Identification of a Novel QTL for Panicle Length From Wild Rice (Oryza minuta) by Specific Locus Amplified Fragment Sequencing and High Density Genetic Mapping

Zhengzheng Zhu†, Xiaoqiong Li‡, Yu Wei, Sibin Guo* and Aihua Sha*

1 Hubei Collaborative Innovation Center for Grain Industry, Yangtze University, Jingzhou, China, 2 Guangxi Key Laboratory of Rice Genetics and Breeding, Rice Research Institute, Guangxi Academy of Agricultural Science, Nanning, China

Wild rice possesses a large number of valuable genes that have been lost or do not exist in cultivated rice. To exploit the desirable gene controlling panicle length (PL) in wild rice Oryza minuta, a recombinant inbred line (RIL) population was constructed that was derived from a cross between the long panicle introgression line K1561 with Oryza minuta segments and a short panicle accession G1025. Specific Locus Amplified Fragment (SLAF) sequencing technology was used to uncover single nucleotide polymorphisms (SNPs) and construct the high-density genetic linkage map. Using 201 RIL populations, a high-density genetic map was developed, and spanned 2781.76 cM with an average genetic distance 0.45 cM. The genetic map was composed of 5, 521 markers on 12 chromosomes. Based on this high-density genome map, quantitative trait loci (QTL) for PL were analyzed for 2 years under four environments. Seven QTLs were detected, which were distributed within chromosomes 4, 9, and 10, respectively. pl4.1 was detected twice, and pl10.1 was only detected once. Although pl9.1 was only detected once, it was very near pl9.2 in the genetic map which was detected three times. Thus, we speculate one major QTL exists in the region of pl9.1 and pl9.2 to control PL (temporarily referred to as pl9). pl9 is a potentially novel allele derived from Oryza minuta, and it can be used for genetic improvement of cultivar rice.

Keywords: introgression lines, yield related traits, RILs, SNPs, linkage map

INTRODUCTION

Rice is the most widely consumed staple food for a large part of the world’s human population, especially in Asia. To meet consumption demand of the growing world population, high-yield variety breeding is one of the major targets for modern breeders. In China, rice yield has been greatly improved during past decades owing to breeding of dwarfism, utilization of hybrid vigor, and cultivation and extension of super rice varieties. However, the narrow genetic basis of the super rice varieties has currently limited further improvements of yield. These yield improvements will likely only be achieved by expanding the available genetic resources.
Cultivated rice (*Oryza sativa* L.) is domesticated from wild rice, which has 22 species that are classified into 10 distinct types according to their genomes. Wild rice species contain many genes that are extremely valuable for genetic improvements of cultivated rice (Guo et al., 2009). Sourcing and using favorable wild rice genes may be an effective way to overcome the yield plateaus in cultivated rice.

*Oryza minuta* (*O. minuta*) is a tetraploid wild relative of cultivated rice, and possesses many resistance and yield improvement related source genes (Rahman et al., 2007). Rahman et al. (2007) developed F₂,₃ families by crossing the introgression line (IL) consisting of at least 14 segments of *O. minuta* with a Korean *japonica* rice variety, and performed QTL mapping for 16 agronomic traits. Of the 22 novel yield-related QTLs detected, 57% were derived from *O. minuta*.

In previous work, we constructed a set of ILs by distant hybridization and backcross, using IR24 (*O. sativa* L.) as the recipient and *O. minuta* as the donor; this was combined with embryo rescue and molecular marker assisted selection. QTLs for 12 agronomic traits such as grain weight, grain number, panicle length, panicle number etcetera were mapped using this set of ILs. A total of 28 QTLs for yield related traits were identified, of which, 46.4% of notable QTLs were from *O. minuta* (Guo, 2009).

