








Deciphering global DNA variations and embryo sac fertility in autotetraploid rice line

Fozia GHOURI^{1,2,3,*} , Jiani ZHU^{1,2,3,*} , Hang YU^{1,2,3,*} , Jinwen Wu^{2,3} , Faheem Shahzad BALOCH⁴ ,
Xiangdong LIU^{1,2,3} , Muhammad Qasim SHAHID^{1,2,3,*} 

¹State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources,
South China Agricultural University, Guangzhou, P.R. China

²Guangdong Provincial Key Laboratory of Plant Molecular Breeding, South China Agricultural University, Guangzhou, P.R. China

³College of Agriculture, South China Agricultural University, Guangzhou, P.R. China

⁴Department of Field Crops, Faculty of Agricultural and Natural Sciences, Abant İzzet Baysal University, Bolu, Turkey

Received: 07.01.2019 • Accepted/Published Online: 16.05.2019 • Final Version: 03.12.2019

Abstract: Autotetraploid rice is a new germplasm resource obtained by doubling chromosomes through colchicine treatment. There have been many studies on the reproductive characteristics of autotetraploid rice, but little is known about global DNA variations and reasons for low embryo sac fertility in autotetraploid rice. Therefore, an autotetraploid rice line (T446) was used for resequencing and embryo sac fertility was observed. Whole-genome resequencing data revealed 87,229 SNPs and 11,022 InDels in the genome of T446 versus E246 (diploid rice), which had an average of 23.37 SNPs and 2.95 InDels per 100 kb. A total of 17,375 and 17,171 structural variations and 131 and 128 copy number variations were identified in the autotetraploid and its diploid counterpart, respectively. We detected 140 large-effects SNPs and InDel variants that might be related to the embryo sac fertility of autotetraploid rice, including 10 genes that may be closely associated with the development of the embryo sac. Of these, *Os02g0292600* and *Os06g0565200* were specifically expressed in the ovary. Mature embryo sac fertility was observed through whole-mount eosin B-staining confocal laser scanning microscopy. Many abnormalities were found in the embryo sac of T446, including embryo sac degeneration, embryo sac without female germ unit, abnormal polar nuclei, and poly-eggs, which, in turn, resulted in low seed set. However, whole-genome polymorphisms and genetic differences were high and exhibited broad prospects for genetic improvement. Genetic mutations in genes associated with embryo sac fertility in polyploid rice require further studies.

Key words: Embryo sac abortion, genome, polyploidy, resequencing, seed set

1. Introduction

Biological resources, particularly plants, are vital not only for the global economy but also for meeting the nutritional requirements of the poor. Greater diversity ensures natural sustainability for all life forms. Moreover, the richer the plant biodiversity, the greater the opportunity for medical findings and adaptive responses to new challenges, such as climate change (Liu et al., 2015; Alp et al., 2016; Galiana-Belaguer et al., 2018; Yaldız et al., 2018). Polyploidy plays a critical role in plant biodiversity and evolution (Doyle et al., 2008; Soltis et al., 2009). Over 70% of all angiosperm species have increased their polyploidy during evolutionary processes (Masterson, 1994). Allopolyploidy has long been thought to be more vital and common in natural populations than in autopolyploid species; hence, relatively little attention has been paid to autopolyploid species (Parisod et al., 2010). Autotetraploid rice was derived from diploid rice by chromosome doubling

through colchicine treatment and has a pronounced potential to enhance yield and nutrition. Autotetraploid rice had higher genetic variation and hybrid vigor than diploid rice, including higher 1000-grain weight, larger kernels, higher protein, higher biomass yield, and better amino acid contents than diploid rice (Song and Zhang, 1992; Luan et al., 2008; Shahid et al., 2011, 2012; Guo et al., 2017). However, low seed set is a major hindrance in the application of autotetraploid rice at a commercial level (Shahid et al., 2013b; Wu et al., 2013).

