  
135 length of the original sequencing data of all samples back to the *japonica* rice  
136 reference genome (ensembl, IRGSP-1.0), and used stringtie (version: 2.1.6;  
137 parameters: -e, Other defaults) to calculate the TPM (transcript per million) of the  
138 protein-coding gene (ensembl, IRGSP-1.0.51). TPM was used as the expression level  
139 of the protein-coding gene for subsequent analysis.

140

141

142 **Principal component analysis**

143 After obtaining gene expression levels, we use principal components to analyze  
144 the differences between samples. First, the gene-specific TPM uses the R method  
145 vegdist to calculate the dissimilarity index between samples, then uses the R method  
146 prcomp to calculate the principal component of the sample for the index, and finally  
147 uses the R method ggbiplot to visualize the data. (Figure 1, grouping, principle  
148 component separates different samples)

149

150

151 **Identification of genes under four types of natural selection**

152 We adopted the method of Yeh et al. According to whether there are significant  
153 differences in gene expression between species and within species, we divided them  
154 into four categories: directional selection: significant differences between species and  
155 insignificant differences within species; stabilizing selection: both difference between  
156 interspecies and difference between intraspecies are not significant; genetic drift:  
157 interspecies and intraspecies differences are both significant; complex scenario:

158 interspecies differences are not significant, and intraspecies differences are significant  
159 (Yeh et al., 2014). We use a linear model to measure whether there are significant  
160 differences in gene expression between species and within species.

161 For interspecies:

$$y_i = \mu + S_i + \varepsilon_i$$

162

163 Among them,  $y_i$  represents the expression level of the protein coding gene of  
164 species  $i$ ,  $\mu$  is the basic expression level,  $S_i$  is the effect of species  $i$ , and  $\varepsilon_i$  is the  
165 residual. If the P value of the linear model is less than 0.05, it is defined as significant  
166 difference between species.

167 For intraspecies:

$$y_j = \mu + L_j + \varepsilon_j$$

168

169 Among them,  $y_j$  represents the expression level of the protein coding gene of  
170 line  $j$ ,  $\mu$  is the basic expression level,  $L_j$  is the effect of line  $j$ ,  $\varepsilon_j$  and is the residual.  
171 If the P value of the linear model is less than 0.05, it is defined as significant  
172 intraspecies difference.

173

174

### 175 **Functional enrichment analysis**

176 In order to find genes and functions related to subspecies differentiation, we used  
177 RiceNETDB (<http://bis.zju.edu.cn/ricenetdb/>) to carry out Gene Ontology (Gene  
178 Ontology) on *indica* and *japonica* species-specific directional selection genes. , GO  
179 enrichment analysis. The three functional branches of GO, the biological process,  
180 molecular function and cellular component, are all used for functional enrichment  
181 analysis.

182

183

184

185 **Results**

186 **Processing and expression analysis of Asian cultivated rice transcriptome data**

187 We obtained the seedling transcriptome data of three *indica* rice lines (Khao Dawk  
 188 Mali 105, Guangluai 4 and Rathuwee) and three *japonica* rice lines (Taipei 309,  
 189 Heukgyeong and Nipponbare) from the research of Wen et al. (Wen et al.,  
 190 2016)(Table 1). After hisat2 (version 2.2.0), the initial read length was posted to the  
 191 reference genome (ensembl, IRGSP-1.0). Among the samples with a total read length  
 192 of 18,818,192 to 20,326,850, it was found that the only response rate of the  
 193 transcriptome sequencing read length was between 63.41 and 68.40%. For the 35,775  
 194 protein-coding genes that have been annotated, the number of genes detected to be  
 195 expressed (TPM greater than 0) ranged from 28,021 to 29,092 (78.33 to 81.32%);  
 196 while higher expression levels were detected (TPM greater than or equal to 5), the  
 197 number of genes is between 16,942-17,807 (47.36-49.77%). It shows that in our  
 198 subsequent analysis, about 50% of the genes can be judged to be credible and selected  
 199 because of their higher expression levels.

200

201 **Table 1.** *Indica* and *Japonica* transcriptome statistics

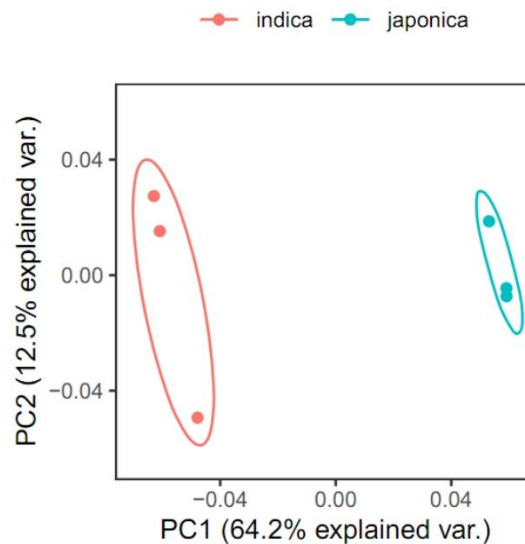
Subspecies	SRR id	Accession name	Total reads	Uniquely mapped	# of detected PCGs (TPM>0)	# of detected PCGs (TPM≥5)
<i>indica</i>	SRR2154082	Guangluai4	20,326,850	13,336,411 (65.61%)	28,021 (78.33%)	16,942 (47.36%)
	SRR2154083	KDM105	20,216,686	12,820,328 (63.41%)	28,316 (79.15%)	17,367 (48.55%)
	SRR2154084	Rathuwee	19,876,912	12,716,807 (63.98%)	28,240 (78.94%)	17,428 (48.72%)
<i>japonica</i>	SRR2154085	Heukgyeong	19,889,198	13,584,731 (68.30%)	28,766 (80.41%)	17,312 (48.39%)
	SRR2154086	Nipponbare	18,818,192	12,871,620 (68.40%)	29,300 (81.90%)	17,807 (49.77%)
	SRR2154087	Taipei309	19,094,551	13,013,267 (68.15%)	29,092 (81.32%)	17,475 (48.85%)

