INVESTIGATION AND GENETIC ANALYSIS OF YIELD TRAITS AND DROUGHT TOLERANCE OF *ORYZA RUFIPOGON* USING ADVANCED BACKCROSSING INBRED POPULATIONS

Luu Ngoc Sinh[1], Nguyen Nhu Toan[2], Peiman Zandi[3], Saikat Kumar Basu[4]

[1,2] Hanoi Metropolitan University, [3] Yibin University, China,

[4] PFS, Lethbridge Alberta Canada T1J 4B3

Abstract: *Rice (Oryza sativa L.) is one of the staple cereal crops of the world and it is one of the main sources of carbohydrates for nearly one-half of the world's population. So, an increase in rice yield to satisfy human needs is a pressing requirement. Agronomic characters such as high plant, length of panicle, length of flag leaf, number of seed per panicle, the weight of thousand rice, length and width of rice and tolerance to salt are the deciding factor for yield in rice. Today's rice is derived from the common wild rice (Oryza rufpogon) and inherited its advantages. However, during the rice domestication process, some valuable features of wild rice, such as tolerance to biotic and abiotic stress, were lost. To fully utilize wild rice germplasm resources, we constructed a set of introgression lines (ILs) using a common wild rice material from Lingshui, China. A set of high-resolution InDel molecular markers with an average interval of 2.39 Mb were designed to carry out marker-assisted selection and identification of segment characteristics. The ILs contained 213 lines including 1.286 introgressed fragments with an average length of 6.511 Mb, covering 93.59% of the donor parent's chromosomes. The agronomical characters of 213 lines were investigated. Many old quantitative trait loci (QTLs) involved in plant height, awn length, seed traits and other characteristics reappeared in our ILs, proving that our system was reliable. Further, many new QTLs were identified. A QTL related to drought tolerance located on chromosome 4 was thoroughly elaborated. This set of ILs provides a new resource for utilizing the excellent features of wild rice.*

Keywords: *Rice, QTL mapping, Quantitative Trait Locus.*

Received 13 March 2023

Revised and accepted for publication 24 July 2023

Contact author: Luu Ngoc Sinh; Email: lnsinh@daihocthudo.edu.vn

1. INTRODUCTION

Backgroud

Rice belongs to the genus Oryza, and comprised more than 22 species widely distributed across the tropical, sub-tropical and temperate regions though only two species are cultivated:

Oryza glaberrima (Steudel) and *Oryza sativa* (L.) (Lu, 1999) of which the later one comprises the most common rice varieties, indica and japonica. It is widely grown in Asia, Africa, and also in Latin America and covers about 40 countries under a broad range of ecosystems between 55^0 N and 36^0 S including irrigated, rainfed upland, and flood-prone areas (Brondani, et al., 2006). About 90% of the world's rice is grown in Asia (Chandler, 1979) and one-third of the world's population consumes rice as a staple food (Lee, et al., 2006).

Relationship between stable QTL with environment effects stable yield

At the start of the 21st century, statistical methods for quantitative trait locus (QTL) analysis are being refined and added to at an astonishing rate. Methods based on the analysis of quantitative traits tend to be more powerful than the equivalent binary trait methods, and, in conjunction with whole-genome screening technologies, are yielding exciting results in agriculture. QTL analysis and a variety of associated innovations will likely continue to provide key landmarks. GWAS are becoming increasingly popular in genetic research, and they are an excellent complement to QTL mapping. QTL studies have a long and rich history and have played important roles in gene cloning and characterization. Data generated by lab-based QTL studies can also be used to direct and inform other efforts, such as population genomics, wherein a large number of molecular markers are scored in the attempt to identify targets of selection and thus genes underlying ecologically important traits (Stinchcombe, 2007). Furthermore, QTL studies can inform functional genomics, in which the goal is to characterize allelic variation and how it influences the fitness and function of whole organisms. Nowadays, the rapid improvement of sequencing technology has offered more powerful and detailed tools for insight into genetic information at a whole genome level.

Molecular characterization of the rice obtained from biolistic transformation with constructs containing GUS, HPT and barcoding regions, each driven by a 35S promoter showed that the transgene was typically found to be inserted at one locus and was present in 1–9 copies of rearranged or truncated sequence (Kohli, et al., 1999). The increase in transgene copy number did not always lead to a concomitant decrease in expression levels or silencing but the integrity of integrated transgenes was a major factor in the onset of silencing. They observed that the presence of truncated sequences of transgenes capable of generating incomplete transcripts, resulting in aberrant RNA species, may be responsible for silencing. Li (2005) reported introgression lines (ILs) of uniformed genetic backgrounds with introgressed fragments from donors of diverse origin in BC breeding combined with appropriate selection pressure have approved to be effective in exploring the 'hidden genetic diversity' in the primary gene pool of rice for genetic improvement.