Our research group carried out microarray analysis to evaluate genetic variation during pollen development in autotetraploid rice (Taichung65-4x) and diploid rice (Taichung65) and detected a total of 1251 differentially expressed genes (Wu et al., 2014). Moreover, small RNA sequencing revealed that partial embryo sac and pollen sterilities were caused by some specific differentially expressed miRNAs in autotetraploid rice (Li et al., 2016,

* Correspondence: qasim@scau.edu.cn

2017). Polyploidy enhanced multiallelic interactions at pollen sterility loci and increased abnormalities in chromosome behavior (Wu et al., 2015). Recently, cytological studies revealed that the pervasive deleterious genetic interactions at pollen sterility loci cause high pollen sterility in polyploid rice hybrids, which could be overcome by double natural genes (Wu et al., 2017; Chen et al., 2018; Chen et al., 2019). Guo et al. (2017) identified several genes related to heterosis and fertility in neotetraploid rice hybrids by RNA sequencing. However, little information about genome-wide DNA variation is available in autotetraploid rice compared to its diploid counterpart, especially regarding embryo sac fertility.

Recently, some NBS-LRR genes were identified by using whole-genome resequencing in two newly developed rice lines from *Oryza rufipogon*, which are native to Dongxiang, Jiangxi Province, China (Liu et al., 2017). By using pangenome analyses, 2.4 million small insertions/deletions (InDels), 29 million single-nucleotide polymorphisms (SNPs), and over 90,000 structural variations, which lead to within- and between-population variations, were detected (Wang et al., 2018). SNPs and InDels have become the markers of choice for high-resolution linkage mapping, population genomics studies, and association mapping (Subbaiyan et al., 2012; Zhang et al., 2016; Nadeem et al., 2018; Wang et al., 2019). The large polymorphisms are termed as structural variations (SVs) and copy number variations (CNVs), and many recent developments in plant genetics have been achieved through advances in high-resolution technologies (Huang et al., 2013; Yu et al., 2018).

To date, there has been no study of autotetraploid rice lines with very low embryo sac fertility based on whole-genome resequencing. In the current study, an autotetraploid rice line and a diploid cultivar were resequenced through NGS and mapped onto the reference genome of Nipponbare. Detection of genome-wide InDels and SNPs among these lines has the potential to provide useful resources for future genetic studies. Moreover, we observed mature embryo sac fertility in autotetraploid rice, which will help us to understand the types of embryo sac abortion. We also investigated the agronomic traits and pollen fertility of autotetraploid lines. The results may provide elite genes associated with the embryo sac in polyploid rice. The information described here can be exploited in future studies to provide novel observations for embryo sac fertility and rice genetics.

2. Materials and methods

2.1. Plant materials and DNA isolation

Autotetraploid rice line T446 was developed from a diploid rice cultivar (E246) in 2004 and self-crossed for more than 25 generations at the farm of South China Agricultural

University (SCAU). The autotetraploid rice line used in this study has been investigated by our research group for the last 13 years and is genetically stable. Genomic DNA was isolated from the leaves of 2-week-old seedlings using the modified cetyltrimethylammonium bromide (CTAB) method (Zhang et al., 2012). Genomic DNA quality was evaluated by NanoDrop 2000 and agarose gel electrophoresis.

2.2. Evaluation of agronomic traits

A total of 15 plants of the autotetraploid line were harvested from the field at maturity. Agronomic traits, including panicle length, seed set, 10-grain length, 10-grain width, and ratio of grain length to width were measured. These agronomic traits were selected and investigated or measured according to the 2012 test guidelines for the registration of a new plant, namely variety distinctness, uniformity, and stability (DUS), in rice (*Oryza sativa* L.) in China.

2.3. Whole-genome resequencing analysis

The genomic resequencing library was prepared according to the standard protocol of Illumina. Pair-end sequencing reads generated by the Illumina HiSeq platform (Biomarker Technologies, Beijing, China) were then stored in a FASTQ file, and their quality was evaluated using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Low-quality reads, i.e. reads with sequencing adapter, reads with >10% N content, and reads with >50% low-quality bases (<10), were removed, and the high-quality reads were mapped onto the Nipponbare reference genomes by Burrows-Wheeler Aligner (BWA) software (Li and Durbin, 2009). MarkDuplicates in Picard (<https://sourceforge.net/projects/picard/>) was used to eliminate PCR duplication. Base recalibrations and realignments near insertion or deletion regions were conducted using the Genome Analysis Toolkit (GATK). Genomic variations were called using GATK and were stored in a VCF file (i.e. diploid vs. Nipponbare and autotetraploid vs. Nipponbare). Reference genome coverage was estimated using SAMtools (McKenna et al., 2010).