202



203

204 Subsequently, we conducted a principal component analysis of six lines from two  
205 cultivated rice subspecies, and found that the difference between the samples of *indica*  
206 and *japonica* is relatively large (the first principal component PC1 can  
207 explain 64.2% of the variation between the samples), while the subspecies differences  
208 in the internal samples are small (the second principal component PC2 only  
209 explains 12.5% of the variation between samples ) (Figure 1). It shows that the data is  
210 sufficient to define the evolution of expression based on inter-species and  
211 intra-species differences (see Materials and Methods).



212

213 **Figure 1.** Principal component analysis of the transcriptome of *indica* and *japonica*  
214 rice samples.

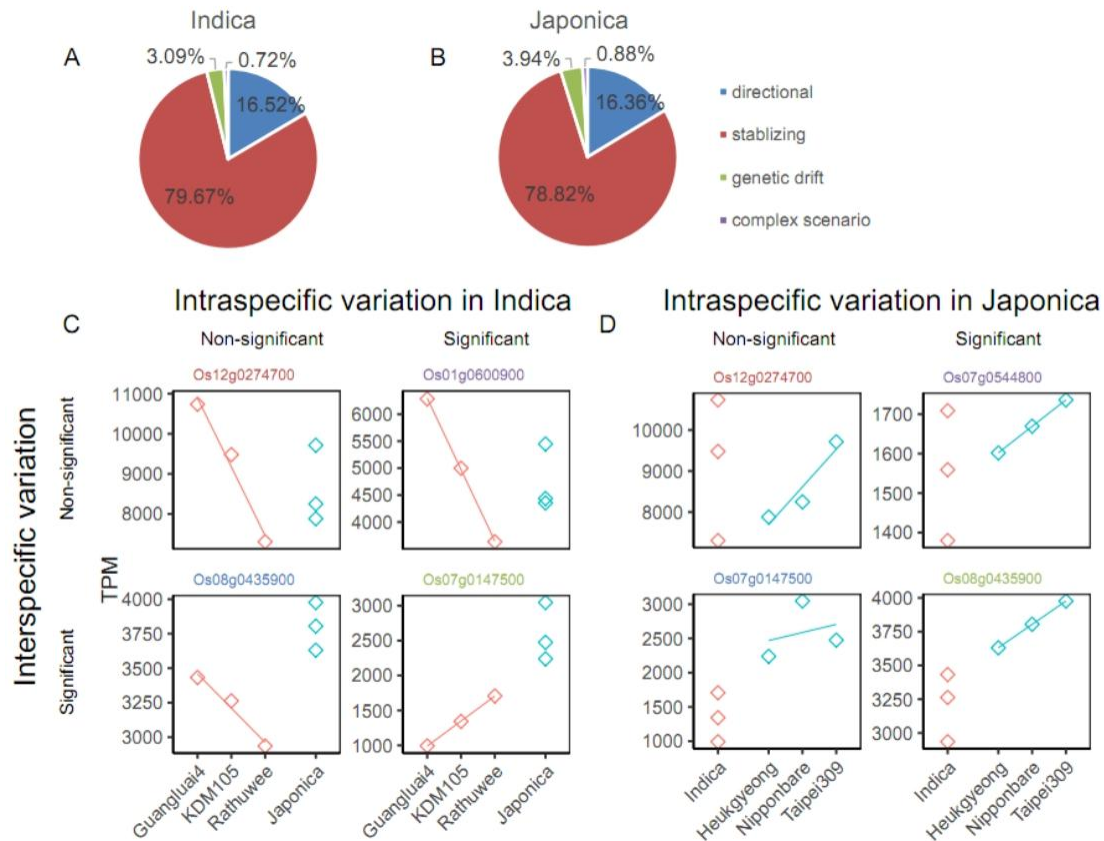
215

216

217 **The expression of most genes in *indica* and *japonica* rice is affected by stabilizing**  
218 **selection**

219 *Indica* rice and *japonica* rice are subspecies of cultivated rice in Asia.  
220 They diverged about 0.55 million years ago (Stein et al. 2018, nature genetics) . There  
221 is only 1/6 difference in genome , and most of them are located on transposons  
222 (Campbell et al., 2020; Gao et al., 2015; Ma & Bennetzen, 2004). Therefore, in  
223 addition to the genomic differences between *indica* and *japonica* rice, differences in