QTL in rice

Qualitative and Quantitative Traits. Qualitative or Mendelian traits have discrete phenotypes and are controlled by a single or few genes. They are caused by mutations that have major effects on the phenotype (macromutations). Phenotypes such as wingless flies, hairless mice and dwarf plants are conditioned by macromutations at single loci (Tanksley, 1993). Loci controlled by macromutations are easy to study because they allow the genotype of a particular locus to be predicted from the phenotype of the individual using Mendelain genetics. However,

phenotypic variation is usually continuous instead of discrete and controlled by several genes with relatively small effects. Characters whose phenotypic variation is determined by several loci are called quantitative traits and their inheritance is polygenic (Tanksley, 1993). The individual loci controlling a quantitative trait are known as polygenes or quantitative trait loci (QTL) (Tanksley, 1993). Polygenes control most agronomically important traits (Jena & Mackill, 2008). These traits could not be studied by classical Mendelain techniques leading to the emergence of a subspecialty of genetics called quantitative genetics.

Quantitative genetics relies on statistics to describe the characteristics of continuous phenotypic distributions. These statistics help to estimate some genetic information including the approximate number of loci affecting a character in a particular mating, the average gene action and the degree to which the various polygenes interact with each other and the environment in determining the phenotype (Tanksley, 1993; Falconer & Mackay, 1996). Classical quantitative genetics tools, therefore, consider only the aggregate effects of all the genes causing the variation and do not take into account the properties of genes individually - their gene frequencies and the magnitude of their effects on the trait of interest (Falconer & Mackay, 1996).

2. MATERIALS AND METHODS

2.1. Materials

Plant materials

Oryza sativa Japonica, *O. sativa* has a compact diploid genome of approximately 500 Mbp;

The *Oryza sativa*, that have important traits for yield 9311;

The donor parent, LSWR, was common wild rice (*O. rufpogon*) collected from Lingshui City, Hainan Province, China (18°52′N, 110°02′E). The LSWR was selected as the donor parent because of its drought resistance, waterlogging resistance and strong root oxygen secretion. The recurrent parent was an elite *Indica* cultivar 9311 (*O. sativa* ssp. *indica* cv. 9311) kindly provided by Professor Huang Niansheng (Lixiahe District Institute of Agricultural Sciences, Jiangsu Province, China).

Indel marker

We compared the reference genome sequences of *O. sativa* ssp. *indica* and *O. sativa* ssp. *japonica* in the database (http://www.gramene.org/) and then searched for variations (insertions or deletions) between the two kinds of genomes. For the convenience of detection, the gap of each InDel was chosen as approximately 20 bp, while the PCR products would be approximately 200 bp. Primers were designed by DNAMAN version 6.0. The accurate physical location and specificity of the primers were determined by the Basic Local Alignment Search Tool (BLAST) of the genome of *O. sativa* ssp. *japonica*, the reference genome used in this study (http://rapdb.dna.afrc.go.jp/download/irgsp1.html). The melting temperatures (Tm) of primers were calculated by the method of SantaLucia (SantaLucia, 1998). The average interval

of the markers was calculated by dividing the cumulative length of 12 chromosomes by the number of markers.

Experimental Equipment and Chemicals

Experimental equipment and Chemicals were used in an experiment from the Biotechnology Research Institute (BRI), Chinese Academy of Agricultural Sciences (CAAS), China, and companies in China, the USA, and Japan.

2.2. Methods

DNA extraction and PCR reaction

DNA extraction from rice leaves

1. Collect 1 to 3 grams of rice leaf tissue, grind in liquid N_2 and transfer the fine powder to a 50 ml centrifuge tube;
2. Immediately add 700 ml CTAB buffer and mix thoroughly;
3. Incubate at 65°C for 2 hours;
4. Add 700 ml chloroform and invert the tube several times to mix thoroughly;
5. Centrifuge at room temperature at 10,000rpm for 10 minutes;
6. Transfer supernatant to new centrifuge tube;
7. Add 500ml equal volume of isopropanol, and invert the tube gently several times;
8. Incubate at –20°C for 60 minutes or overnight;
9. Spool out DNA fibre with pasture pipette hook or centrifuge at 12,000 rpm for 10 minutes to collect pellet;
10. Discard the supernatant and wash the DNA pellet with 75% ethanol;
11. Centrifuge at 12,000 rpm for 5 min;
12. Discard the supernatant and let the pellet dry for 10-15 min;
13. Resuspend DNA with Water or TE;
14. Test DNA with NanoDrop 200 software.

PCR reaction

A 10-μL PCR reaction was set up as follows: 5 μL of 2 × *Taq* PCR Star Mix with Loading Dye (Genstar Biotech Co., Ltd., Beijing, China), 1 μL of forward primer (10 μM), 1 μL of reverse primer (10 μM), 1 μL of DNA sample (50 ng/μL) and 2 μL of ddH2O.