2.4. Annotation of SNPs, InDels, SVs, and CNVs

Annotations of SNPs and InDels were performed using SnpEff software. The distribution of variations was analyzed in all chromosomes. The following SNPs and InDels were filtered: two or more SNPs in a 5-bp or shorter window, SNPs near (5 bp or less) InDels, and two or more InDels in a 10-bp or shorter window. Furthermore, we retained the SNPs and InDels with a coverage depth range of 11–100. BreakDancer software was used to call the SVs (Chen et al., 2009), and all SVs and CNVs were annotated based on the Nipponbare reference genome annotation GFF file. FREEC software was used to call CNV variations (Boeva et al., 2012).

2.5. Functional annotation and predicted protein-protein interaction

Gene ontology (GO) enrichment analysis (<http://www.geneontology.org>) was performed with the *Oryza sativa* setting as a species background in Nipponbare. The protein-protein interaction network was predicted using the STRING website tool (Szklarczyk et al., 2015).

2.6. Embryo sac observation

The mature florets with open glumes were collected from the field between 0900 and 1200 hours and fixed in Carnoy's solution (ethanol: acetic acid = 3:1) for at least 24 h. Then the mature florets were washed and kept in 70% ethanol at 4 °C. Lemma and palea were cut under a binocular dissecting microscope, and dissected ovaries were hydrated with a series of ethanol (50%, 30%, and 10%) and distilled water. Whole mount eosin B confocal laser scanning (WE-CLSM) was performed according to Shahid et al. (2010) and Li et al. (2017) with some minor modifications. The ovaries were treated with 2% aluminum potassium sulfate for 30 min and then stained with eosin B (10 mg/L dissolved in 4% sucrose solution) for 10–14 h at room temperature. The ovaries were rinsed again with 2% aluminum potassium sulfate for 30 min, which produced a clear image by removing dye from the ovary wall. The samples were washed three times with distilled water and dehydrated sequentially with ethanol solutions (30%, 50%, 70%, 90%, and 100%) for 30 min. The dehydrated samples were dipped into a mixture of ethanol absolute and methyl salicylate (1:1) for 2–3 h and then cleaned for 1–2 h in pure methyl salicylate solution. The samples were kept on a microscope slide and observed under a confocal laser scanning microscope (Heidelberg, Germany). The direction of the embryo sac was maintained so that the polar nuclei, antipodal cells, and egg apparatus had maximum slice area. Four to six pictures of different sizes were taken.

2.7. Pollen fertility observation

The mature florets were collected from the field early in the morning (0600 to 0800 hours) and fixed in Carnoy's solution (ethanol:acetic acid, 3:1) for 24 h, and the samples were stored in 70% ethanol at 4 °C after washing two times with 70% ethanol. Pollen fertility of the autotetraploid rice line was investigated by using 1% I₂-KI (Shahid et al., 2013a). First, we prepared a solution of KI by dissolving 8 g of KI in 20 mL of water. Then we added 1 g of I₂ to that solution and mixed it by stirrer until all I₂ dissolved.

Finally, more water was added to the solution to bring the volume up to 100 mL. I₂ does not dissolve in polar solvents such as water, so we used KI to prepare the I₂ solution. The anthers were dissected from the floret and placed in a small drop of I₂-KI on a microscope slide. The slide was covered with a cover slip after 1–2 min and investigated under a microscope (Motic BA200). Pollen fertility was divided into five categories, i.e. normal pollens, stained abortive pollens, typical abortive pollens, small pollens and empty abortive pollens.

3. Results

3.1. Agronomic traits and pollen fertility of autotetraploid rice line

Pollen fertility and some agronomic or yield-related traits of an autotetraploid rice line were observed in this study. The seed set of T446 was 7.50%, 10-grain length was 11.62 cm, 10-grain width was 3.42 cm, and length-to-width ratio was 3.40 (Table 1). Pollen fertility was 25.38% in the autotetraploid rice line. For abortive pollen types (Figures 1A–1D), the stained abortive pollen type was the most common (Figure 1B, arrows), reaching 59.93%, followed by typical abortive pollens (Figure 1A, arrows), which accounted for 10.31%, and empty abortive pollens (Figure 1C, arrows) and small pollens (Figure 1D, right arrow), which were 1.68% and 2.71%, respectively (Table 1).