224 gene expression levels are also an important factor explaining the differences in traits  
225 between this subspecies (Rieseberg et al., 2002). In order to find out which genes  
226 cause differences between subspecies, we divided genes into four categories based on  
227 differences in gene expression between and within subspecies: genes subject  
228 to directional selection, genes subject to stabilizing selection gene, genetic drift, and  
229 complex scenario. Our results show that among the 35,775 genes in *indica* and  
230 *japonica*, most of the genes are subject to stabilizing selection  
231 (*indica*: 28,501, accounting for 79.67%; *japonica*: 28,198 , accounting for 78.82%),  
232 which is similar to the two subspecies. The genomic differences are relatively small  
233 and consistent (Figure 2A,B). For example, the expression difference of  
234 the gene Os12g0274700 between and within the two subspecies is relatively small and  
235 not significant (Figure 2C , D ). However, there are still quite a few genes subject to  
236 targeted selection, 16.25% (5,910/35,775) in *indica* rice and 16.36% ( 5,854/35,775)  
237 in *japonica* rice (Figure 1A, B ). For example, in *indica* rice, the gene Os08g0435900 ,  
238 which is subject to targeted selection, has a large difference in expression between  
239 species, while the difference between *indica* species is not significant  
240 (Figure 2C). This gene is chlorophyll ab binding protein P4 (chlorophyll ab binding  
241 protein P4), which is located in the chloroplast, and is highly expressed in flag leaves  
242 before flowering (Wang et al., 2015) . The gene Os07g0147500 , which is subject to  
243 directional selection in *japonica* rice, has a larger expression difference between  
244 *japonica* rice and *indica* rice, but the expression difference in *japonica* rice is not  
245 significant (Figure 2D). This gene is a 10kDa polypeptide located in the  
246 photosystem II of the chloroplast , and is highly expressed in flower buds, leaves after  
247 anthesis, and grain-filled seeds (Wang et al., 2015) . In addition, genetic drift genes  
248 whose expression is affected by random factors, that is, genes that are not  
249 significantly different between species and within species, account for  
250 about 3.09~3.94% of the two subspecies ; the expression of very few genes differs  
251 within species Significant but not significant differences between species, it is a  
252 complex situation, accounting for 0.72 to 0.88% (Figure 2A, B).



254

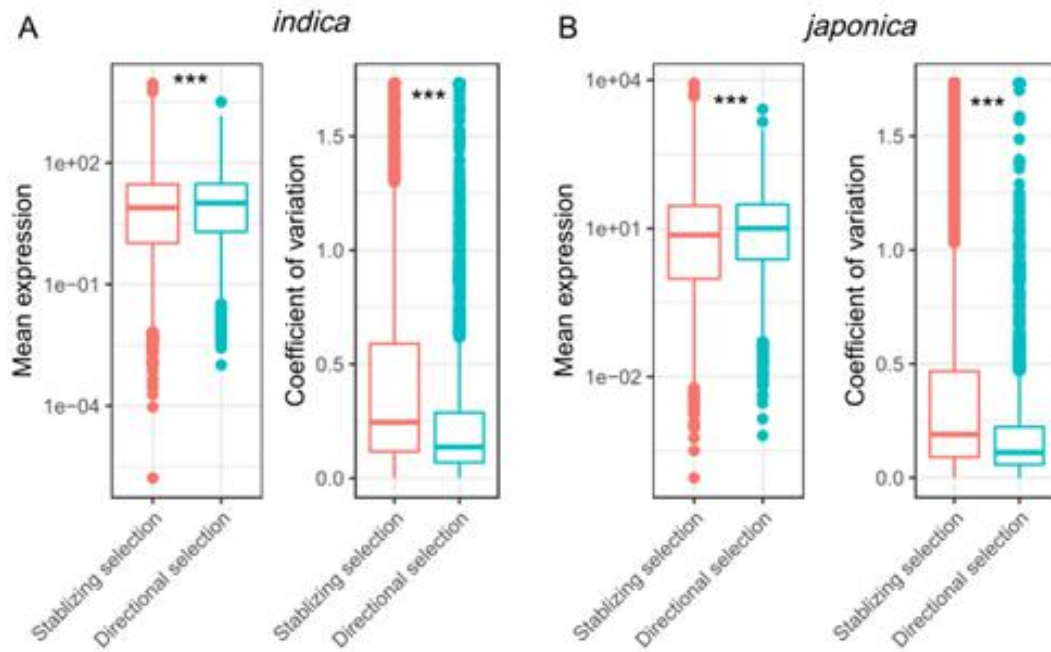
255 **Figure 2.** Evolutionary pattern of gene expression in *indica* and *japonica* rice. The  
 256 ratio of the four types of expression selection in *indica* (A) and *japonica*  
 257 (B). Graphical representation of the expression of four different selection types of  
 258 genes in *indica* (C) and *japonica* (D). The inter-species and intra-species significance  
 259 were obtained by one-way analysis of variance (one-way ANOVA). Directional  
 260 selection: genes with significant differences between species and insignificant  
 261 differences within species; stabilizing selection: genes with insignificant differences  
 262 between species and within species; genetic drift: between species Genes with  
 263 significant differences and intraspecies differences, and complex scenarios: Genes  
 264 with insignificant differences between species and significant differences within  
 265 species. Red: *Indica* rice strain; Blue: *Japonica* rice strain.

266

267

268 **The difference in expression level of directional selection genes and stabilizing**  
 269 **selection genes**