Step	Temperature for reaction	Time
1	Predenaturation at 94 °C	2 min
2	Denaturation at 94 °C	30 s
3	Annealing at 60 °C	30 s
4	Extension at 72 °C	12 s
5	Go to step 2	35 cycles
6	Final extension at 72 °C	5 min

The PCR products were detected by electrophoresis in 4% agarose gels.

Samples that shared a banding pattern identical to LSWR were recorded as A, and those identical to 9311 were recorded as B. Samples with multiple bands corresponding to both identified bands were treated as hybrids and recorded as H, as determined by the MapMaker software *(Lander, et al., 1987)*. Graphical genotype visualization was achieved by ABHgenotypeR, a package of R *(Reuscher & Furuta, 2016)*. We assumed that the total fragment belonged to LSWR when the marker genotypes on both ends were A; belonged to 9311 when both ends were B; and were half LSWR and half 9311 when the two ends had different genotypes. The length of introgressed fragments was calculated by this assumption *(Young & Tanksley, 1989)*. The coverage was obtained by calculating the ratio between the cumulative introgressed fragment length and the whole genome length.

Backcross

Backcross breeding is an effective method to transfer one or a few genes controlling a specific trait from one line into a second—usually elite—breeding line. The parent with the desired trait, called the donor parent, provides the desired trait and may not perform as well as an elite variety in other areas. The elite line, called the recurrent parent, usually performs well in all other areas. Backcrossing involves making an initial cross between the donor and recurrent parents.

Measuring Seed Shape through Image Analysis

We have measured seed shape through image analysis for Tanabana, et al (Tanabata, et al., 2012), In this study, we developed a high-throughput phenotyping program called SmartGrain that uses image analysis to determine seed shape. SmartGrain automatically recognizes all seeds within a digital image, detects outlines, and then calculates L, W, seed area (AS), perimeter length (PL), and other parameters. To validate the software, we used it in QTL analysis for rice (Oryza sativa) seed shape, which is difficult to automatically measure because of pedicels and awns. The pedicel is the stalk supporting a spikelet on a panicle branch and the awn is a filiform extension of varying length protruding from the top of a lemma (Chang & Bardenas, 1965)

Measure Panicle length

The samples were measured and repeat three times. The data were analysed by a mathematical formula, Measure the panicle length of the tallest plant on the hill from the neck node (panicle base node) to the end of the panicle.

Step 1: After finishing the measurement of the culm length, measure the panicle length above the same culm.

Step 2: Stretch out the panicle on the meter stick.

Step 3: Measure and record by 0.2 cm.

Construction of the IL population

The crossing was carried out between LSWR and 9311, during which 9311 was regarded as “female” and LSWR as“male.” Then, we obtained BC1F1 individuals by backcrossing to 9311. This step was repeated three times until the BC_3F_1 individuals were obtained; then, selfing and MAS were applied. We obtained advanced progenies after seven generations of selfing or

backcrossing, from which homozygous individuals with only a few fragments of the donor parent's chromosome were selected as ILs.

Phenotype investigation

Investigated important agronomic traits of ILs planted in a paddy field located in Langfang, for 3 years, including architectural traits such as plant height and fag leaf length; yield and quality traits such as grain length, grain width, grain length/grain width, grain perimeter, grain area, thousand-kernel weight and thousand-seed weight (Tanabata, et al., 2012).

The panicles of LSWR were fully awned with an awn length longer than 2 cm, while 9311 was unwanted. The progenies of the two parents showed separated traits regarding awns. We set a standard to evaluate the level of awns, wherein absent would be recorded as 0 similar to 9311, partly awned panicles with a relatively short awn length of less than 1 cm would be recorded as 1, fully shortawned panicles or partly awned panicles with a relatively long awn length of 1–3 cm would be recorded as 2, fully long-awned panicles would be recorded as 3, and fully very long-awned with an awn length more than 3 cm would be recorded as 4, similar to LSWR. Some qualitative traits such as the colours of the apiculus, stigma, pericarp and husk were also investigated.

Evaluation of drought tolerance at the seedling stage

Uniformly germinated seeds of ILs were sown in 96-well plates floating on Yoshida nutrient solution (Yoshida, et al., 1976) and cultured in a greenhouse at 30 °C (light) and 22 °C (dark) under 16-h-light/8-h-dark conditions. Then, seedlings at the 4th leaf stage were treated with 20% PEG (polyethylene glycol, with a molecular weight of 4000), a drought simulation reagent. After 10 days of drought treatment, phenotypic changes in the treated plants were carefully investigated and photographed. Each treatment was repeated three times.

