Transcriptome Analysis and Functional Identification of Xa13 and Pi-ta Orthologs in Oryza granulata

Rui Yang, Jing Li, Hui Zhang, Fei Yang, Zhigang Wu, Xiaoxuan Zhuo, Xingyu An, Zaiquan Cheng, Qianchun Zeng, and Qiong Luo*

ABSTRACT Oryza granulata Nees & Arn. ex Watt, a perennial wild rice species with a GG genome, preserves many important genes for cultivated rice (Oryza sativa L.) improvement. At present, however, no genetic resource is available for studying O. granulata. Here, we report 91,562 high-quality transcripts of O. granulata assembled de novo. Moreover, comparative transcriptome analysis revealed that 131 single-copy orthologous pairs shared by O. granulata and Oryza meyeriana (Zoll. & Moritzi) Baill. that may have undergone adaptive evolution. We performed an analysis of the genes potentially involved in plant–pathogen interactions to explore the molecular basis of disease resistance, and isolated the full-length cDNAs of Xa13 (OgXa13) and Pi-ta (OgPi-ta) orthologs from O. granulata. The overexpression of OgPi-ta in Nipponbare and functional characterization showed enhanced the resistance of transgenic Nipponbare to rice blast resulting from the presence of the OgPi-ta gene. OgPi-ta, an alternatively spliced transcript of the Pi-ta blast resistance gene in O. sativa, encodes a 1024-amino acid polypeptide with a C-terminal thioredoxin domain. This study provides an important resource for functional and evolutionary studies of the genus Oryza.

RICE is one of the oldest domesticated crops and is a staple food for over half of the world’s population (Hajira et al., 2016). However, in recent years, rice production has been seriously affected by biotic and abiotic stresses caused by adverse climatic changes and the breakdown of resistance genes in elite cultivars (Normile, 2008; Reddy et al., 2017). Genetic resources can be used to improve the resistance or tolerance of plant species to biotic and abiotic stresses; however, these are extremely limited in the cultivated rice gene pool. The Oryza genus consists of the AA, BB, BBCC, CC, CCDD, EE, FF, GG, HHJJ, and HHKK genomes and includes 22 wild species in addition to two cultivated species (O. sativa and Oryza glaberrima Steud.) (Jena, 2010; Khush, 1997; Lu et
al., 2009). Over the course of long-term competition for survival and natural selection under some extreme environmental conditions, wild rice types have developed agronomically important genes for improving resistance to plant hopper, blast, and bacterial blight (Bon et al., 2011; Li et al., 2011; Lv et al., 2013; Qu et al., 2006; Yan et al., 2004), as well as tolerance to drought, soil acidity, and Fe and Al toxicity (Brar and Khush, 1997; Jena, 2010). They are largely untapped resources. There is an urgent need to broaden the gene pool of cultivated rice by transferring valuable genes from wild species into modern cultivars to enhance their resistance or tolerance to biotic and abiotic stresses and thus eventually increase the yield potential of modern cultivars (Jena, 2010).

Oryza granulata and O. meyeriana, two diploid wild rice species in the O. meyeriana complex defined as having the GG genome, have been categorized as basal lineages of Oryza (Gong et al., 2000; Jena, 2010; Sharma and Shastry, 1965; Vaughan, 1994; Zou et al., 2008), and possess the second largest genome (~882 Mb) of the diploid species in the genus Oryza (Ammiraju et al., 2007). Both linages contain valuable genes that can confer resistance to bacterial blight, blast, and hopper and tolerance to shade and drought (Bon et al., 2011; Brar and Khush, 1997; Jena, 2010; Yan et al., 2004). As early as 1991, hybrids were produced from the cross between O. sativa and O. meyeriana to develop bacterial-blight-resistant rice varieties (Brar and Khush, 1991). Nevertheless, to date, none of the favorable genes have been successfully transferred from wild species of the GG genome into cultivated rice varieties, mainly because of highly incompatibility and sterility issues (Jena, 2010; Zou et al., 2008).

In the past few decades, the rapid development of sequencing technology has led to the availability of both genome and transcriptome data for various organisms. Transcriptome sequencing is an excellent tool for identifying genes associated with important agronomic traits, efficiently developing genetic markers, studying gene expression, and resolving gene networks. The technique is particularly useful for investigating organisms with little available genomic information (Wang et al., 2009). Although a transcriptomic database was recently established for O. meyeriana (He et al., 2015), there are still only a few genomic resources available for studying wild species of the GG genome, compared with numerous resources for the AA genome. Moreover, closely related species can exhibit considerable differences in their genome size, gene number, and collinearity (Ammiraju et al., 2008; Hu et al., 2011). Oryza granulata and O. meyeriana have a long history of natural evolution in different habitats; exhibit many distinct characteristics in relation to grain length and width, flag-leaf width and length, culm length, and leaf blade size (Song et al., 1995; Tateoka, 1962); and are expected to carry unique genetic resources that could improve cultivated rice. Oryza granulata is placed on the basal lineage in Oryza (Zou et al., 2008) and is available in the IRRI gene bank as a representative germplasm of the GG genome. Thus, RNA-sequencing and a comprehensive survey of the expressed sequence tags in O. granulata, as well as comparative transcriptome analyses of O. granulata and O. meyeriana, will provide more valuable genomic information relating to the O. meyeriana complex. This will allow us to take full advantage of the available rice genome resources for biological research and rice breeding, and help researchers fully elucidate the changes that have occurred in the GG genome, as well as the underlying mechanisms of Oryza genome evolution.