270 Species differentiation is affected by two types of genes: stabilizing selection and  
271 directional selection. Genes with small differences between species and within species  
272 when genes are subjected to stabilizing selection are the continuous selection pressure  
273 for interspecific traits in the population, and this natural selection will not lead to  
274 population differentiation (Charlesworth et al., 1982); genes subjected to directional  
275 selection are genes with large differences in expression between species and small  
276 differences in expression within species. They are the continuous strengthening of  
277 selective pressure for a trait among populations, and are the main genes that cause  
278 differences in traits between species (Mitchell-Olds et al., 2007; Rieseberg et al.,  
279 2002). Through previous research and analysis, we found that *indica* rice and  
280 *japonica* rice are mainly subject to two types of genes, namely stabilizing selection  
281 and directional selection, respectively. We analyze the differences in their expression  
282 levels between subspecies strains. We get TPM median of *indica* gene level is 7.3773,  
283 TPM median of gene subjected to stabilizing selection is 3.6448 . The expression  
284 level of directional selection genes in *indica* rice was significantly greater than that of  
285 stabilizing selection genes (Figure 3A Wilkerson rank sum test,  $p < 0.001$ ). In addition,  
286 the coefficient of variation (0.1364) of directional selection genes was significantly  
287 lower than that of stabilizing selection genes (0.2451) (Figure 3A Wilkerson rank sum  
288 test,  $p < 0.001$ ). We also found the same situation in *japonica*  
289 rice. The median TPM expression level of the directional selection gene in *japonica*  
290 rice was 9.4966, and the gene expression level of the stabilizing selection gene  
291 was 3.4431. The expression level of directional selection genes was significantly  
292 greater than that of stabilizing selection genes (Figure 3B Wilkerson rank sum  
293 test,  $p < 0.001$  ). Similarly, the coefficient of variation of *japonica* gene subjected to  
294 directional selection (0.1106) is significantly lower than the coefficient of variation of  
295 gene subjected to stabilizing selection (0.1904 ) (Figure 3B Weierkesen rank sum  
296 test,  $P < 0.001$ ). Therefore, from the data of gene expression level and variation  
297 level, we can conclude that the expression level of targeted selection genes is higher  
298 and the variation level is lower. Directional selection genes play a more important role  
299 in the process of subspecies differentiation than stabilizing selection genes.



300

301 **Figure 3.** The expression and variation levels of directional selection and stabilizing  
302 selection in *indica* (A) and *japonica* (B). \*\*\* Wilcoxon rank sum test,  $p < 0.001$ . Red:  
303 stabilizing selection gene; green: directional selection gene.

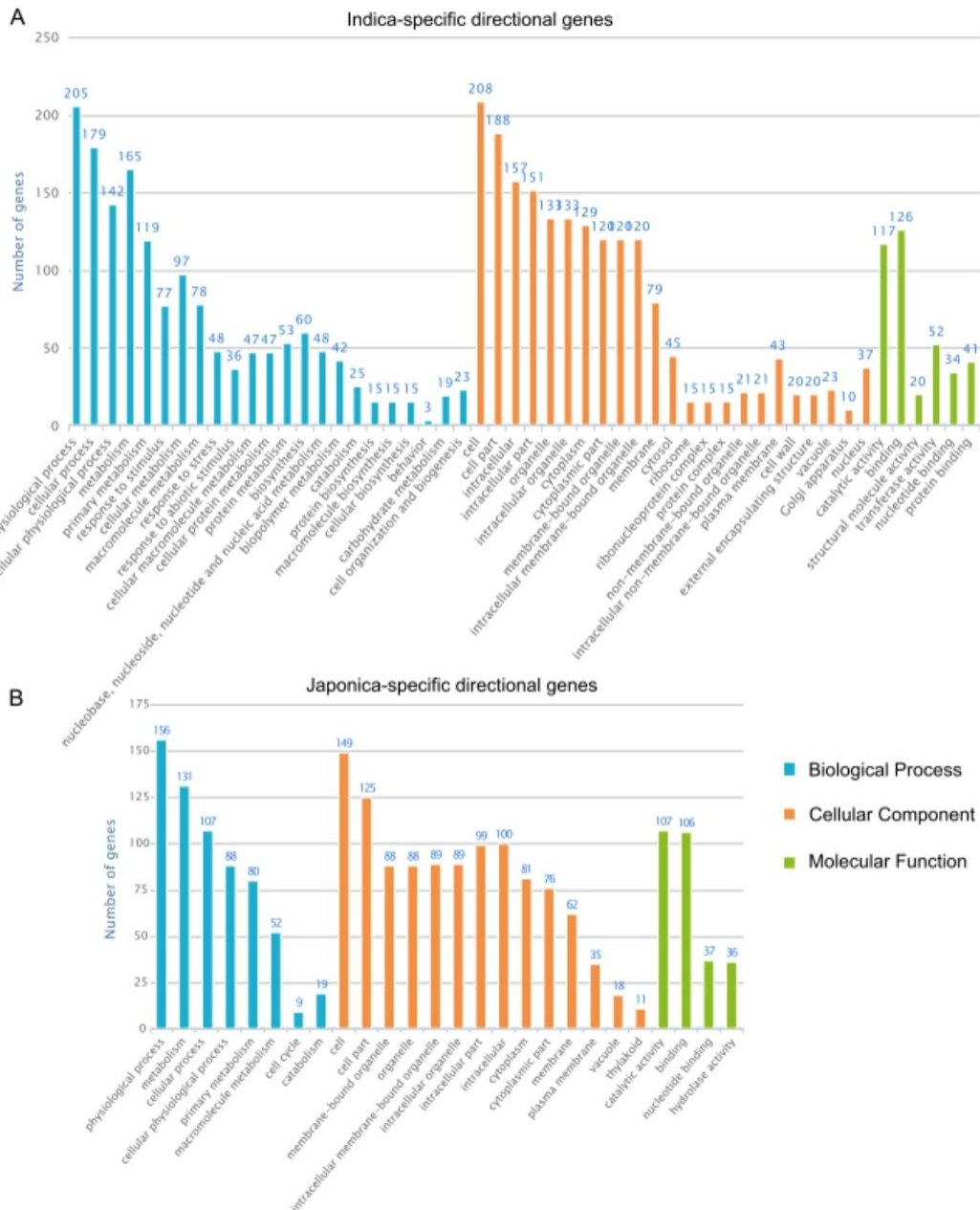
304

305

306 ***Indica* rice species-specific targeted selection genes will enrich biological**  
307 **pathways to cope with environmental pressures**