Drought tolerance QTL confirmation on chromosome 1

The *NAL1* gene containing a 3-kb promoter region and a 1-kb downstream region was sequenced between IL line No.28 and 9311. To generate an overexpression vector, *NAL1* full-length cDNA (1749 bp) was amplified by PCR from IL28 with the primers OP1 containing a *Sma*I digestion site and OP2 containing an *Xba*I digestion site. The PCR product was cloned into the pEASY-Blunt simple cloning vector (Transgen Biotech) and sequenced. The vector was introduced into *Agrobacterium tumefaciens* (stain EHA105) by electroporation. *Agrobacterium*-mediated transformation was performed using vigorously growing calli derived from mature embryos of *Japonica* Nipponbare following an improved protocol described previously (Yang, et al., 2004). For stress tolerance evaluation of *NAL1*-overexpressing plants, seedlings of NAL1-OX and *Nipponbare* at 4th leaf stage were treated. The relative water contents and malondialdehyde (MDA) contents were measured immediately before and after drought stress treatments.

Drought-sensitive mutant *dsm1* screening and complementation

From the rice T-DNA enhancer trapping library (Yang, et al., 2004), a drought-sensitive mutant (*dsm1)* was achieved at the seedling stage. Map-based methods were used to clone the drought-resistance gene. To generate a complementation vector, full-length *NAL1* cDNA was cloned into the *p*Cambia2300 vector carrying a ubiquitin promoter to generate pUbi::NAL1.

The pUbi::NAL1vector was transformed to *Agrobacterium tumefaciens* (stain EHA105) by electroporation. The mutant *dsm1* calli were induced, and complementary lines were created. For stress tolerance evaluation of *NAL1* complementary plants, seedlings of the complementary line (NAL1-CP) and Nipponbare at the 4th leaf stage were treated. The relative water contents and malondialdehyde (MDA) contents were measured immediately before and after drought stress treatments.

All of the experiments were conducted at least three times. The results are presented as the mean ± standard deviation (SD).

3. RESULTS

Genetic analysis for grain weight. The grain weights of 213 ILs were evaluated by measuring two indicators, thousand-kernel weight (TKW) and thousand-seed weight (TSW), after drying. TKW and TSW were positively correlated (r =0.817, p=7.466e−20), where TKWs ranged from 15.55 to 23.77 g and TSWs ranged from 19.96 to 34.54 g. ILs that showed the four lowest grain weights (both TKW and TSW) were highly linked with short grain length or width. Among them, IL29, IL36, IL40 and IL87 had TKWs of 18.24 g, 17.44 g and 17.32 g and TSWs of 23.21 g, 24.11 g and 21.54 g, respectively. As a control, the TKW and TSW of 9311 were 20.83 g and 28.62g. The 3 lines had an overlapped region on Chr. 3 where GS3 is located (Fan et al. 2006). IL23 had a TKW of 15.84 g and a TSW of 19.96 g, which are 23.79% and 30.31% lower than those of 9311, respectively. GIF1 located on Chr.1 was within the introgressed fragment of IL23 and may confer the low-weight phenotype of grainsc (Wang, et al., 2008).

ILs with high coverage were constructed. We chose 213 lines with fewer introgressed fragments as individual ILs. We identified that each individual contains a minimum of one chromosome fragment of the donor parent, a maximum of 3 fragments and an average of 1.286 fragments. The length of introgressed fragments ranged from 1.075 to 17.405 Mb, with an average length of 6.511 Mb. The accumulative fragments overlapped and interlocked to cover 93.59% of the donor parent's chromosome. There are 88 lines carrying heterozygous fragments, which need further selfing or backcrossing to become immortal populations. Since these residual heterozygous lines (RHLs) will segregate in the spring, the RHLs containing the target fragments could be used in fine-mapping loci corresponding to the traits in further experiments.

Agronomic traits investigation. We investigated the agronomic traits of 214 ILs with the donor parent's chromosome fragments and found that all showed stable and heritable phenotypes that differed greatly from each other. These ILs also have some specific traits that differ from those of cultivated rice, such as height, late flowering, red awns, purple apiculus and black husks. Red and grey blocks indicate the chromosome fragments of LSWR and 9311, respectively. Yellow blocks indicate uncharacterized genotypes (colour figure online).

Genetic analysis of plant height. The plant height of recurrent parent 9311 is approximately 115.9 cm, whereas 4 ILs (No. 7, 8, 9 and 30) had a height over 170 cm, which is 81%, 74%, 89% and 52.2% higher than that of 9311, respectively. The introgressed fragments of IL7, IL8, IL9 and IL30 have an overlapped region of chromosome 1 (linked with marker Ind1-351–Ind1-370). Near this region, Asano reported that the *Sd1* gene had been subjected to artificial selection during rice evolution (Asano, et al., 2011). We speculated that the *Sd1* gene

was responsible for the ILs showing extreme plant height phenotypes. The sequencing result verified that 4 ILs contain the wild *Sd1* locus.