In the present study, de novo transcriptome assembly of O. granulata was performed on the Illumina paired-end sequencing platform (GENE DENOVO, Guangzhou, China). We obtained 91,562 high-quality transcripts from five cDNA libraries prepared from the roots, stems, sheaths, panicles, and leaves of O. granulata. Multiple plant resistance-related genes were identified, including 324 genes enriched in the plant–pathogen interaction pathway, 2597 nucleotide-binding-site (NBS)-coding genes, 1166 receptor-like protein (RLP)-coding genes, and 209 receptor-like kinase (RLK)-coding genes. The xa13 (a recessive bacterial blight resistant gene) and Pi-ta (a dominant blast resistant gene) homologs (Bryan et al., 2000; Chu et al., 2006a, 2006b) of O. granulata were cloned and the function of the OgPi-ta gene was investigated by heterologously expressing it in the susceptible Nipponbare cultivar followed by pathogen infection. Finally, comparative transcriptome analysis was conducted to explore the genomic differences between O. granulata and O. meyeriana at the transcriptional level. This study will greatly benefit gene mining and genome assembly of the O. meyeriana complex and provide an important resource for functional and evolutionary studies of plant species in the genus Oryza.

MATERIALS AND METHODS

Plant Materials

Two wild rice species, O. granulata and O. meyeriana, and a susceptible rice variety, Nipponbare, were used in this study. Orz ya granulata was obtained from IRRI and O. meyeriana was collected from its native habitat in Xishuangbanna, Yunnan Province, China, in 2013. The rice cultivar Nipponbare, a variety that is susceptible to rice blast and bacterial blight, is maintained in our laboratory.

RNA Extraction

For RNA extraction, the roots, stems, sheaths, panicles, and leaves of O. granulata and mixture of panicle and leaf samples of O. meyeriana were harvested at the booting stage. Total RNA from each sample was isolated with a TransZol UP reagent (TransGen, Beijing, China) following the manufacturer’s instructions.

RNA Sequencing and Assembly

The extracted RNA samples were used for the cDNA synthesis. Polyadenylation mRNA was isolated with oligo-dT beads (Qiagen, Düsseldorf, Germany). All mRNA was broken into short fragments (200 nucleotides) in a
Functional Annotation and Classification

All unique assembled gene sequences were searched for within well-known databases, including the National Center for Biotechnology Information’s nonredundant protein sequences (NR) (http://www.ncbi.nlm.nih.gov/, accessed 14 June 2018), the Eukaryotic Orthologous Groups (KOG) database (ftp://ftp.ncbi.nih.gov/pub/COG/KOG/kyva, accessed 18 June 2018) (Roman et al., 2003), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/, accessed 14 June 2018) (Kanehisa et al., 2008), and SwissProt (a manually annotated and reviewed protein sequence database) (http://www.expasy.ch/sprot/, accessed 14 June 2018) (Magrane and UniProt Consortium, 2011), by using the BLASTX algorithm (Altschul et al., 1997) with an E-value cutoff of <10^{-5}. The protein sequences found in the databases with the highest similarity scores were used to obtain the functional annotations of related unigenes. The Blast2GO (Conesa et al., 2005) program was used to obtain the annotation results for unigenes in Gene Ontology (http://www.geneontology.org/, accessed 14 June 2018) (Botstein et al., 2000), a database containing a series of terms describing the function of genes based on three categories: molecular functions, biological processes, and cellular compounds. The statistical tool WEGO (http://wego.genomics.org.cn, accessed 18 June 2018) was used to classify the annotation of the unigenes in the Gene Ontology database (Ye et al., 2006).

Protein Coding Sequence Prediction

The unigene annotation results were used to determine whether a particular coding sequence (5’–3’) was present by translating the coding sequences into amino acid sequences. The unigene sequences that did not align with any database were predicted with ESTScan (http://myhits.isb-sib.ch/cgi-bin/estscan, accessed 14 June 2018) (Christian et al., 1999), as well as their nucleotide sequence (5’–3’) direction and the amino acid sequence of the predicted coding region.

Cloning of Full-Length cDNA

Mixed panicle, leaf, stem, and root samples were collected from O. granulata and O. meyeriana at the booting stage for RNA extraction with the TransZol UP reagent (TransGen). 3’- and 5’- rapid amplification of cDNA ends (RACE)-PCR was performed with the SMARTer RACE 5’/3’ Kit (TaKaRa, Dalian, China), according to the manufacturer’s instructions. For 3’ RACE, the RNA was reverse-transcribed with the 3’ coding sequence primer and treated with SMARTScribe reverse transcriptase. Polymerase chain reaction was performed with the 3’ Gene Special Primer and 3’ Universal Primer Mix provided by the SMARTer RACE 5’/3’ Kit. For 5’ RACE, the RNA was reverse-transcribed using 5’ coding sequence primer and treated with SMARTer II A oligonucleotide and SMARTScribe reverse transcriptase. Polymerase chain reaction was performed using 5’ RACE Gene Special Primer and 5’ Universal Primer Mix. The products were cloned into a pEASY-Blunt simple vector (TransGen) and sequenced. The primers are listed in Supplemental Table S1.

Gene Cloning and Genome Structure Analysis

Genome DNA was extracted from fresh leaves of O. granulata. The primers of OgXa13 (forward and reverse) and OgPi-ta (forward and reverse) were designed based on 3’ or 5’ RACE-PCR product sequences and previously published data (Yang et al., 2006) to amplify the genomic DNA fragments of OgXa13 and OgPi-ta from O. granulata. The primers are listed in Supplemental Table S1. All the cDNA-genomic alignments were conducted with Vector NTI version 11.5 (Thermo Fisher Scientific, Waltham, MA).