308 In order to measure the phenotypic difference between *indica* and *japonica* in  
309 gene expression levels, we focused on species-specific directional selection of  
310 genes. Among the 5,910 *indica*-directional selection genes identified, 298 were  
311 *indica*-specific; among the 5,854 *japonica*-directional selection genes, 242 were  
312 *japonica*-specific. Subsequently, we will conduct gene function enrichment analysis  
313 for these subspecies-specific targeted selections (see Materials and Methods) to obtain  
314 functional pathways related to phenotypic differences. In addition to enriching a large  
315 number of common functional pathways, we found that *indica* rice-specific targeted  
316 genes also specifically enriched two types of biological pathways (Figure 4). The first  
317 category is related to coping with environmental stress, such as the response to  
318 stimulus, response to stress, response to abiotic stimulus (Figure 4A). This is related

319 to the different geographical distribution, temperature and light adaptation of the two  
320 subspecies of *indica* and *japonica* (Foll & Gaggiotti, 2006; Mahgoub, 2019; Wu et al.,  
321 2019). The second category is the biosynthesis and metabolism, especially the  
322 Biopolymer Metabolism, Carbohydrate Metabolism, Protein biosynthesis,  
323 biosynthesis Macromolecule, etc (Figure 4A). This type of pathway may be related to  
324 the difference in starch and protein composition between *indica* and *japonica* rice. For  
325 example, *indica* rice has a higher amylose content than *japonica* rice (DuPont &  
326 Altenbach, 2003; Padilla-González et al., 2019). In addition, *indica*  
327 subspecies-specific targeted selection genes also have characteristics in the location of  
328 cellular components, for example, they are specifically located  
329 in non-membrane-bound organelles (intracellular non-membrane-bound organelle,  
330 external encapsulating structure). The molecule function also exist specificity, such as  
331 the activity of the specific structure of the molecules enriched (Structural Molecular  
332 Activity), transferase activity, etc.



333

334 **Figure 4.** Gene Ontology (GO) functional enrichment analysis of species-specific  
 335 directional selection genes in *indica* (A) and *japonica* (B). Blue: biological processes;  
 336 orange: cellular components; green: molecular functions.

337

### 338 Discussion

339 In the process of natural differentiation or artificial selection of species, the core gene  
 340 regions affected by selection pressure are of concern to a wide range of researchers. If  
 341 gene flow continues to affect species undergoing natural differentiation or artificial

342 selection, under different selection pressures, different regions of the same genome  
343 will express different evolutionary results (He et al., 2011) , such as gene  
344 introgression. Differences in genomic DNA and changes in epigenetics are the two  
345 main factors that cause differences between species. Unlike in the past, which mainly  
346 focused on the change of DNA sequence to study species differentiation (He et al.,  
347 2011) , in recent years, more studies have focused on the transcriptome level and the  
348 expression level to explore the factors of species evolution. These studies have found  
349 that gene expression has a wide range of adaptability and is related to species  
350 differentiation (Guo et al., 2016; Wen et al., 2016; Yeh et al., 2014). In the past, some  
351 scholars studied the effects of changes in the species transcriptome on the evolution of  
352 gene expression, and gene chips. But our research has data advantages. We study the  
353 evolution of gene expression, which makes it easier to accurately find genes that  
354 differentiate between species than pure differential expression. In addition, we use  
355 more types of strains and effective methods to obtain more accurate  
356 transcriptomes (Ding et al., 2007; Goff et al., 2002; Huang et al., 2012; Wang et al.,  
357 2014) .

358

359 If the effects on rice differentiation are studied by observing changes in genome, this  
360 will have limitations and deficiencies, such as only studying wild rice and neglecting  
361 artificial selection (Guo et al., 2016) , or lack of measuring the internal expression  
362 variation of multiple strains between subspecies (Guo et al., 2016), or lack of  
363 sequencing data of tissue or developmental stages at that time (Wen et al., 2016). We  
364 use the linear model method to study the interspecific and intraspecific expression  
365 differences of the seedling transcriptome data of six lines of *japonica* rice and *indica*  
366 rice, and divide the genes into four selected gene types. We found that most of the  
367 subspecies with a higher degree of genome similarity are genes that are subjected to  
368 stabilizing selection, followed by genes that are subject to targeted selection. Later, in  
369 further comparing the expression levels and expression variation levels of genes  
370 subject to stabilizing selection and targeted selection, we found that targeted selection  
371 genes have higher expression levels and lower variation levels than stabilizing



372 selection genes. This illustrates the importance of directional selection of genes in  
373 species differentiation. Finally, in the enrichment analysis of genes subject to targeted  
374 selection, we found that these genes are related to environmental pressure. This can  
375 indicate that the directional selection genes of *japonica* rice may have been affected  
376 by environmental pressure, which led to the differentiation of *indica* rice. The study  
377 of species evolution from the perspective of the transcriptome can provide different  
378 perspectives and depths.

379

380 Our research has two main shortcomings. First, from the data level, we only collected  
381 the transcriptome data of *indica* and *japonica* rice seedlings but missed the  
382 transcriptome data of other tissues or developmental stages. Future research can use  
383 more extensive and comprehensive organization and developmental stages to carry  
384 out more in-depth research. Second, there are many models for the evolution of  
385 cultivated rice. There are single-source models, that is they are domesticated together  
386 and then differentiated. There are also multiple independent domestication models,  
387 that is *indica* rice originates from *japonica* rice, and there are independent-origin  
388 models, that is completely independent domestication. In addition, wild rice, the  
389 ancestor of cultivated rice, is also controversial (He et al., 2011; Stein et al.,  
390 2018). However, we only studied the transcriptome of two cultivated rice. Future  
391 research can focus on the transcriptome of different wild rice. Although our results are  
392 insufficient, we have discovered the contribution of the transcriptome of cultivated  
393 rice seedlings to species differentiation. This lays the foundation for future research  
394 on the impact of changes in gene expression levels on species differentiation and  
395 gives certain prospective results, especially the impact of artificial selection on  
396 cultivated rice.