Genetic analysis of panicle length. Panicle length is one of the important factors influencing rice yield. We found that IL59 has a panicle length of 23.5 cm, the shortest among all the ILs, and 10.65% lower than that of 9311(26.3 cm). IL59 has only one introgressed fragment with a length of 13.79 Mb on Chr.10 (marker Ind10-038–Ind10-154). There were no reported loci controlling panicle length near this region. IL7 (32.2 cm), IL8 (32.3 cm), IL9 (36.5 cm) and IL30 (31.9 cm) had the longest panicles, corresponding to their highest plant height, indicating that the variation of long panicles was caused by natural variation in plant height.

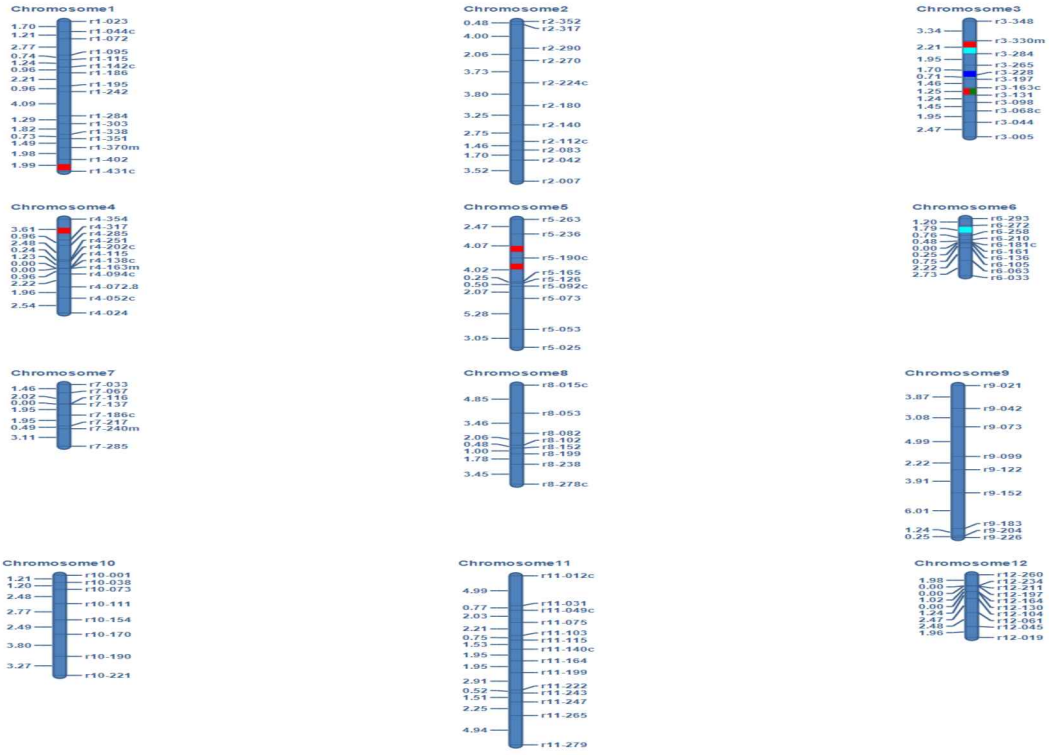
Genetic analysis of grain length. We detected grain lengths, ranging from 5.56 mm and 7.48 mm. The grain length of IL20, IL48 and IL 63, which have introgressed fragments containing In3-197, was 5.814 mm, 5.806 mm and 5.562 mm, which was 13.76%, 13.88% and 17.50% lower than that of 9311, respectively. In this region, the gene *GS3* has been reported (Fan, et al., 2006). Another IL had a significantly shorter grain length (14.18% shorter) compared to 9311 (No. 25, 5.786 mm). IL25 has a single introgressed fragment from Chr. 4 (Ind4- 072–Ind4-202), harbouring the previously reported *GIF1*gene (Wang, et al., 2008a). IL25 showed increased grain chalkiness, confirming our speculation that *GIF1*, which results in grain incomplete filling, caused the low grain length of IL25.

The grain length of IL46 was 7.200 mm, which was 6.79% higher than that of 9311. It has a single introgressed fragment from Chr. 7 (marker Ind7-240–Ind7-285), harbouring the previously reported *GL7*/*GW7*gene (SK, et al., 2015). IL16, IL75 and IL77 showed a significantly increased grain length compared to 9311 (at least 7.37% higher) and shared a 4.50-Mb overlapped region on Chr. 12 (maker Ind12-234–Ind12-260), where no known genes involving grain length are located.