Analyses of OgPi-ta Function

A 3075-bp cDNA fragment was amplified with OgPi-ta (forward and reverse) primers and cloned into the pCAMBIA1300-OXHLL vector by homologous recombination to generate the overexpression construct pCAMBIA1300-OXHLL-OgPi-ta (OgPi-ta-ox). After sequence verification,
the final construct of OgPi-ta-ox was introduced into the susceptible rice variety Nipponbare via *Agrobacterium*-mediated transformation, as described previously (McGurl et al., 1994). Transformants were selected according to a previously described protocol (Chen et al., 2010). Positive transgenic plants were initially detected on the basis of the presence of the marker gene for hygromycin in the vector, and a 598-bp fragment was amplified with hygromycin forward and reverse primers. To confirm positive transgenic plants accurately, a 220-bp fragment was amplified with the gene-specific primers OgPi-ta-G (forward and reverse) for *OgPi-ta* transgenic plants. The primers are listed in Supplemental Table S1.

**Gene Expression Analysis**

Gene expression analysis was performed by quantitative real-time PCR. The total RNA of each sample was obtained as described above. cDNA was synthesized with the ReverTra Ace qPCR RT Master Mix kit (TOYOBO, Osaka, Japan) and used as the template for quantitative real-time PCR, which was performed with GoTaq qPCR Master Mix (Promega, Madison, WI) on a Bio-Rad CFX96 Real-Time System (BioRad, Berkeley, CA), with β-actin as the reference gene. The primer sequences are listed in Supplemental Table S1.

**Pathogen Inoculation and Disease Evaluation**

For evaluation of blast disease, seedlings at three to four leaf stage were spray-inoculated with isolate 1817–2 with spore suspensions of 10⁵ spores mL⁻¹ in a dew growth chamber, as previously described (Li et al., 2015). The *Magnaporthe oryzae* strain 1817–2 containing the avirulence gene *AVR-Pita* was kindly provided by Prof. Cailin Lei, Institute of Plant Protection, Chinese Academy of Agricultural Sciences. To ensure successful inoculation, seedlings were also transplanted into a greenhouse and injection-inoculated at the tillering to booting stage. Disease evaluation was performed 7 d after spraying or injection, for which lesion types on leaves were observed and scored from 0 (resistant) to 5 (susceptible), according to a standard reference scale (Silue et al., 1992; IRRI, 2002): 0, no disease lesions observed; 1, small brown specks of pinpoint size with a diameter less than 0.5 mm; 2, small, roundish to slightly elongated necrotic spots, ~1 to 2 mm in diameter with a distinct brown margin or yellow halo; 3, narrow or slightly elliptical lesions, 1 to 2 mm in breadth, more than 3 mm long with a brown margin; 4, typically broad spindle-shaped lesions with yellow, brown, or purple margins, ~3 to 4 mm in diameter with a significant number of lesions are on the upper leaves; 5, disease lesions with diameters over 4 mm.

**RESULTS**

**Sequence and De Novo Assembly of the *O. granulata* Transcriptome**

Five cDNA libraries were generated from the roots, stems, leaves, panicles, and sheaths of *O. granulata*, and assembled *de novo* with Illumina paired-end deep sequencing technology. We obtained 412,084,330 clean reads with a total of 59,667,208,274 nucleotides (Supplemental Table S2). All high-quality reads from each of the five samples were combined for transcriptome assembly with Trinity software (Grabherr et al., 2011). A total of 91,562 unigenes were obtained, which included a total of 88,913,258 nucleotides, had an average length of 971 bp, and for which the shortest sequence length of 50% of the contigs was 1758 bp (Table 1); 26,902 unigenes were longer than 1000 bp. The size distributions of the unigenes are presented in Supplemental Fig. S1.

**Functional Annotation and Classification of *O. granulata* Unigenes**

Upon obtaining sequences for the assembled unigenes, we searched each sequence against a number of public databases (KOG, KEGG, SwissProt, and NR) by using the BLASTX algorithm with an E-value cutoff of <10⁻⁵. We found that 46,393 unigenes had at least one significant match to an existing gene model, 45,982 (99.11%) of which shared significant similarity with known proteins in NR, 29,148 (62.83%) in Swiss-Prot, 24,314 (52.41%) in KOG, and 16,223 (34.97) in KEGG (Fig. 1A). We then counted the proportional distribution of unigenes with an E-value of <10⁻⁵ in the NR and SwissProt databases. Unigenes with an E-value of <10⁻⁵ constituted the largest proportion of BLAST hits in NR (56.47%) (Supplemental Fig. S2A) and Swiss-Prot (47.83%) (Supplemental Fig. S2B). Further analysis of sequence homology among different species showed that 26,133 (56.83%) unigenes were best matched to sequences from *O. sativa*, 8566 (16.83%) from *Oryza brachyantha* A.Chev. & Roehr., and 1659 (3.61%) from *Setaria italicica* (L.) P.Beauv. (Supplemental Fig. S2C).

Among the 46,393 annotated unigenes, 34,511 were assigned to 48 Gene Ontology functional terms (Fig. 1B). In the cellular component category, cell (22,956; 27.42%), cell part (22,954; 27.42%), and organelle (21,090; 25.19%) were the most common matches. In the biological process category, metabolic process (18,496; 26.19%), cellular process (17,383; 24.62%), and single-organism process (11,273; 15.96%) were the main matches. In the molecular function category, binding protein-encoding genes (18,010; 46.88%) and catalytic activity related protein-encoding genes (16,254; 42.31%) were the largest proportion, representing 89.19% of Gene Ontology terms.