3.2. Detection and distribution of genomic polymorphisms in autotetraploid and diploid rice

Through whole-genome resequencing, diploid and autotetraploid rice generated 76,528,170 and 76,664,648 sequencing reads (raw reads), respectively. The reads with adapters and reads with N content exceeding 10% and reads with >50% low-quality values were filtered. Total clean reads of diploid and autotetraploid rice were 74,493,049 and 74,527,181, respectively. The percentage of the total number of bases (Q30) in which the mass value was ≥30 accounted for 86.73% and 86.06% of the two materials, and the average mapping ratios were 88.46% and 88.45%, respectively. The average coverage depths were 42 X and 43 X, and GC contents were 41.70% and 41.86%, respectively (Table 2). These sequencing results showed that the whole-genome resequencing data had high quality and good consistency and could be used for subsequent biological analysis.

Total SNPs in diploid and autotetraploid rice were 2,537,718 and 2,535,390, respectively (Table 3). In diploid

Table 1. Agronomic traits and pollen fertility of autotetraploid rice line.

Panicle length (cm)	Seed set (%)	10-grain length (cm)	10-grain width (cm)	Length-to-width ratio	Normal pollens (%)	Pollen abortion types (%)			
						Typical abortive	Stained abortive	Empty abortive	Small pollens
22.64 ± 1.05	7.50 ± 1.14	11.62 ± 0.09	3.42 ± 0.12	3.40 ± 0.08	25.38 ± 2.19	10.31 ± 1.32	59.93 ± 3.67	1.68 ± 0.04	2.71 ± 0.10

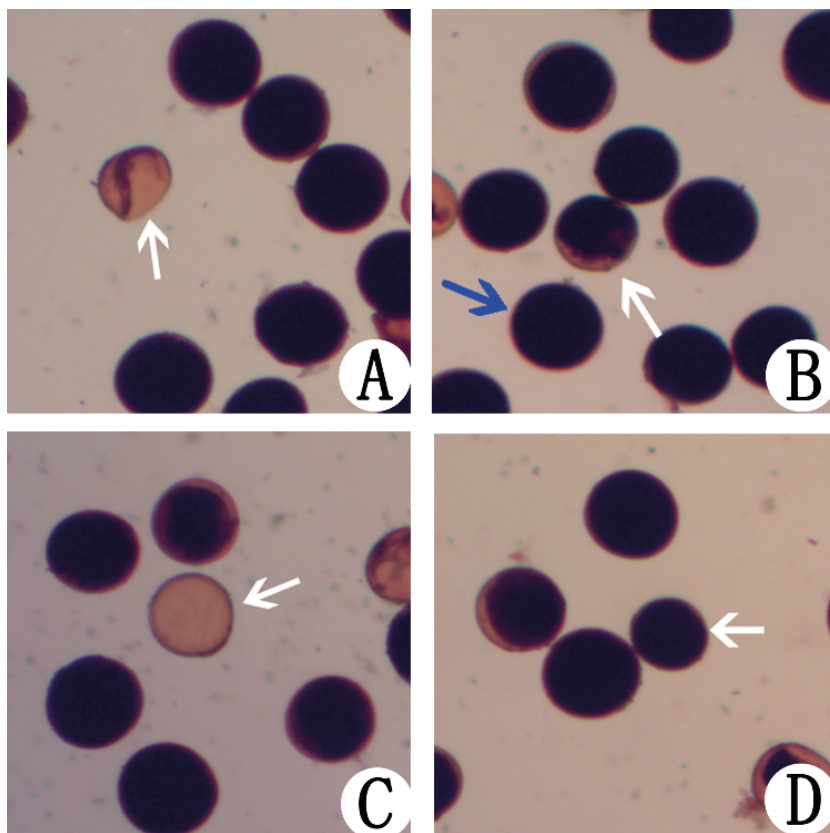


Figure 1. Pollen fertility of autotetraploid rice line. A) Typical abortive pollens (arrow); B) stained abortive pollens (white arrow) and normal pollens (blue arrow); C) empty abortive pollens (arrow); D) small pollen.

Table 2. Summary of the sequencing data of diploid and autotetraploid rice.