Rice yield is determined by panicles per plant, spikes per panicle, seed setting rate and grain weight (Xing and Zhang, 2010). Panicle length (PL) correlates with rice yield, and thus, can be used as criteria for yield breeding (Zhang et al., 2015). A number of QTLs for PL have been mapped, and are located on almost all of the 12 rice chromosomes (Guo and Hong, 2010; Liu et al., 2011; Zhang et al., 2015; Bai et al., 2016; Sun et al., 2017). Some genes/QTLs responsible for PL regulation have been cloned. Most of them regulate the growth of primary or second branches and spikes by affecting meristematic activity; these include *Ghd7, Ghd7.1, Short Panicle 1, Dense and Erect Panicle1, Dense and Erect Panicle2* and *OsRAMOSA2* (Xue et al., 2008; Huang et al., 2009; Li S. et al., 2009; Li et al., 2010; Yan et al., 2013; Lu et al., 2017). Some genes control PL by regulating cell wall components and nutrients necessary for growth; examples of these are *Dense and Erect Panicle3, OsCD1*, and *OsARG* (Qiao et al., 2011; Luan et al., 2011; Ma et al., 2013). Other genes regulate PL by modulating hormone metabolism, such as *LP, OsPIN5b*, and *OsGRF4* (Li et al., 2011; Lu et al., 2015; Sun et al., 2016). In addition, *LON PANICLE1 (LP1)* encodes a protein of unknown function containing a Remorin_C domain (Liu et al., 2016). Two single nucleotide polymorphisms (SNPs) of LP1 lead to amino acid changes that affect PL.

In this study, we performed QTL analysis for PL using a recombinant inbred line (RIL) population derived from the cross of G1025 and K1561. K1561 is one of a set of 192 ILs that were developed from backcross progenies (BC₄F₂) derived from a cross between IR24 and *O. minuta* (Guo, 2009), which shows excellent agronomic traits such as long panicles and high 1000-grain weight. G1025 is an excellent restorer line widely used in Guangxi Province of China. Parents and RILs were planted in single plant and each with three lines. The field managements were conducted with regularly cultural methods. Ten main panicles were harvested to measure PL with ruler after mature. The mean values of ten plants were used as input data to identify PL QTLS (Supplementary Table S1).

## MATERIALS AND METHODS

### Plant Materials and Field Trials

The parental lines G1025 and K1561 along with the 201 RILs were planted in 2014 in Nanning (NN), China in both the early season (February to July) and the late season (July to November), and only in the early season in 2015. In addition, they were planted in 2015 in Wuhan (WH), China from May to October. K1561 is a long panicle variety developed from the backcross progenies of IR24 and *O. minuta* (Guo, 2009), and G1025 is a short panicle restorer line widely used in Guangxi Province of China. Parents and RILs were planted in single plant and each with three lines. The field managements were conducted with regularly cultural methods. Ten main panicles were harvested to measure PL with ruler after mature. The mean values of ten plants with two replicates were used as input data to identify PL QTLS.

### DNA Preparation

The DNA from two parental lines and RILs was extracted following the CTAB procedure (Murray and Thompson, 1980) and was treated with RNase. DNA quality was determined by electrophoretically resolving on a 1.0% agarose gel, and

| Sample ID | SLAF number | Total depth | Average depth |
|-----------|-------------|-------------|---------------|
| G1025     | 112,450     | 4,050,703   | 36.02         |
| K1561     | 110,284     | 3,912,556   | 35.48         |
| Average of RILs | 108,127 | 644,993 | 5.95 |

### TABLE 2 | Number of specific length amplified fragment (SLAF) markers on each chromosome.

| Chr ID | SALF number | Polymorphic SLAF |
|--------|-------------|------------------|
| Chr01  | 14,754      | 1,866            |
| Chr02  | 12,930      | 2,052            |
| Chr03  | 13,889      | 2,041            |
| Chr04  | 10,704      | 1,830            |
| Chr05  | 10,439      | 1,627            |
| Chr06  | 10,544      | 1,275            |
| Chr07  | 10,195      | 2,234            |
| Chr08  | 8,439       | 1,770            |
| Chr09  | 7,025       | 1,119            |
| Chr10  | 6,821       | 1,300            |
| Chr11  | 6,821       | 1,181            |
| Chr12  | 6,983       | 1,176            |
| Scaffold| 10,656      | 2,078            |
| Total  | 130,000     | 21,549           |
the concentrations were measured with a NanoDrop™
spectrophotometer.