---

**Table 1. Summary of the assembly data for *O. granulata***

<table>
<thead>
<tr>
<th>Number of genes</th>
<th>GC content</th>
<th>N50</th>
<th>Max. length</th>
<th>Min. length</th>
<th>Average length</th>
<th>Total assembled bases (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>91,562</td>
<td>48.76%</td>
<td>1758</td>
<td>16,688</td>
<td>201</td>
<td>971</td>
<td>88,913,258</td>
</tr>
</tbody>
</table>
Fig. 1. The annotation and functional classification of assembled unigenes. (A) Venn diagram of the annotation results of unigenes searched against four protein databases. (B) Gene Ontology (GO) classifications. (C) Histogram of Cluster of Orthologous Groups (COG) classifications. (D) The top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms. Classifications are based on the categories: metabolism (red), genetic information processing (orange), environmental information processing (green), cellular processes (blue), and organismal systems (purple).
Next, 24,314 of the 45,982 unigenes with NR hits were assigned to 25 KOG categories (Fig. 1C). General function prediction was the largest group (8214; 33.78%), followed by posttranslational modification (5132; 21.11%), and signal transduction mechanisms (4329; 17.80%). As previously reported, lipid transport and metabolism play an essential role in the biogenesis of the photosynthetic apparatus in developing chloroplasts (Benning, 2009). Chloroplasts can generate reactive oxygen species including superoxide anions, hydrogen peroxide, hydroxyl radicals, and singlet oxygen, which are important for stress responses (Zhu, 2016). In O. granulata, 1363 unigenes involved in lipid transport and metabolisms were identified, which may contribute to drought and shade tolerance in O. granulata.

**Functional Classification via the KEGG Pathway**

To further understand the transcriptome data, 91,562 unigenes were checked against referential canonical pathways in KEGG with a threshold of $E < 10^{-5}$ and 9168 (10.01%); unigenes with significant matches in the database were assigned to five major categories involving 128 KEGG pathways. Among the top 20 KEGG pathway terms, the most highly represented term was ribosome (755; 8.24%), followed by C metabolism (568; 6.20%), and protein processing in the endoplasmic reticulum (500; 5.45%) (Fig. 1D). In particular, 324 unigenes were enriched in the plant–pathogen interaction KEGG pathway in O. granulata (Fig. 1D). These results will help us identify the genes involved in these pathways.

**Analysis of the Plant–Pathogen Interaction Pathway in O. granulata**

_Oryza granulata_ is resistant to bacterial blight and rice blast (Jena, 2010; Yan et al., 2004). During coevolution with pathogens, plants have developed a set of defense strategies against various pathogens, with an innate immune machinery consisting of pathogen-associated molecular pattern-triggered immunity (PTI), conferred by pattern recognition receptors and pathogen-effector-triggered immunity (ETI) conferred by race-specific resistance (R) proteins (Spoel and Dong, 2012). We found that 324 unigenes were enriched in the plant–pathogen interaction KEGG pathway in _O. granulata_ (Fig. 1D; Supplemental Table S3). The increase in intracellular Ca$^{2+}$ concentration is an important trigger for the production of reactive oxygen species and NO, ultimately leading to a series of defense responses, including localized programmed cell death or the hypersensitive response, cell wall reinforcement, and stomata closure. In _O. granulata_, 151 unigenes encoding proteins involved in this process, including the cyclic nucleotide-gated channel, Ca-binding protein, Ca-dependent protein kinase, respiratory burst oxidase and nitric oxide synthase (Fig. 2; Supplemental Table S3). The activation of flagellin-sensing 2 (FLS2) triggers tandem signaling of mitogen-activated protein kinase cascades, which trigger defense-related genes to induce antimicrobial compounds, such as phytoalexin. In _O. granulata_, 31 unigenes encoding proteins were involved in this process, including FLS2, mitogen-activated protein kinase kinase kinase 1, mitogen-activated protein kinase kinase 1/2 and mitogen-activated protein kinase kinase 4/5, WRKY transcription factor 22/29 and WRKY transcription factor 25/33, glycerol kinase, and pathogenesis-related protein 1 (Fig. 2; Supplemental Table S3).

Pathogen effectors from diverse kingdoms are recognized by R proteins and activate defense responses, resulting in disease resistance and usually an associated hypersensitive cell death response at the infection site. In _O. granulata_, 142 unigenes encoding proteins are involved in the ETI process, including PRM1-interacting protein 4, disease-resistance protein RPM1, disease-resistance protein RPS2, serine or threonine protein kinase PBS1, the suppressor of the G2 allele of skp1 (SGT1), enhanced disease susceptibility 1, Heat Shock Protein 90, Pto-interacting protein 1, and Mla12 resistance protein RAR1 (Fig. 2; Supplemental Table S3).

**Identification of RLK, RLP, and NBS-Leucine-Rich Repeat Genes**

Receptor-like kinases and RLPs play important roles as pattern recognition receptors and regulators of PTI (Boller and Felix, 2009; Gomez-Gomez and Boller, 2000; Song et al., 1995). Nucleotide-binding-site-leucine-rich-repeat (NBS-LRR) proteins are the largest class of R proteins in plants (Dangl and Jones, 2001). We performed BLASTp analysis of all predicted amino acid sequences of _O. granulata_ against the PRGdb database (http://prgdb.crg.eu/wiki/Main_Page, accessed 15 June 2018), and identified 1166 RLP-encoding unigenes, 209 RLK-encoding unigenes, and 2597 NBS-encoding unigenes. These included 2597 NBS-encoding unigenes, 788 NBS-LRR, 441 coiled-coil NBS-LRRs, 605 interleukin-1 receptor NBS-LRRs, 113 coiled-coil NBS, and 650 NBS unigenes (Supplemental Table S4). However, we did not identify any unigenes encoding interleukin-1 receptor NBS protein in _O. granulata_.