Genetic analysis of drought tolerance. Since the LSWR we selected harboured many stress resistance characteristics, the drought tolerance of 77 IL lines was investigated using the PEG-stimulated method at the seedling stage. The results showed that 3 ILs (No. 15, No. 28, and No. 63) had conferred drought tolerance to some degree according to the wilt rate degree. IL15 has donor parent fragments located on Chr. 2 (19.0–30.4 Mb) and Chr. 11 (8.9–15.2 Mb). IL28 has donor parent fragments located on Chr. 4 (29.7–34.9 Mb) and Chr. 9(8.0–21.5 Mb). IL63 has donor parent fragments located on Chr. 3 (18.0–24.7 Mb), Chr. 4 (29.7–33.8 Mb) and Chr. 10 (13.3–20.6 Mb). Two lines (No. 28 and No. 63) had an overlapping region on Chr. 4 (29.7–33.8 Mb), where Ding reported a drought resistance QTL identified using a population derived from upland *Japonica* variety IRAT109 and lowland *indica* variety Zhenshan 97 (Ding, et al., 2011). We speculated that *NAL1*, which was previously reported to control leaf width and affect vein patterning and polar auxin transportation (Chen, et al., 2012; Qi, et al., 2008), is the most promising candidate gene. To validate whether the *NAL1* gene was responsible for drought tolerance in IL28 and IL63, we first detected the expression level of *NAL1* in IL28 and IL63. The result showed the expression level of *NAL1* in IL28 and IL63 was higher than that in 9311(Figure S4). *NAL1* may be one of the candidate genes responsible for drought tolerance in IL28 and IL63. Thus, we sequenced the *NAL1* gene including the 3-kb promoter and 1 kb downstream of the stop codon in the two lines and 9311. The sequencing result showed that 7 SNPs located upstream of the *NAL1* gene and no InDels were found compared with 9311 (Fig.

3b). The *NAL1* open reading frame (ORF) was no different among No. 28, No. 63 and 9311. Thus, we speculated that the promoter SNPs may be responsible for No. 28 and No. 63 drought tolerance. To further verify that the *NAL1* gene mediated the drought stress response, we transformed the *NAL1* overexpression vector (full codon sequence, CDS) in the *Japonica Nipponbare* background because it is easy to transform. More than 20 overexpressing lines were obtained in the T1 progeny. During the 20% PEG drought stress treatment, the relative water content and malondialdehyde (MDA) content were measured. The result showed that T 1 seedlings overexpressing the *NAL1* gene had a higher relative water content and lower malondialdehyde (MDA) content compared with the control after the drought treatment.T1 seedlings overexpressing the *NAL1* gene could significantly improve the drought tolerance during the seedling stage. Meanwhile, a drought-sensitive mutant *dsm1* from our rice mutant library was also screened and used to verify that *NAL1* participated in drought resistance (Yang, et al., 2004). Through rough and fine mapping, the *DSM1* gene was narrowed to a 3-kb region including four genes We sequenced the ORFs and found only a single base change in the third ORF (LOC_Os04g52479) within exon 3. This polymorphism changed a leucine (Leu) residue into a histidine (His) residue We performed 20% PEG treatment on complementary line1 (cp1) and Nipponbare (Nip). After 4 days of treatment, the cp1 and Nip plants wilted. After rehydration for 4 days, a photograph was taken and presented. The complementary test verified that *NAL1* could complement the drought-sensitive phenotype of the *dsm1* mutant during PEG treatment at the seedling stage. The result further demonstrated that the *NAL1* gene was perhaps involved in the drought stress response in rice, which also indirectly supports that *NAL1* was responsible for the drought resistance phenotype.

QTLs linkage traits yield



4. DISCUSSION

In this study, a novel set of ILs carrying chromosome fragments of LSWR, a variety of common wild rice (*O. rufpogon*) from Lingshui City, Hainan Province, China (18°52′N, 110°02′E), was constructed by backcrossing, selfing and MAS. The elite rice variety 9311 (*O. sativa* ssp. *indica*) was used as a recurrent parent. A set of InDel molecular markers were designed based on the variation between the genome sequences of *Indica* and *Japonica* rice, and a physical map with high resolution was constructed (Gof, et al., 2002; Yu, et al., 2002). This set of markers contains 156 primer pairs that were evenly distributed across the 12 chromosomes with an average physical interval of approximately 2.39 Mb. We applied MAS using these markers and finally selected 77 individuals. The ILs carried 1.285 chromosome segments from the donor parent on average with an average length of 6.511 Mb, overlapping to cover 93.59% of the whole genome of wild rice. Meanwhile, we investigated several important agronomic traits of the ILs including plant height, grain length, grain width, grain weight and awn length. Many QTLs corresponding to those traits were identified, some of which were found in previous studies and some have not been reported previously. Furthermore, a drought tolerance QTL on chromosome 4 was thoroughly investigated. We believe that a set of ILs would serve as a valuable resource for utilizing the excellent characteristics of wild rice.