**SLAF Sample Preparation and Sequencing**
Fresh leaves were harvested from a single plant of the parents and RILs, and genomic DNA was isolated for analysis using specific-locus amplified fragment sequencing (SLAF-seq) (Sun et al., 2013). Different combinations of restriction enzyme were tested using *in silico* digestion prediction to obtain expected SLAF tags per genome with even distribution in unique genomic regions. Two restriction enzymes (*Hae* III and *Hpy* 166II) were chosen for their uniform distribution and prevalence among simulations of fragment alignments to the reference genome of *Nipponbare* (*Oryza sativa L. japonica*). After digestion, different length fragments of genomic DNA were simulated *in silico*. *Arabidopsis thaliana* (ecotype Columbia) was used as the control genome to verify the restriction digest protocol for accuracy using SOAP software (Li R. et al., 2009).

Restriction digests were performed with 10 µg of genomic DNA from each parent line and RIL, and were followed by ligation reactions which included the addition of A to the 3’end and ligation with the Dual-index adapter. PCR was performed with the diluted restriction-ligation samples, and the PCR products were then purified with a Quick Spin column (Qiagen, Hilden, Germany) and electrophoretically resolved on a 2% (w/v) agarose gel. Fragments with correct lengths were isolated with a Gel Extraction Kit (Qiagen) and prepared for sequencing. Fragments of 264–314 bp were isolated to be used as SLAF tags. The fragments were sequenced with the Illumina HiSeq 2500 system (Illumina, Inc., San Diego, CA, United States) by Biomarker Technologies Corporation (Beijing, China).

**SLAF Sequence Comparison, Polymorphic Analysis, and Identification of the Associated Markers**

The SLAF-seq data were processed as per Sun et al. (2013). In brief, poor reads with a quality score < Q30 (<99.9% confidence) were filtered out. The pair-end reads from SLAF-seq were clustered according to sequence similarity, and the reads could be inferred using BLAT with one to one alignment (-tileSize = 10, -stepSize = 5). Identical reads were merged, and reads with > 90% similarity were placed into one SLAF locus (Sun et al., 2013). Alleles were defined in each SLAF locus with minor allele frequency (MAF) evaluation.

In order to ensure a good quality genetic map, SLAFs with a sequence depth of under 10× parents, below 30% complete degrees, serious partial separation (p < 0.05), and heterozygous loci of the two parents were removed. Polymorphic SLAFs with more than three SNPs were removed. A SLAF-tag with the same genotype as the female (G1025) or male (K1561) parent was designated as a SLAF marker and used.
for the subsequent association analysis. Diploid species have up to 4 base types, and a polymorphic locus between the parents will have a minimum of 2 base types. Those polymorphic SLAF markers were assorted into eight segregation patterns as following: ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc, and cc × ab. A heterozygous locus with different base types in both parents is indicated with ab × cd.

The results show that the locus is heterozygous in both parents, but the parents share one identical base. hk indicates that the locus is heterozygous in the female parent but homzygous in the male parent, and the parents share one identical base. cc × ab indicates that the locus is heterozygous in the female parent but homzygous in the male parent. aa × bb indicates that the locus is heterozygous in the male parent, and the parent base type is inconsistent. cc × ab indicates that the locus is heterozygous in the male parent but homzygous in the female parent, and the parent base type is inconsistent. The segregation patterns of aa × bb were used for map construction because the RIL mapping population were derived from two homozygous parents with a genotype of aa or bb.

**Linkage Map Construction and QTL Analysis**

JoinMap4.0 software (Van Ooijen, 2006) was used to perform linkage analysis with all the genotype data from the RIL mapping population. The SLAFs can be divided into 12 linkage groups (LGs) with positions of the reference genome. The MLOD value between two adjacent markers was determined (Vision et al., 2000), and SLAFs with MLOD values less than 5 were filtered out. Marker HighMap software (Liu et al., 2014) was used to analyze the linear array of markers in each LG, and the genetic distances between pairs of adjacent markers were estimated. QTLs analysis was conducted using the Internal Mapping method with R/qtl software, and 95% Bayes credible interval method (Sen and Churchill, 2001) was used to calculate the confidence intervals of QTLs. The LOD score threshold for significance was determined using 1,000 permutations with R/qtl.