**Comparative Transcriptome Analysis of O. granulata and O. meyeriana**

_Oryza granulata_ and _O. meyeriana_ are the two wild rice species in the _O. meyeriana_ complex. Previous studies showed that the two species have different characteristics, including length and width, flag-leaf width and length, culm length, and leaf blade size (Gong et al., 2000; Vaughan, 1994). Additionally, F1 hybrids derived from the cross between _O. granulata_ and _O. meyeriana_ had an average spikelet fertility of only 31.28%, implying, to a certain extent, that reproductive barriers exist between _O. granulata_ and _O. meyeriana_ (Gong et al., 2000). To further understand the genomic differences between _O. granulata_ and _O. meyeriana_, we performed a comparative transcriptome analysis with the transcriptomics data generated from panicles and leaves of _O. granulata_ and _O. meyeriana_. We assembled the clean reads with Trinity to construct unique consensus sequences, which resulted.
in the detection of 79,810 and 67,743 unigenes from *O. granulata* and *O. meyeriana*, respectively. Furthermore, we searched for orthologous unigenes between *O. granulata* and *O. meyeriana* using BLASTp and OrthoMCL (Li et al., 2003). We found 45,331 unigenes belonging to 16,565 gene families with an E-value less than 10⁻⁷, which were considered to be the homologous unigenes shared by *O. granulata* and *O. meyeriana*. Among the 45,331 homologous unigenes, 11,766 shared by *O. granulata* and *O. meyeriana* had a one-to-one orthologous relationship (Supplemental Table S5).

To perform an evolutionary analysis, we used 2865 single-copy orthologous unigene pairs (Supplemental Table S6) shared by *O. granulata* and *O. meyeriana* [excluding identical sequences and those with synonymous substitutions (Ks) > 0.1] to detect variations in the ratio of nonsynonymous to synonymous substitutions (Ka/Ks) with the Ka/Ks Calculator Toolbox, version 2.0 (Wang et al. 2010b). The mean values of Ka, Ks, and Ka/Ks...

---

**Fig. 2.** Genes predicted to be involved in the plant-pathogen interaction pathway. The red box indicates unigenes expressed in *O. granulata*.

**Fig. 3.** Comparative transcriptome analysis of *O. granulata* and *O. meyeriana*. (A) Venn diagram showing the distribution of unigenes and gene families (numbers in parentheses) between *O. granulata* and *O. meyeriana*. (B) Distribution of the ratios of nonsynonymous substitutions to synonymous substitutions (Ka/Ks). Blue dots indicate Ka/Ks > 1, green dots indicate 0.5 < Ka/Ks < 1, red dots indicate Ka/Ks < 0.5, black dots indicate Ka/Ks < 0.1.
Ks were 0.0126, 0.0219, and 0.5936, respectively. Of the orthologous pairs examined, 1311 (45.76%) had a Ks / Ks > 0.5 (including 359 orthologous pairs with Ks / Ks > 1, 952 orthologous pairs with 0.5 < Ks / Ks < 1), and 1534 orthologous pairs had Ks / Ks < 0.5 (Fig. 3B). Genes with an overall Ks / Ks of >0.5 are more likely to have been subjected to adaptive evolution (Swanson et al., 2004). These results indicate that 1311 (45.76%) unigenes may have been subjected to positive selection under natural conditions in O. granulata and O. meyeriana.

To verify the transcriptome analysis results, three single-copy orthologous unigenes with a Ks / Ks of >0.5 were selected for further analysis, including the orthologs of the rice genes Pi-ta (rice blast resistance gene) (Bryan et al., 2000), ZEP1 (rice meiosis gene) (Wang et al., 2010a), and GW2 (rice grain width gene) (Song et al., 2007). The full-length cDNA (FLcDNA) sequences of the three orthologous genes were amplified from O. granulata and O. meyeriana by 3’5’-RACE-PCR. For all genes, the cDNA sequences obtained by 3’5’-RACE-PCR were identical to those obtained from the transcriptome analysis in both O. granulata and O. meyeriana (Supplemental Fig. S3A–C), indicating that the results of the transcriptome analysis are reliable. Moreover, amino acid sequences alignment of the three orthologous proteins between O. granulata and O. meyeriana identified eight amino acid differences in the NBS, LRR, and thioredoxin (TRX) domains of Pi-ta (Supplemental Fig. S3A), two amino acid differences in the structural maintenance of chromosomes domain of ZEP1 homologs (Supplemental Fig. S3B), and one amino acid difference in the GW2 homologs (Supplemental Fig. S3C). The results suggest that all three genes were subjected to differential adaptive evolution in O. granulata and O. meyeriana and might be good candidates for further study.

Isolation of xa13 and Pi-ta Orthologs in O. granulata

Although O. granulata and O. meyeriana are known to be resistant to bacterial blight and rice blast, so far, few disease resistance genes have been characterized in these two species. To obtain further insights into the characteristics of their disease resistance, we performed cloning and functional analysis of the xa13 (recessive bacterial blight R gene) (Chu et al., 2006a, 2006b) and Pi-ta (dominant rice blast R gene) (Bryan et al., 2000) orthologs in O. granulata.

To isolate the orthologs of rice xa13 and Pi-ta in O. granulata, we first performed a BLAST search against the protein coding region prediction database of O. granulata by using the rice xa13 and Pi-ta amino acid sequences from GenBank (AF171278 and AAC00132). A 912-bp fragment (Unigene 0037134) sharing 92.8% nucleotide identity with xa13 and a 1686-bp fragment (Unigene 0053633) sharing 81.3% nucleotide identity with Pi-ta were obtained. Subsequently, we conducted 3’5’-RACE using specific RACE primers (Supplemental Table S1) to amplify the FLcDNA sequences in O. granulata. The FLcDNA of the xa13 ortholog is 1441 bp (named OgXa13) and the FLcDNA of the Pi-ta ortholog is 3581 bp (named OgPi-ta).