Sequencing parameters	Diploid	Autotetraploid
Raw reads	76,528,170	76,664,648
Clean reads	74,493,249	74,527,181
Average depth (fold)	42	43
Mapped (%)	88.46	88.45
Properly mapped (%)	83.59	83.53
Coverage ratio, 1 X (%)	91.48	91.53
Coverage ratio, 5 X (%)	87.60	87.65
Coverage ratio, 10 X (%)	84.55	84.58
GC (%)	41.70	41.86

rice, the ratio of transition (Ti) to transversion (Tv) was 2.61 (1,834,764 Ti and 702,954 Tv). In autotetraploid rice, the ratio of Ti to Tv was 2.61 (1,833,367 Ti and 702,023 Tv). In diploid rice, the SNP heterozygosity rate was 7.03% (178,391 heterozygous SNPs and 2,359,327 homozygous

SNPs), while the SNP heterozygosity rate was 6.91% (175,149 heterozygous SNPs and 2,360,241 homozygous SNPs) in autotetraploid rice.

In diploid rice, the InDel number was 360,444, including 176,069 insertions and 184,375 deletions. Heterozygous InDels and homozygous InDels were 348,660 and 11,784, respectively. A total of 6974 InDels were found in coding regions, including 3446 insertions and 3528 deletions and 6753 heterozygous and 221 homozygous InDels. In autotetraploid rice, the InDels were 359,805, including 175,666 insertions and 184,139 deletions. Heterozygous InDels and homozygous InDels were 347,996 and 11,809, respectively. In total, 6994 InDels were detected in coding regions, including 3461 insertions and 3533 deletions, 6772 heterozygous InDels, and 222 homozygous InDels (Table 4).

There was little difference in whole-genome mutations for the number of SNPs and InDels in diploid and autotetraploid rice compared to the Nipponbare reference genome. In order to further investigate the distribution of SNPs and InDels on each chromosome, the distribution on each chromosome was counted with 100 kb as a basic

Table 3. SNPs identified in diploid and autotetraploid rice.

Material	SNPs	Ti	Tv	Ti/Tv	Heterozygous	Homozygous	Het rate
Diploid	2,537,718	1,834,764	702,954	2.61	178,391	2,359,327	7.03%
Autotetraploid	2,535,390	1,833,367	702,023	2.61	175,149	2,360,241	6.91%

Ti: Transitions; Tv: transversions.

Table 4. Detection of InDels in diploid and autotetraploid rice.

Location	Type of variation	Diploid	Autotetraploid
CDS	Insertion	3446	3461
	Deletion	3528	3533
	Homozygous	6753	6772
	Heterozygous	221	222
	Total	6974	6994
Genome	Insertion	176,069	175,666
	Deletion	184,375	184,139
	Homozygous	348,660	347,996
	Heterozygous	11,784	11,809
	Total	360,444	359,805

unit. SNPs and InDels were 687.14 and 96.76 for each 100 kb of the diploid line and 686.55 and 96.60 for each 100 kb of autotetraploid rice, respectively. In the diploid and autotetraploid lines, the greatest SNP density (i.e. 775.08 SNPs/100 kb and 774.95 SNPs/100 kb) was observed on chromosome 10 (Table 5).

To understand whole-genome sequence variations in autotetraploid rice, the genes involved in the mutation sites, and whether the variations of these genes sequences are related to abnormalities of the embryo sac and agronomic traits in T446, the data of control diploid line were statistically compared with variations in T446. The differential mutations between three groups (autotetraploid vs. diploid rice, autotetraploid vs. reference genome, and diploid vs. reference genome) were observed, and data exhibiting no difference between lines were not included in the comparison. As a result, 87,229 SNPs and 11,022 InDels were detected between the whole genomes of diploid and autotetraploid rice lines, with an average of 23.37 SNPs and 2.95 InDels per 100 kb (Table 6). Compared with the Nipponbare reference genome, the number of SNPs + InDels in T446 was 2,895,190, which was 29.47 times the T446/E246 original polymorphism.