Introgression lines, or chromosome segment substitution lines, refer to a series of lines that carry different chromosomal segments of the donor parents in the same genetic background, and the accumulated segments basically cover the entire donor parental genome. Different ILs differ only in a few chromosomal regions, which can effectively reduce the interference of genetic background variation, maximally eliminate the epistatic effect from the donor parent and improve the identification probability of the mini-effect QTLs. Therefore, ILs have obvious advantages in QTL identification, QTL fine mapping, QTL effect analysis and cloning of the underlying QTL genes.

Compared with existing ILs, our ILs have some unique advantages. First, we chose LSWR and *Indica* variety 9311 as parents, taking into account the excellent characteristics of both varieties. LSWR, a wild rice with inferior economic traits, perhaps contains special genes, such as its drought resistance, waterlogging resistance and strong root oxygen secretion, which have been lost in cultivated rice, while variety 9311 is an elite *Indica* rice whose genome has been sequenced (Yu, et al., 2002). Such a combination provided sufficient polymorphism contributed to QTL analysis and would apply the mapping population to breeding directly. In the previous study, Qiao reported the construction of introgression lines using *O. rufpogon* as a donor parent in *O. sativa* background (Qiao, et al., 2016). Compared with its CSSLs, the set of ILs in our study has fewer lines due to fewer and longer introgressed fragments, which guarantee a lower repetition rate of genotypes from donor parent between each line, thus providing sufficient effective information even on a small scale. Under the premise of high coverage, the limited number of introgressed lines and chromosome fragment avails genetic breeding program since it can greatly alleviate workload and effectively reduce the genetic noise and epistasis effect. It is kept more compact and is easily handled. Second, each line of our ILs contains an average of 1.27 introgressed segments, which made the phenotype identification reliable and would largely reduce the complexity and increase the accuracy of QTL mapping. Third, compared

with other molecular markers, such as SNP and SSR markers used in the development of ILs reported previously, InDel markers with high resolution and polymorphism were chosen in this study. This would significantly increase the sensitivity and accuracy of identifying the introgression segments of the ILs. Fourth, our results showed that not only some old QTLs or genes corresponding to plant height, awn length, seed weight and other traits were detected in the IL population, but also some new QTLs were identified, providing directions mining candidate genes. Fifth, 88 backcross inbred lines containing heterozygous segments were found. If a QTL was located within the heterozygous region, the heterozygous lines were similar to F1 progeny and could be used directly for cloning gene analysis. This would avoid crosses between lines and save time. Sixth, ILs with special fine traits such as large grain size and high tolerance to drought were screened out and would provide important germplasm resource materials for green super rice breeding.

Most previously identified genes, such as *Sd1*, *An2* and *GL7*, were found in our ILs, further verifying the reliability of our system. Additionally, some new loci were discovered to be involved in traits such as grain length, awn length and panicle length. Cloning these genes would reveal the domestication characteristics of rice. Among the reported rice domestication-associated genes, *GL7* (Wang, et al., 2015b; Wang, et al., 2015c), *CYP78A13* (Xu, et al., 2015) and *GS3* (Fan, et al., 2006) related to Asian cultivated rice domestication and *GL4* (Wu, et al., 2017) related to African rice domestication have been identified to contribute to grain length; *An-1* (Luo, et al., 2013b), *An-2*/*LABA1* (Gu, et al., 2015; Hua L, 2015) and *RAE2*/*GAD1* (Bessho-Uehara, et al., 2016; Jin, et al., 2016) are involved in the development of awn and *Dep1*/*qPE9-1* (Huang, et al., 2009; Zhou, et al., 2009) is involved in panicle traits. We identified 3 ILs that showed drought tolerance at the seedling stage by PEG stimulation; a candidate gene *NAL1* was cloned. Mutant and overexpression experiments demonstrated that *NAL1* confers drought tolerance in seedlings, which also further showed that the *NAL1* gene was a multi-effect gene participating in the plant's multiple development processes. However, some details need to be further clarified, for example, which SNP was responsible for drought tolerance and whether the *NAL1* gene suffers from domestication selection. Those lines with favourable characteristics could also be transferred to good elite cultivars by MAS breeding. We believe that this set of ILs would have great value for rice breeding and functional genomics research.

Author contribution statement GQ developed molecular markers, identified genotypes and wrote the manuscript; HMN and SNL carried out phenotyping and backcross; YW performed map-based cloning and functional verification; ZZ conceived and designed the experiments, analysed data and revised the manuscript.