On the basis of the FLcDNA sequence information, we designed specific primer sets (Supplemental Table S1) located in the corresponding 5’ and 3’ termini to amplify their genomic sequences in O. granulata. The complete genome sequences of OgXa13 and OgPi-ta are 2244 bp and 5040 bp, respectively. The result of the sequence analysis with Vector NTI version 11.5 (Thermo Fisher Scientific) showed that a 2244-bp genome sequence of OgXa13 consisted of five exons and four introns with the same genome structure as xa13 in rice, encoding a 304-amino acid polypeptide and sharing 94.2% amino acid similarity with xa13 (Fig. 4A). Similarly, a 5040-bp genome sequence of OgPi-ta also consisted of five exons and four introns with a different genome structure from Pi-ta in rice (Fig. 5A), encoding a 1024-amino acid polypeptide and sharing only 72.6% amino acid similarity with Pi-ta.

In rice, xa13 encodes a 307-amino acid protein containing two Medicago truncatula Nodulin 3 or saliva domains (Yuan et al., 2010). In comparison to Xa13 and xa13, 16 amino acid substitutions, one amino acid insertion, and four amino acid deletions were identified in the O. granulata orthologous protein (Fig. 4A). Among them, three amino acid substitution variations were found to be located in the conserved MtN3.1 domain interacting with COPT1 and COPT5 in rice (Yuan et al., 2010). Analysis of the promoter sequence of OgXa13 in O. granulata found that the crucial promoter sequence [UPTRbox] is identical to that of the susceptible allele Xa13 (Fig. 4B). In rice, the resistance gene xa13 (recessive) and the susceptible allele Xa13 (dominant) encode the same protein. Mutations in the promoter of Xa13 cause it to be downregulated during host–pathogen interactions, resulting in the fully recessive xa13, which confers race-specific resistance (Chu et al., 2006b).

Previous studies have shown that the Pi-ta gene in O. sativa consists of two exons and one intron, encoding a protein belonging to the NBS-LRR class (Bryan et al., 2000) and confers resistance to races of M. oryzae carrying the avirulence gene AVR-Pita in a gene-for-gene manner (Marc et al., 2002). A single amino acid difference within the leucine-rich domain of the Pi-ta protein was found to be responsible for reaction specificity with the corresponding fungal elicitor AVR-Pita (Bryan et al., 2000). Compared with Pi-ta in rice, OgPi-ta encodes a protein with a centrally located NBS domain and shorter leucine-rich domain region containing only eight repeats (Pi-ta contains 10 imperfect repeats of various length ranging from 16 to 75 amino acids with a consensus LxxLxxL) and the addition of a TRX domain at their C-terminus (Fig. 5B,C). Two leucine repeats missing from the OgPi-ta protein have been shown to harbor the amino acid alanine-918 (Fig. 5B,C), which has been associated with the gene-for-gene specificity of the Pi-ta–AVR-Pita interaction (Marc et al., 2002). Previous gene expression analysis has revealed that transcript variants of Pi-ta encoding the TRX domain have the highest level of expression compared with the other full length or truncated transcripts in a resistant rice variety containing...
**Pi-ta** (Costanzo and Jia, 2009), but their function was not investigated further. Furthermore, in the specific case of Cf-9, the TRX domain has been shown to act as an independent adaptor protein recruiting a protein kinase (ACIK1) upon elicitation by the Avr9 peptide (Nekrasov et al., 2006). This implies that transcripts with open reading frames coding for the TRX domain may encode for totally independent disease resistant specificities, or constitute essential components that interact directly or indirectly with the Pi-ta protein to mediate signal perception and transduction.

**Functional Characterization of OgPi-ta**

To explore the biological function of OgPi-ta, we generated the overexpression construct OgPi-ta-ox and introduced it into the blast susceptible rice variety Nipponbare through *Agrobacterium* -mediated transformation, with an empty vector as a control. Transformed OgPi-ta-ox and empty vector control plants, as well as wild-type Nipponbare, were inoculated with *M. oryzae* strain 1817–2 carrying Avr-Pita. Transgenic plants expressing OgPi-ta-ox enhanced the resistance of Nipponbare to blast strain 1817–2, with smaller disease lesions observed on the leaves than in the wild-type (Nipponbare) and empty vector control (Fig. 6A,B). All plants were also transplanted to the greenhouse and injection-inoculated with blast strain 1817–2 at the tillering to booting stage. The blast disease symptoms exhibited by these plants were relatively consistent with those observed after spray-inoculation (Supplemental Fig. S4). Gene expression analysis revealed that the expression level of OgPi-ta in four transgenic plants was significantly increased (Fig. 6C). These results suggest a role for OgPi-ta in disease resistance against rice blast. Other functions of the OgPi-ta protein, such as effect on growth, development and environmental stress in rice, should be investigated in future studies.
**DISCUSSION**

*Oryza granulata* is a perennial diploid wild rice species in the *O. meyeriana* complex with a GG genome (Gong et al., 2000; Sharma and Shastry, 1965; Vaughan, 1984, 1994; Zou et al., 2008). This plant is usually found in well-drained soil under the trees in deciduous forests, in bamboo (*Bambusa* spp.) thickets, on limestone hills or mountains in tropical Asia, including Nepal, India, southern China, the Philippines, Malaysia, and Indonesia (Gong et al., 2000; Sharma and Shastry, 1965; Tateoka, 1962; Vaughan, 1989). It is a reservoir of useful genes for cultivated rice improvement, such as resistance to bacterial blight and rice blast and tolerance to drought and shade (Bon et al., 2011; Brar and Khush, 1997; Jena, 2010; Yan et al., 2004). However, no genetic resources are available at present for investigating *O. granulata*. In this study, by taking advantage of transcriptome sequencing technology, we produced 91,562 high-quality transcripts from *O. granulata* and annotated 46,393 of these transcripts. A comparative analysis with genes known to play a role in the plant-pathogen interaction pathway led to the identification of 324 genes involved in this pathway. Furthermore, by performing BLASTp analysis of the deduced amino acid sequences of *O. granulata* against the Plant Resistance Genes Database (PRGdb; http://prgdb.crg.eu, accessed 15 June 2018), we also identified 1166 RLP-encoding unigenes, 209 RLK-encoding unigenes, and 2597 NBS-encoding unigenes. These high-quality reference transcriptomes are indicated by black boxes and introns are indicated by white boxes; 5' untranslated region (UTR) and 3'UTR are indicated by gray boxes. The red arrow indicates the nucleotide site that distinguishes resistant and susceptible alleles in rice. The leucinerich repeat (LRR) domain is a yellow box. Thioreredoxin (TRX) is indicated by a red box. The red arrow indicates the residue of the amino acid alanine-918, which distinguishes resistant and susceptible alleles in rice. (C) Amino acid sequences of *OgPtA* and *PtA* proteins. The N-terminal region is a gray box and the C-terminal region is a yellow box. The amino acid sequence of *PtA* was obtained from GenBank (AAK00132, http://www.ncbi.nlm.nih.gov/, accessed 15 June 2018).
sequences from *O. granulata* provide an important resource for functional and evolutionary studies of the genus *Oryza*.