Large segment genomic variations such as SVs can affect genome stability. Numerous SVs were detected in autotetraploid and diploid rice. A total of 17,375 and 17,171 SVs were identified in autotetraploid rice and its

diploid counterpart, respectively (Table 7). In the exon, intron, and intergenic regions, 2330, 3012, and 18,070 SVs were detected in both types of rice, respectively (Table 8). Among these SVs, 1132 were detected in the exon region of T446. Moreover, 2255, 9087, and 407 insertions, deletions, and inversions were identified in autotetraploid rice, respectively. Three types of SVs, insertions (INs), deletions (DELs), and inversions (INVs), were distributed across twelve chromosomes. The number of DELs was much greater than INs and INVs in both types of rice. The number of total SVs in T446 was greater than the total in E246, which was primarily caused by the increase in deletions and interchromosomal translocations (CTXs). However, a higher number of insertions was found in E246 than T446. A total of 128 CNVs, including 78 copy numbers gained and 50 copy numbers lost with variations in length ranging from 50 kb to 3.3 Mb, and 131 CNVs, including 78 copy numbers gained and 53 copy numbers lost ranging from 50 kb to 4.1 Mb, were detected in the diploid and autotetraploid rice lines, respectively. The total number of CNVs in T446 was slightly higher than in E246. The CNVs were distributed unevenly on the 12 rice chromosomes (Figure 2).

3.3. SNPs and InDels: functional annotation in autotetraploid and diploid rice

SnEff software was used to annotate locations of the mutation sites in the genome, including the intergenic region, the gene region, and the CDS region, based on the position of the mutation site on the reference genome and the position of the gene on the reference genome. SnEff was also used to annotate the effects of mutation, including synonymous mutations and nonsynonymous mutations.

Excluding the mutations with coverage of <5, 86,123 SNPs and 10,853 InDels were detected between autotetraploid rice and its diploid counterpart, and 9159 and 1694 InDels were detected in the intergenic region and within the gene, respectively. In the gene region, there were 1006 InDels in the intron, 464 InDels in the untranslated region, and 224 InDels in the CDS region (Figure 3A). The MISSENSE InDels were involved in a single gene. The STOP_GAINED and START_LOST InDels were involved in three and two genes, respectively. A total of 81,686 SNPs were identified in the intergenic region and accounted for 94.85% of the total SNPs, while 4517 SNPs were detected within the gene and accounted for 5.15% of the total SNPs

Table 5. SNPs and InDel variations along the chromosomes of autotetraploid and diploid rice.

Chromosome	Diploid rice				Autotetraploid			
	Total SNPs	SNPs/100 kb	Total InDels	InDels/100 kb	Total SNPs	SNPs/100 kb	Total InDels	InDels/100 kb
chr01	256,107	591.87	40,187	92.87	255,947	591.50	40,172	92.84
chr02	248,821	692.38	37,438	104.18	248,630	691.84	37,361	103.96
chr03	230,896	634.09	35,594	97.75	230,578	633.22	35,530	97.57
chr04	217,102	611.51	30,360	85.51	216,531	609.90	30,236	85.17
chr05	226,276	755.30	31,359	104.68	225,791	753.68	31,327	104.57
chr06	222,369	711.61	31,803	101.77	221,991	710.40	31,707	101.47
chr07	201,371	678.07	28,355	95.48	201,329	677.93	28,327	95.38
chr08	195,922	688.82	27,199	95.63	195,713	688.09	27,215	95.68
chr09	161,220	700.57	21,994	95.57	161,165	700.33	22,039	95.77
chr10	179,874	775.08	23,647	101.89	179,844	774.95	23,618	101.77
chr11	205,530	708.21	26,260	90.49	205,642	708.59	26,155	90.12
chr12	192,219	698.17	26,248	95.34	192,225	698.19	26,117	94.86
Total	2,537,707	8245.67	360,444	1161.16	2,535,386	8238.62	359,804	1159.16

Table 6. Statistics of DNA polymorphisms in autotetraploid compared to diploid rice and reference genome.

Comparison groups	SNPs		InDels	
	Total	SNPs/100 kb	Total	InDels/100 kb
T446/E246	87,229	23.37	11,022	2.95
E246/Nipponbare	2,537,707	679.90	360,444	96.57
T446/Nipponbare	2,535,386	679.28	359,804	96.40

E246: Diploid rice; T446: autotetraploid rice.

Table 7. The number of structural variations (SVs) and copy number variations (CNVs) in diploid and autotetraploid rice.

Line	Total SVs	Insertion	Deletion	Inversion	ITX	CTX	Total CNVs	CNG	CNL
E246	17,171	2382	8875	406	441	5067	128	78	50
T446	17,375	2255	9087	407	443	5183	131	78	53

E246: Diploid rice; T446: autotetraploid rice; ITX: intrachromosomal translocation; CTX: interchromosomal translocation; CNG: copy number gained; CNL: copy number lost.