397

#### 398 **Funding**

399 This research received no external funding.

400

#### 401 **Conflict of interest**

402 The authors declare no competing interests.

403

404 **References**

405

406 Sasaki T, Burr B. International Rice Genome Sequencing Project: The effort to  
407 completely sequence the rice genome. *Curr Opin Plant*  
408 *Biol* . 2000;3:138–142.

409 Ting, Y. (1957). The origin and evolution of cultivated rice in China. *Acta Agron.*  
410 *Sinica* 8, 243–260

411 Campbell, MT, Du, Q., Liu, K., Sharma, S., Zhang, C., & Walia, H. (2020).  
412 Characterization of the transcriptional divergence between the subspecies of  
413 cultivated rice (*Oryza sativa*). *BMC genomics* , 21 , 1-16.

414 Charlesworth, B., Lande, R., & Slatkin, M. (1982). A NEO-DARWINIAN  
415 COMMENTARY ON MACROEVOLUTION. *Evolution* , 36 (3),  
416 474-498. <https://doi.org/10.1111/j.1558-5646.1982.tb05068.x>

417 Ding, J., Araki, H., Wang, Q., Zhang, P., Yang, S., Chen, JQ, & Tian, D. (2007).  
418 Highly asymmetric rice genomes. *BMC genomics* , 8 ,  
419 154. <https://doi.org/10.1186/1471-2164-8-154>

420 DuPont, FM, & Altenbach, SB (2003). Molecular and biochemical impacts of  
421 environmental factors on wheat grain development and protein  
422 synthesis. *Journal of cereal science* , 38 (2), 133-146.

423 Foll, M., & Gaggiotti, O. (2006). Identifying the environmental factors that determine  
424 the genetic structure of populations. *Genetics* , 174 (2),  
425 875-891. <https://doi.org/10.1534/genetics.106.059451>

426 Gao, D., Jiang, N., Wing, RA, Jiang, J., & Jackson, SA (2015). Transposons play an  
427 important role in the evolution and diversification of centromeres among  
428 closely related species [Original Research]. *Frontiers in Plant*  
429 *Science* , 6 (216). <https://doi.org/10.3389/fpls.2015.00216>

430 Garris, AJ, Tai, TH, Coburn, J., Kresovich, S., & McCouch, S. (2005). Genetic  
431 structure and diversity in *Oryza sativa* L. *Genetics* , 169 (3), 1631-1638.

432 Goff, SA, Ricke, D., Lan, TH, Presting, G., Wang, R., Dunn, M., Glazebrook, J.,  
433 Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C.,  
434 Katagiri, F., Lange, BM, Moughamer, T., Xia, Y., Budworth, P., Zhong, J.,  
435 Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, WL, Chen, L.,  
436 Cooper, B., Park, S., Wood, TC, Mao, L., Quail, P., Wing, R., Dean, R., Yu,  
437 Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin,  
438 A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T.,  
439 Miller, RM, Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R.,  
440 Feldhaus, J., Macalima, T., Oliphant, A., & Briggs, S. (2002). A draft  
441 sequence of the rice genome (*Oryza sativa* L. ssp.  
442 *japonica*). *Science*, 296 (5565),  
443 92-100. <https://doi.org/10.1126/science.1068275>

444 Guo, J., Liu, R., Huang, L., Zheng, X.-M., Liu, P.-L., Du, Y.-S., Cai, Z., Zhou, L.,  
445 Wei, X.-H., & Zhang, F.-M. (2016). Widespread and adaptive alterations in  
446 genome-wide gene expression associated with ecological divergence of two  
447 *Oryza* species. *Molecular biology and evolution*, 33 (1), 62- 78.

448 He, Z., Zhai, W., Wen, H., Tang, T., Wang, Y., Lu, X., Greenberg, AJ, Hudson, RR,  
449 Wu, C.-I., & Shi, S. (2011). Two evolutionary histories in the genome of rice:  
450 the roles of domestication genes. *PLoS genetics*, 7 (6), e1002100.

451 Huang, X., Kurata, N., Wei, X., Wang, ZX, Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu,  
452 H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T.,  
453 Zhang, L., Wang, Y., Feng, L., Furuumi, H., Kubo, T., Miyabayashi, T., Yuan,  
454 X., Xu, Q., Dong, G., Zhan, Q., Li, C., Fujiyama, A., Toyoda, A., Lu, T.,  
455 Feng, Q., Qian, Q., Li, J., & Han, B. (2012). A map of rice genome variation  
456 reveals the origin of cultivated rice. *Nature*, 490 (7421), 497-  
457 501. <https://doi.org/10.1038/nature11532>

458 Lu, B.-R., Cai, X., & Xin, J. (2009). Efficient *indica* and *japonica* rice identification  
459 based on the InDel molecular method: Its implication in rice breeding and  
460 evolutionary research. *Progress in Natural Science*, 19 (10),  
461 1241-1252. <https://doi.org/https://doi.org/10.1016/j.pnsc.2009.01.011>