*Oryza meyeriana* is another perennial diploid wild rice species in the *O. meyeriana* complex. This plant has a somewhat smaller distribution area than *O. granulata*, primarily restricted to the Philippines, Malaysia, Thailand, and southern China. Morphological analysis showed that 14 of the 20 quantitative characteristics, such as grain length and width, flag-leaf width and length, culm length, and leaf blade size, are significantly different between *O.*
granulata and O. meyeriana (Gong et al., 2000; Vaughan, 1994). Investigation of the genetic diversity between O. granulata and O. meyeriana revealed a genetic similarity coefficient of 0.488, which is lower than the genetic similarity of other Oryza species (Qian et al., 2006). F1 hybrids derived from a cross between O. granulata and O. meyeriana had an average spikelet fertility of only 31.28%, suggesting that may be some reproductive barriers between O. granulata and O. meyeriana (Gong et al., 2000). Our comparative transcriptome analysis of O. granulata and O. meyeriana revealed that 25.94% of the unigenes were unique to O. granulata and did not have homologs in O. meyeriana, whereas 32.03% of the unigenes were unique to O. meyeriana and did not have homologs in O. granulata, indicating that there are genomic differences between O. granulata and O. meyeriana. In addition, analysis of the K/K ratios of 2865 single-copy orthologous unigene pairs shared by O. granulata and O. meyeriana suggested that 1311 (45.76%) of the orthologous pairs (K/K > 0.5) had been subjected to adaptive evolution under natural conditions in O. granulata and O. meyeriana (Yang and Bielawski, 2000). We also amplified the FLcDNA sequences of three single-copy orthologous genes to rice Pi-ta (Bryan et al., 2000), ZEP1 (Wang et al., 2010a), and GW2 (Song et al., 2007) from O. granulata and O. meyeriana using 3’/5’-RACE PCR. Multiple amino acid differences were found in the functional domains of three homolog proteins between O. granulata and O. meyeriana, which also reflected their differences at the genome level. To adapt to the pressure of environmental selection during long-term evolution, O. granulata and O. meyeriana have developed unique genetic resources specific to the conditions in which they live. These results suggest that both species can be used for different purposes in scientific research and breeding.

Oryza granulata and O. meyeriana are resistant to bacterial blast and rice blast (Cheng et al., 2016; Luo et al., 2010; Chen et al., 2016; Yan et al., 2004). During coevolution with pathogens, plants have developed a set of defense strategies against various pathogens, with an innate immune machinery consisting of PTI conferred by pattern recognition receptors and ETI conferred by R proteins (Spoel and Dong, 2012). Analysis of the plant–pathogen interaction pathway uncovered 324 unigenes encoding 24 proteins involved in the plant–pathogen interactions in O. granulata. Among these, 182 unigenes encoding 13 proteins including cyclic nucleotide-gated channel, Ca-binding protein, Ca-dependent protein kinase, respiratory burst oxidase, nitric oxide synthase, FLS2, mitogen-activated protein kinase kinase 1, mitogen-activated protein kinase kinase 1/2, mitogen-activated protein kinase kinase 4/5, WRKY transcription factor 22/29, WRKY transcription factor 25/33, glycerol kinase and pathogenesis-related protein 1 are enriched in the PTI process and 142 unigenes encoding 11 proteins including CERK1, PRM1-interacting protein 4, disease-resistance protein RPM1, disease-resistance protein RPS2, serine or threonine protein kinase PBS1, the suppressor of the G2 allele of skp1, enhanced disease susceptibility 1, Heat Shock Protein 90, Pto-interacting protein 1, Mla12 resistance protein RAR1, and WRKY transcription factor 12 are enriched in the ETI process, for which the numbers are similar to those in O. sativa (Supplemental Fig. S5) and O. meyeriana (Supplemental Fig. S6).

To date, few disease resistance genes have been characterized in O. granulata and O. meyeriana. In the current study, two orthologs in O. granulata, OgXa13 (orthologous to the rice bacterial blight R gene xa13) (Chu et al., 2006a, 2006b; Yuan et al., 2010) and OgPi-ta (orthologous to the rice blast R gene Pi-ta) (Bryan et al., 2000) were cloned. The resistance gene xa13 (recessive) and the susceptible allele Xa13 (dominant) encode the same protein. Having one to three amino acid changes in xa13 is not a determinant of xa13-mediated resistance (Chu et al., 2006b), the insertions and deletions or substitution mutations within the −69 to −86 region of the Xa13 promoter downregulated its expression during host–pathogen interactions, resulting in the fully recessive xa13 confers race-specific resistance, such that the lower the expression level of either xa13 or Xa13, the more resistant the plants become (Chu et al., 2006b; Yuan et al., 2011). The crucial promoter sequence [the UPT box (Romer et al., 2010)] in OgXa13 of O. granulata is identical to that of the susceptible allele Xa13 (Fig. 4B) and the expression of OgXa13 was also induced in O. granulata by Xanthomonas oryzae pv. oryzae strain PXO99 inoculation (Supplemental Fig. S7A). However, O. granulata treated with X. oryzae pv. oryzae strain PXO99 infection via the leaf-clipping method (Yuan et al., 2010) showed high resistance to X. oryzae pv. oryzae strain PXO99 (Supplemental Fig. S7B), suggesting that O. granulata contains many other resistance genes conferring the resistance to X. oryzae pv. oryzae strain PXO99. Further isolation and functional characterization of genes conferring resistance against bacterial blight disease will help us clarify the molecular mechanisms of O. granulata.