(Figure 3B). There were 2367, 838, and 1312 SNPs in the intron, untranslated region, and CDS region, respectively. In the CDS region, 519 SNPs led to synonymous mutations and 758 SNPs were related to nonsynonymous mutations. The nonsynonymous SNPs were involved in 331 genes, 6 genes were involved in start codon loss, and stop codon acquisition was involved in 17 genes. After excluding the duplicated genes, nonsynonymous SNPs, missense

mutations, loss of start codons, and STOP_GAINED mutations were involved in 369 variant genes, and 22 GO pathways were significantly enriched.

3.4. Candidate genes associated with embryo sac fertility and predicted protein-protein interactions in T446

Among the 369 mutated genes with high-effect mutations, 140 genes associated with embryo sac development were screened through the RiceXPro website to search for gene

Table 8. Annotation of structural variations (SVs) in diploid and autotetraploid rice.

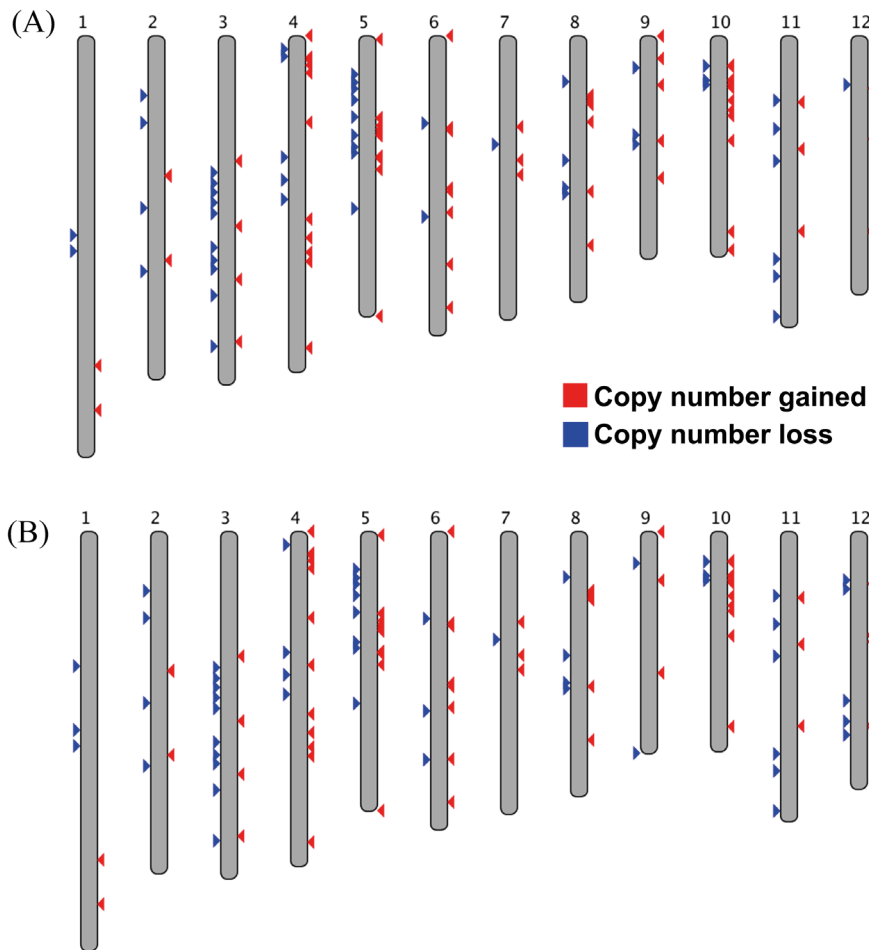
Type	Exon		Intron		Intergenic	
	E246	T446	E246	T446	E246	T446
DEL	666	687	1012	1047	7197	7353
INS	298	217	474	446	1610	1592
INV	234	228	17	16	155	163
Total	1198	1132	1503	1509	8962	9108

E246: Diploid rice; T446: autotetraploid rice; DEL: deletion; INS: insertion; INV: inversion.

expression profiles. The STRING website was used to predict the protein–protein interaction network of these 140 genes related to embryo sac fertility (Figure 4). Here, three interaction networks were detected, including an interaction between *LOC_Os03g38740* and 12 genes, an

interaction between *LOC_Os05g47770* and eight genes, and another interaction between *LOC_Os12g42070* and nine genes. Known interactions were screened out from the processed database or experimentally verified, and 13, 7, and 7 genes interacted with proteins/genes, respectively (Figure 5).