**RESULTS**

**SLAF-Seq Data and SLAF Markers**

A total of 302.01 Mbp of raw data were generated by DNA high-throughput sequencing. The average of Q30 ratios was 91.92%, and the average of guanine-cytosine (GC) contents was 42.15%. There were 9,353,528 reads from the female parent, 9,125,218 reads from the male parent, and the average number of reads from the RIL population was 1,410,603.

The numbers of SLAFs in the male and female parents were 112,450 and 110,284, respectively, correspondingly, the male and female parents had average sequencing depths of 36.02- and 35.48-fold, respectively. The average number of SLAFs was 108,127 in the RIL population, and the coverage was an average of 5.95-fold (Table 1). Of the 130,000 high-quality SLAFs detected, 21,549 (or 16.58%) were polymorphic (Table 2). A total of 20,443 of the 21,549 polymorphic SLAFs were classified into eight segregation patterns (Figure 1). The RIL population was derived from the cross of two parents with an aa or bb homozygous genotype; thus, only the RIL plants with an aa × bb segregation pattern were used for high-density genetic linkage map construction, and a total of 18,360 markers were of this type. Of these 18,360 markers, the high-density genetic map was constructed with 5,960 markers which were homozygous in the two parents and had sequence depth of more than 10-fold with over 70% integrity of SLAF tags.

**Genetic Map Construction and Its Basis Information**

After linkage analysis, 5,521 of the 5,960 SLAF markers were mapped on the genetic map. On average, these markers had 53.32-fold sequence depths in G1025 (the female parent), 53.20-fold in K1561 (the male parent), and 8.65-fold in each RIL line. The final genetic map included 5,960 markers of which 1,410,603.

**TABLE 4 | Quantitative trait loci (QTL) analysis of rice panicle length in RILs.**

| Year | Season | QTL | Chr | Marker interval | Interval (cM) | LOD | Additive effect | r² (%) | The donor parent |
|------|--------|-----|-----|---------------|---------------|-----|----------------|-------|-----------------|
| 2014 | NN-E   | P24.1| 4   | Marker2465781-Marker2467059 | 160.40-160.90 | 9.86 | −1.08 | 14.84 | G1025           |
| 2014 | NN-L   | p91.1| 9   | Marker768445-Marker797853 | 152.74-153.60 | 4.60 | −1.13 | 7.67 | G1025           |

*NN-E, during the early season in Nanning; NN-L, during the late season in Nanning; WH, in Wuhan.*
The largest gap on this map was 14.50 cM, which is located in LG4.

**Evaluation of the Genetic Map**

The quality of the genetic map was evaluated by drawing heat maps for the 5,521 SLAF markers using pair-wise recombination values. Adjacent markers in the heat maps have low recombination frequencies, which indicate that the constructed linkage map has good accuracy ([Supplementary Figure S3](#)).

The haplotype map is reflective of genotyping errors that have occurred by crossover events in an advanced population. According to the haplotype map, most of the recombination blocks were defined with under 0.1% heterozygosity, less than 0.1% missing, and an even distribution of markers on each chromosome ([Supplementary Figure S4](#)).

**FIGURE 3** | Quantitative trait loci (QTL) analysis with panicle length (PL) using the high-density genetic map. (A) QTL analysis of PL in the early season of Nanning (NN) in 2014; (B) QTL analysis of PL in the late season of Nanning (NN) in 2014; (C) QTL analysis of PL in NN in 2015; (D) QTL analysis of PL in Wuhan (WH) in 2015. The horizontal line on the chart represents the LOD threshold.
Evaluation of collinearity between the genetic map and the rice reference genome was performed by mapping SLAF markers to the rice reference genome, and curves between physical distances and genetic distances were evident for most chromosomes except Chr8 (Supplementary Figure S5). The high collinearity suggests that the 5,521 SLAF markers were accurately mapped to the 12 chromosomes.