The genome structure of OgPi-ta, a new transcript variant of the Pi-ta gene in O. granulata, is quite different from the Pi-ta gene in rice but, similar to KPt, one of the alternative splice transcripts of the Pi-ta gene in the japonica rice cultivar rice Katy, which harbors the dominant blast resistant Pi-ta allele (Costanzo and Jia, 2009; Lee et al., 2009b). OgPi-ta encodes a protein with a centrally located NBS domain, a shorter leucine-rich domain region with two missing repeats, and an additional TRX domain at the C-terminus. The two missing leucine repeats in OgPi-ta protein harbor the amino acid alanine-918, which has been associated with the gene-for-gene specificity of the Pi-ta–AVR-Pita interaction (Fig. 5B,C). Previous studies showed that transcript variants encoding the TRX domain in a resistant rice variety containing Pi-ta had the highest level of expression compared with the other full-length or truncated transcripts (Costanzo and Jia, 2009). In the specific case of Cf-9, the TRX domain had been shown to act as an independent adaptor protein, recruiting a protein kinase (ACIK1) on elicitation with the Avr9 peptide (Nekrasov et al., 2006).
This suggests that transcripts with open reading frames coding for the TRX domain may encode for totally independent disease-resistant specificities or constitute essential components interacting directly or indirectly with the Pi-ta protein to mediate signal perception and transduction. However, there is no evidence as yet to support this hypothesis. We generated the overexpression construct containing the OgPi-ta gene (OgPi-ta-ox) and introduced it into the blast-susceptible rice variety Nipponbare through Agrobacterium-mediated transformation. The transgenic Nipponbare plants expressing OgPi-ta demonstrated enhanced resistance to the rice blast strain 1817−2 carrying the AVR-Pi-ta gene (Fig. 6A,B; Supplemental Fig. S4). These results provide evidence, to some extent, that the Pi-ta protein with a TRX domain plays a role in plant disease resistance. Furthermore, functional investigations of the OgPi-ta gene should be performed to provide molecular evidence to help us understand its function. Moreover, TRXs in association with NBS-LRR proteins had been reported to exhibit defense responses to biotic and abiotic stresses in tomato (Solanum lycopersicum L.) Cf-9, victorin sensitivity in oats (Avena sativa L.) and Arabidopsis thaliana (L.) Heynh. mediated by the LOV1 protein (Rivas et al., 2004; Nekrasov et al., 2006; Sweat and Wolpert, 2007). Reads per kb per million reads analysis (Mortazavi et al., 2008) have shown that OgPi-ta is highly expressed in O. granulata root (Supplemental Fig. S8) and further suggests that the OgPi-ta gene may play roles in the response to various environmental stresses.

Supplemental Information

Supplemental Table S1: Primers used for PCR amplification.

Supplemental Table S2: Summary of data output quality of the various samples from O. granulata.

Supplemental Table S3: The unigenes enriched in the plant–pathogen interaction pathway in O. granulata.

Supplemental Table S4: RLK, RLP, and NBS-LRR-encoding unigenes in O. granulata.

Supplemental Table S5: The single-copy orthologous unigenes shared by O. granulata and O. meyeriana.

Supplemental Table S6: The single-copy orthologous unigene pairs used for the detection of K_d/K_s ratios between O. granulata and O. meyeriana.

Supplemental Fig. S1: Length distribution of the unigenes in O.granulata.

Supplemental Fig. S2: The distributions of similarity and proportional E-value frequency of annotated unigenes in protein databases.

Supplemental Fig. S3: Amino acid sequence alignment of three orthologous proteins between O. granulata and O. meyeriana.

Supplemental Fig. S4: Plants expressing OgPi-ta-ox were inoculated with M. oryzae strain 1817−2 by injection at the tillering to booting stage.

Supplemental Fig. S5: Genes predicted to be involved in the plant–pathogen interaction pathway in O. sativa.

Supplemental Fig. S6: Genes predicted to be involved in the plant–pathogen interaction pathway in O. meyeriana.

Supplemental Fig. S7: Investigation of OgXa13 expression and the resistance of O. granulata followed by infection with X. oryzae pv. oryzae strain PXO99.

Supplemental Fig. S8: The expression levels of OgPi-ta in the root, stem, sheath, panicle, and leaf of O. granulata.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENTS

We thank Prof. Chengzhi Liang (Institute of Genetics and Development Biology, Chinese Academy of Sciences) for sequencing. Dr. Li Liu for critical reading of the manuscript, Prof. Cailin Lei and Huanbin Zhou (Chinese Academy of Sciences) for kindly providing us with the M. oryzae and X. oryzae pv. oryzae strains, Dr. Xiaojun Chen providing help with RNA extraction, and IRRI for providing us with O. granulata materials. This work was supported by the National Key R&D Program of China (2016YFD0100600) and the National Natural Science Foundation of China (U1302261).

REFERENCES


Orzya meyeriana against Xanthomonas oryzae pv. oryzae. Sci. Rep. 6:38215. doi:10.1038/srep38215