REFERENCES

1. Ahn, S., Anderson, J., Sorrells, M. & and Tanksley, S., 1993. Homoeologous relationships of rice, wheat and maize chromosomes.. *Mol Gen Genet,* Volume 241, pp. 483 - 490.
2. Ahn, S. & Tanksley, S., 1993. Comparative linkage maps of the rice and maize genomes.. *Proc Natl Acad Sci USA,* Volume 90, pp. 7980-7984.
3. Angeles, E. et al., 1998. Advanced breeding lines with resistance to rice tungro viruses. *International Rice Research Notes,* 23(1), pp. 17-18.

4. Asano, K. et al., 2011. Artifcial selection for a green revolution gene during japonica rice domestication.. *Proc Natl Acad Sci USA,* Volume 108, pp. 11034-11039.
5. Ashikari, M. et al., 2005. Cytokinin oxidase regulates rice grain production.. *Science,* Volume 309, p. 741–745.
6. Babu.R, Nair.SK, Prasanna.BM & and Gupta.HS, .., 2004. Integrating maker-assisted selection in crop breeding-Prospects and challenges.. *Current Science,* Volume 87, pp. 607-619.
7. Babu, P. et al., 2002. Annotation and BAC/PAC localization of nonredundant ESTs from drought-stressed seedlings of an indica rice.. *J Genet,* Volume 81, pp. 25-44.
8. Barakat, A., Carels, N. & and Bernardi, G., 1997. The distribution of genes in the genomes of Gramineae.. *Proc Natl Acad Sci USA,* Volume 94, pp. 6857-6861.
9. Bennetzen, J. & Freeling, M., 1993. Grasses as a single genetic systemgenome composition, collinearity and compatibility.. *Trends in Genetics,* Volume 9, pp. 259- 261.
10. Bessho-Uehara, et al., 2016. Loss of function at RAE2, a previously unidentifed EPFL, is required for awnlessness in cultivated Asian rice.. *Proc Natl Acad Sci USA,* Volume 113, pp. 8969-8974.

KHẢO SÁT VÀ PHÂN TÍCH DI TRUYỀN TÍNH TRẠNG NĂNG SUẤT VÀ KHẢ NĂNG CHỐNG CHỊU HẠN HÁN CỦA DÒNG LÚA *ORYZA RUFIPOGON* BẰNG CÁCH SỬ DỤNG CÁC QUẦN THỂ LAI NGƯỢC

Tóm tắt: *Lúa (Oryza sativa L.) là một trong những cây ngũ cốc chủ yếu trên thế giới, lúa là một trong những nguồn cung cấp carbohydrate chính cho gần một nửa dân số thế giới. Vì vậy, việc tăng năng suất lúa gạo để đáp ứng nhu cầu của con người là yêu cầu cấp thiết. Các đặc tính nông học như chiều cao cây, chiều dài bông, chiều dài lá cờ, số hạt trên bông, trọng lượng nghìn hạt, chiều dài, chiều rộng hạt gạo và khả năng chịu mặn là những yếu tố quyết định đến năng suất lúa. Gạo ngày nay có nguồn gốc từ lúa hoang phổ biến (Oryza rufpogon) và thừa hưởng những ưu điểm của nó. Tuy nhiên, trong quá trình thuần hóa cây lúa, một số đặc tính quý giá của lúa hoang như khả năng chống chịu stress sinh học và phi sinh học đã bị mất đi. Để sử dụng đầy đủ nguồn gen lúa hoang dã, chúng tôi đã xây dựng một tập hợp các dòng xâm nhập (IL) bằng cách sử dụng nguyên liệu lúa hoang phổ biến được thu nhập từ Lingshui, Trung Quốc. Một tập hợp các dấu hiệu phân tử InDel có độ phân giải cao với khoảng cách trung bình là 2,39 Mb được thiết kế để thực hiện lựa chọn và xác định các đặc điểm của phân khúc được hỗ trợ bởi dấu hiệu đánh dấu. Các IL chứa 213 dòng bao gồm 1,286 đoạn xâm nhập với chiều dài trung bình là 6,511 Mb, chiếm 93,59% nhiễm sắc thể của bố mẹ cho. Chúng tôi đã khảo sát đặc tính nông học của 213 dòng. Nhiều locus tính trạng định lượng (QTL) cũ liên quan đến chiều cao cây, tính trạng hạt giống và các đặc điểm khác đã xuất hiện trở lại trong IL của chúng tôi, chứng tỏ rằng hệ thống của chúng tôi đáng tin cậy. Hơn nữa, nhiều QTL mới đã được xác định. Một QTL liên quan đến khả năng chịu hạn nằm trên nhiễm sắc thể số 4 đã được xây dựng kỹ lưỡng. Bộ IL này cung cấp một nguồn tài nguyên mới để sử dụng các tính năng tốt của lúa hoang .*

Từ khóa: *Lúa, lập bản đồ QTL, Locus tính trạng định lượng.*