**QTL Analysis Based on the High-Density Genetic Map**

Phenotypic data of the two parents and RILs planted in NN and WH in 2014 and 2015 are presented in Figure 2 and Supplementary Table S1. A total of seven QTLs for PL were detected in RILs (Table 4 and Figures 3, 4), and mapped to unique positions. The proportion of phenotypic variance explained by a single QTL ($r^2$) ranged from 7.67% to 26.99% and LOD scores were from 4.60 to 16.61. The two QTLs located on chromosomes 4 accounted for 14.84% and 14.01% of the phenotypic variation, respectively. Four QTLs were located on chromosome 9, and accounted for 24.46%, 13.63%, 26.99%, and 13.15% of the phenotypic variation, respectively. One QTL was located on chromosome 10, and accounted for 7.67% of phenotypic variation. $pl4.1$ was repeatedly detected early in the season in Nanning in 2014 and 2015, and $pl10.1$ was only detected in Wuhan in 2015. $pl9.1$ was only detected late in the season in

**FIGURE 4** Genetic linkage map with locations of QTLs on chromosomes 4, 9, and 10.
season of Nanning, and \textit{pl9.2} could be detected repeatedly early in the season in Nanning in 2014, 2015, and in Wuhan in 2015. Furthermore, \textit{pl9.2} had a larger LOD and it was quite nearby \textit{pl9.1} in the genetic map. We cannot judge whether \textit{pl9.1} and \textit{pl9.2} are the same QTL or two distinguishing QTLs at present, so we temporarily designated \textit{pl9.1} and \textit{pl9.2} as a single major candidate QTL that controls PL (referred as to \textit{PL9}).

In order to evaluate whether the phenotypes were affected by the detected QTLs, we selected 13 out of the 201 RILs to compare their phenotypes with their genotypes. The 13 RILs were homozygous at the loci of \textit{pl4.1}, \textit{pl4.2}, and \textit{pl10} for the genotypes of G1025 (date not shown), but they were homozygous at the loci of \textit{pl9.1} and \textit{pl9.2} for the genotypes of either G1025 or K1561 (Figure 5). The PL phenotypes of the 13 RILs were consistent with their respective genotypes (Table 5 and Figure 5).

### DISCUSSION

By combining locus-specific amplification and high-throughput sequencing, SLAF-seq is an effective technology for \textit{de novo} SNP discovery and large-scale genotyping. It has distinguishing characteristics of deep sequencing, but employs a reduced representation strategy, pre-designed reduced representation scheme, and double barcodes. Therefore, it can ensure genotyping accuracy, reduce sequencing costs, reasonably and reliably predict fragmentation efficiency, and allow for the simultaneous genotyping of large populations (Sun et al., 2013). In addition, marker development with SLAF-seq was less time-consuming, less expensive, and more efficient relative to other conventional methods (Zhu et al., 2016).

\textit{O. minuta} possess a large number of desirable source genes which can be used for yield improvement. In a previous study, a set of ILs were constructed by crossing IR24 and \textit{O. minuta}, and an excellent IL, K1561, was selected from this set (Guo, 2009). In order to further exploit and utilize the favorable genes of \textit{O. minuta} in this study, we conducted SLAF-seq to develop SNP markers and to identify QTLs for PL by using the population of...
PL is an important yield-related trait that determines the number of spikelets. Based on the high-density genetic linkage map and the phenotypes of G1025, K1561, and the RILs, the seven detected QTLs were distributed on chromosomes 4, 9, and 10. *pl4.1* was repeatedly detected in the early seasons of NN in 2014 and 2015, but was absent in the late seasons of NN in 2014 and WH in 2015. *pl10.1* was only detected in WH in 2015. *pl9.1* was only detected in the late season of NN, but it is very close to *pl9.2* in the genetic map; *pl9.2* was stably detected during the early seasons of NN in 2014 and 2015, and in WH in 2015 with a higher LOD score. The region containing *pl9.1* and *pl9.2* has been temporarily denoted as a major QTL candidate that controls PL (referred to as *PL9*), because we cannot judge whether *pl9.1* and *pl9.2* are the same QTL or two separate QTLs at present. The phenotypes were affected by genotypes at *pl9.1* and *pl9.2* by analysis of 13 RILs, indicating that *Pl9* most likely controls PL.