- 462 Lu, BR, Zheng, KL, Qian, HR, & Zhuang, JY (2002). Genetic differentiation of wild  
463 relatives of rice as assessed by RFLP analysis. *Theor Appl Genet* , 106 (1),  
464 101-106. [https:// doi.org/10.1007/s00122-002-1013-2](https://doi.org/10.1007/s00122-002-1013-2)
- 465 Ma, J., & Bennetzen, JL (2004). Rapid recent growth and divergence of rice nuclear  
466 genomes. *Proceedings of the National Academy of Sciences of the United*  
467 *States of America* , 101 (34), 12404-12410. [https://](https://doi.org/10.1073/pnas.0403715101)  
468 [doi.org/10.1073/pnas.0403715101](https://doi.org/10.1073/pnas.0403715101)
- 469 Mahgoub, AMMA (2019). The impact of five environmental factors on species  
470 distribution and weed community structure in the coastal farmland and  
471 adjacent territories in the northwest delta region, Egypt. *Heliyon* , 5 (4),  
472 e01441. [https://doi. org/https://doi.org/10.1016/j.heliyon.2019.e01441](https://doi.org/https://doi.org/10.1016/j.heliyon.2019.e01441)
- 473 Melo, D., & Marroig, G. (2015). Directional selection can drive the evolution of  
474 modularity in complex traits. *Proceedings of the National Academy of*  
475 *Sciences* , 112 (2), 470-475. <https://doi.org/10.1073/pnas.1322632112>
- 476 Mitchell-Olds, T., Willis, JH, & Goldstein, DB (2007). Which evolutionary processes  
477 influence natural genetic variation for phenotypic traits? *Nature Reviews*  
478 *Genetics* , 8 (11), 845-856. <https://doi.org/10.1038/nrg2207>
- 479 Molina, J., Sikora, M., Garud, N., Flowers, JM, Rubinstein, S., Reynolds, A., Huang,  
480 P., Jackson, S., Schaal, BA, & Bustamante, CD (2011) . Molecular evidence  
481 for a single evolutionary origin of domesticated rice. *Proceedings of the*  
482 *National Academy of Sciences* , 108 (20), 8351-8356.
- 483 Nielsen, R. (2005). Molecular signatures of natural selection. *Annu Rev Genet* , 39 ,  
484 197-218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- 485 Oka, H.-I., & Morishima, H. (1982). Phylogenetic differentiation of cultivated rice,  
486 XXIII. Potentiality of wild progenitors to evolve the *indica* and *japonica* types  
487 of rice cultivars. *Euphytica* , 31 (1), 41-50 .
- 488 Onishi, K., Takagi, K., Kontani, M., Tanaka, T., & Sano, Y. (2007). Different patterns  
489 of genealogical relationships found in the two major QTLs causing reduction  
490 of seed shattering during rice domestication. *Genome* , 50 (8),  
491 757-766. <https://doi.org/10.1139/g07-051>

- 492 Padilla-González, GF, Frey, M., Gómez-Zeledón, J., Da Costa, FB, & Spring, O.  
493 (2019). Metabolomic and gene expression approaches reveal the  
494 developmental and environmental regulation of the secondary metabolism of  
495 yacón ( *Smallanthus sonchifolius*, Asteraceae). *Scientific Reports* , 9 (1),  
496 13178. <https://doi.org/10.1038/s41598-019-49246-2>
- 497 Purugganan, MD, & Fuller, DQ (2009). The nature of selection during plant  
498 domestication. *Nature* , 457 (7231),  
499 843-848. <https://doi.org/10.1038/nature07895>
- 500 Rieseberg, LH, Widmer, A., Arntz, AM, & Burke, JM (2002). Directional selection is  
501 the primary cause of phenotypic diversification. *Proceedings of the National*  
502 *Academy of Sciences of the United States of America* , 99 (19),  
503 12242-12245. <https://doi.org/10.1073/pnas.192360899>
- 504 Second, G. (1982). Origin of the genic diversity of cultivated rice (*Oryza* spp.): study  
505 of the polymorphism scored at 40 isozyme loci. *The Japanese journal of*  
506 *genetics* , 57 (1), 25-57.
- 507 Stein, JC, Yu, Y., Copetti, D., Zwickl, DJ, Zhang, L., Zhang, C., Chougule, K., Gao,  
508 D., Iwata, A., Goicoechea, JL, Wei, S ., Wang, J., Liao, Y., Wang, M.,  
509 Jacquemin, J., Becker, C., Kudrna, D., Zhang, J., Londono, CEM, Song, X.,  
510 Lee, S. , Sanchez, P., Zuccolo, A., Ammiraju, JSS, Talag, J., Danowitz, A.,  
511 Rivera, LF, Gschwend, AR, Noutsos, C., Wu, C.-c., Kao, S. -m., Zeng, J.-w.,  
512 Wei, F.-j., Zhao, Q., Feng, Q., El Baidouri, M., Carpentier, M.-C., Lasserre, E.,  
513 Cooke , R., Rosa Farias, D. d., da Maia, LC, dos Santos, RS, Nyberg, KG,  
514 McNally, KL, Mauleon, R., Alexandrov, N., Schmutz, J., Flowers, D., Fan, C.,  
515 Weigel, D., Jena, KK, Wicker, T., Chen, M., Han, B., Henry, R., Hsing, Y.-i.  
516 C., Kurata, N., de Oliveira, AC, Panaud, O., Jackson, SA, Machado, CA,  
517 Sanderson, MJ, Long, M., Ware, D., & Wing, RA (2018). Genomes of 13  
518 domesticated and wild rice relatives highlight genetic conservation , turnover  
519 and innovation across the genus *Oryza*. *Nature Genetics* , 50 (2),  
520 285-296. <https://doi.org/10.1038/s41588-018-0040-0>