Up until now, many QTLs for PL have been mapped on nearly all of the twelve chromosomes, however, only a few QTLs were located on chromosome 9. *DEP1* is a pleiotropic QTL responsible for dense panicle, high grain number per panicle, and erect panicle, and was defined between the markers RM3700 and RM7424. *DEP1* (*LOC_Os09g26999*) has been cloned and encodes a protein containing a phosphatidylylthanolamine-binding protein (PEBP)-like domain (Huang et al., 2009). *qPL-9-1* and *qPL-9-2* were mapped between markers RM6570-RM5652 and RM5652-RM410 by using 254 RILs derived from a cross between cultivars Xiushui 79 and C Bao, respectively (Guo and Hong, 2010). The *qPL-9-2* (*LP*) was further narrowed to a 90 kb region of DNA between markers L04 and RM7289, and the candidate gene *LOC_Os09g28300* has been cloned (Liu et al., 2016). *MPL9*, was mapped between markers RM215 and RM1013 by using RILs derived from the cross of Tqing and Minghui 63, but it was detected during only 1 year (Liu et al., 2011). One QTL for PL was mapped between markers S9058.3-RM7175 by using 178 F$_2$ RILs from a cross of *japonica* rice line “SNUSG1” and *indica* rice line “Milyang23” (Lim et al., 2014). The mapping or cloning of these QTLs or genes was conducted based on the population derived from the crosses of cultivated varieties.

The mapping population in this study was derived from the cross of a restorer line G1025 and an IL with the segments of *O. minuta* K1561. One major QTL, *pl9*, was detected within 304 Kb between SLAF Markers 768,445 and 775,977. It is a potentially novel allele from *O. minuta*. *pl9* is within the region of rice genome SSR markers RM24780-RM215 (International Rice Genome Sequencing Project, 2005), which is nearby but not the same as *MP9* (Liu et al., 2011), and it was different with other QTLs or genes that have been identified on the chromosome 9. Other QTLs for PL have been detected in *O. minuta*, and include *pl6, pl7, pl8* identified on chromosome 6, 7, 8, respectively (Rahman et al., 2007). In addition, QTLs for PL have also been identified in other wild rice species. For example, the chromosome 6 located *spd6* is responsible for small panicles and dwarfness, and was identified in *Oryza rufipogon* (Shan et al., 2009). Results such as these indicate that potential alleles controlling PL might exist in the genome of wild rice.

**CONCLUSION**

In conclusion, we identified a novel QTL for PL (*pl9*) from *O. minuta* by construction of the high-density SNP map and QTL association analysis. Refined mapping of the *pl9* locus will be conducted to identify the candidate genes. On the other hand, the *pl9* allele from *O. minuta* can also be directly used for genetic improvement of cultivar rice.

**AUTHOR CONTRIBUTIONS**

ZZ extracted DNA, prepared the library, and wrote the manuscript. XL performed SNP data analyses and wrote the manuscript. YW constructed the linkage map. SG and AS designed the experiments and supervised the study.

**FUNDING**

This work was supported by grants from the National High Technology Research and Development of China (No. 2014AA10A603), Guangxi Natural Science Foundation (No. 2017GXNSFDA198039) and Guangxi Key Laboratory of Rice Genetics and Breeding Open Foundation (No. 160-380-16-1).

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01492/full#supplementary-material

**FIGURE S1** High-density genetic map constructed by 5521 SLAF Markers.

**FIGURE S2** Distribution of Markers on chromosomes that were applied to genetic map construction.

**FIGURE S3** Heat map of the high-density genetic map. Each cell represents the recombination rate of two markers. Yellow, red, and purple represent the recombination rate ranging from lower to higher.

**FIGURE S4** Haplotype map of the genetic map. Blue represents K1561, red represents G1025, green indicates heterozygous type, and grey represents deletions.

**FIGURE S5** The collinearity of 12 chromosomes with the rice reference genome. The x-axis indicates the genetic distance of rice chromosomes, and the y-axis represents the linear order of the physical position in the rice genome. All 5521 SLAF markers in these chromosomes are plotted as dots on the Figure. Different colors indicate different chromosomes.

**TABLE S1** Panicle length data of two parents and RILs.