- 521 Subramanian, A., Tamayo, P., Mootha, VK, Mukherjee, S., Ebert, BL, Gillette, MA,  
522 Paulovich, A., Pomeroy, SL, Golub, TR, Lander, ES, & Mesirov, JP (2005 ).  
523 Gene set enrichment analysis: a knowledge-based approach for interpreting  
524 genome-wide expression profiles. *Proceedings of the National Academy of*  
525 *Sciences of the United States of America* , 102 (43),  
526 15545-15550. <https://doi.org/10.1073/pnas.0506580102>
- 527 Ting, C.-T., Tsaur, S.-C., & Wu, C.-I. (2000). The phylogeny of closely related  
528 species as revealed by the genealogy of a speciation gene,  
529 Odysseus. *Proceedings of the National Academy of Sciences* , 97 (10),  
530 5313-5316.
- 531 Wang, H., Niu, QW, Wu, HW, Liu, J., Ye, J., Yu, N., & Chua, NH (2015). Analysis  
532 of non-coding transcriptome in rice and maize uncovers roles of conserved  
533 lncRNAs associated with agriculture traits. *Plant J* , 84 (2),  
534 404-416. <https://doi.org/10.1111/tpj.13018>
- 535 Wang, MX, Zhang, HL, Zhang, DL, Qi, YW, Fan, ZL, Li, DY, Pan, DJ, Cao, YS, Qiu,  
536 ZE, Yu, P., Yang, QW, Wang, XK, & Li, ZC (2008). Genetic structure of  
537 *Oryza rufipogon* Griff. in China. *Heredity (Edinb)* , 101 (6),  
538 527-535. <https://doi.org/10.1038/hdy.2008.61>
- 539 Wang, X., Kudrna, DA, Pan, Y., Wang, H., Liu, L., Lin, H., Zhang, J., Song, X.,  
540 Goicoechea, JL, Wing, RA, Zhang, Q ., & Luo, M. (2014). Global genomic  
541 diversity of *Oryza sativa* varieties revealed by comparative physical  
542 mapping. *Genetics* , 196 (4),  
543 937-949. <https://doi.org/10.1534/genetics.113.159970>
- 544 Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for  
545 transcriptomics. *Nat Rev Genet* , 10 (1), 57-63. <https://doi.org/10.1038/nrg2484>
- 547 Wen, M., Xie, M., He, L., Wang, Y., Shi, S., & Tang, T. (2016). Expression  
548 variations of miRNAs and mRNAs in rice (*Oryza sativa*). *Genome biology*  
549 *and evolution* , 8 (11), 3529-3544.

- 550 Wu, Z., Xu, X., Zhang, J., Wiegleb, G., & Hou, H. (2019). Influence of environmental  
551 factors on the genetic variation of the aquatic macrophyte *Ranunculus*  
552 *subrigidus* on the Qinghai-Tibetan Plateau. *BMC Evolutionary Biology* , 19 (1),  
553 228. <https://doi.org/10.1186/s12862-019-1559-0>
- 554 Yeh, S.-D., von Grotthuss, M., Gandasetiawan, KA, Jayasekera, S., Xia, X.-Q., Chan,  
555 C., Jayaswal, V., & Ranz, JM (2014). Functional divergence of the miRNA  
556 transcriptome at the onset of *Drosophila* metamorphosis. *Molecular biology*  
557 *and evolution* , 31 (10), 2557-2572.  
558 *Douding Network* . <https://www.docin.com/p-2595022643.html>
- 559 *Flower Encyclopedia* . <https://nong.huabaike.com/n/5849.html>
- 560 Jiang Jian , Li Jinqun , Xu Zhengjin , Zhang Longbu , & Jin Chenghai .  
561 (2001). Differences in the main morphological and anatomical characteristics  
562 of *indica* and *japonica* rice and their relationship with economic traits . *Jilin*  
563 *Agricultural Sciences* , 26 (6), 11-15.
- 564 Lu Baorong , Cai Xingxing , & Jin Xin . (2009). High-efficiency molecular  
565 identification methods for *indica* and *japonica* rice and their significance in  
566 rice breeding and evolution . *Advances in Natural Science* , 019 (006),  
567 628-638.
- 568 Miao Xiangwei , & Wang Dexin . (2009). Research on the correlation between rice  
569 subspecies characteristics and economic traits . *Seed World* (12), 28-31.
- 570 Ruan Bosheng . (2008). *Creation of a mutant bank of japonica*  
571 *rice " Nipponbare " and genetic analysis of some important traits Zhejiang*  
572 *University ]*.
- 573 *Rice* *Seedling* *Nursery*  
574 *report - Douding* . <https://www.docin.com/p-1870714240.html>
- 575 Xu , Liu Hongguang , Zhuchun Jie , Yang Li , GUO Yan-hua , Wang Jiayu , Yang  
576 Gan Hua , Xu into , Zhengjia Kui , & Chen Wenfu . (2007). Effects of  
577 ecological condition of Subspecies Characteristics and Economic Characters  
578 of *Indica* rice hybrids . *Zirankexue progress* .

579 Xu Zhengjin , Li Jinquan , Jiang Jian , Jing Yanhui , Zhang Wenzhong , Chen Wenfu ,  
580 & Zhang Longbu . (2003). Subspecies characteristic traits of *indica* and  
581 *japonica* hybrids and their relationships with economic traits . *Acta*  
582 *Agronomica Sinica* , 029 (005), 735 -739.  
583  
584