Of the 140 genes associated with embryo sac fertility, 10 genes are highly or specifically expressed in the ovary and may be closely related to the fertility of the embryo sac: *Os01g0880100*, *Os02g0292600*, *Os05g0331800*, *Os05g0332000*, *Os06g0565200*, *Os07g0219400*, *Os07g0216600*, *Os08g0127900*, *Os11g0512200*, and *Os11g0644900*. Among these, *Os02g0292600* and *Os06g0565200* were specifically expressed in the ovary. These 10 genes were functionally annotated and the results are shown in Table 9. *Os02g0292600* was annotated as an acyl hydrolase family protein and the product is a nuclear lipase. The expression products of *Os05g0331800*, *Os05g0332000*, and *Os07g0219400* were annotated as encoding alcohol-soluble protease precursor substances.

**Figure 2.** The distribution of copy number variations (CNVs) in diploid (A) and autotetraploid rice (B) along 12 rice chromosomes.

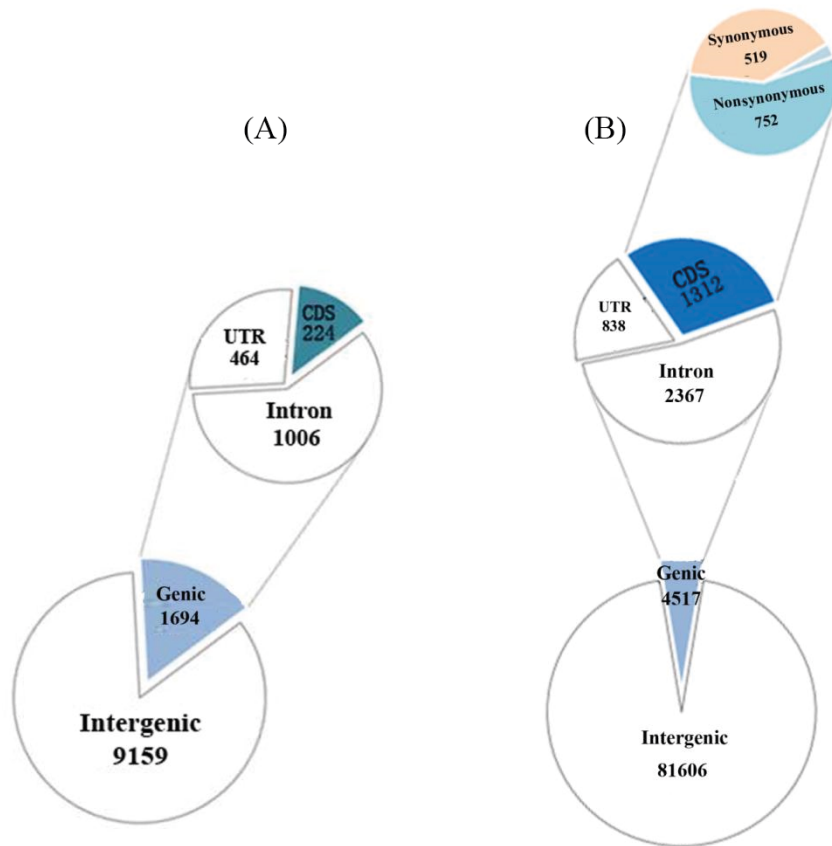


Figure 3. Statistics of DNA polymorphism annotation: A) InDel annotation between autotetraploid and its diploid rice; b) SNP annotation between autotetraploid rice and its diploid counterpart.

Os06g0565200 was annotated as encoding the DNA-binding domain protein of the heat shock transcription factor and the specific transcriptional regulator of the heat shock promoter element. *Os07g0216600* might encode an amylase and protein inhibitor of the seed, and its expression was detected in the aleurone layer of the seed. *Os08g0127900* was annotated as encoding gluten. *Os11g0644900* was annotated as encoding antimicrobial peptides that belong to a defensin-like gene family, and the expression product of *Os11g0512200* was apical meristem.

3.5. Embryo sac fertility and abnormalities in autotetraploid rice line

More than 650 mature embryo sacs of autotetraploid rice were observed and the results showed that the proportion of normal embryo sacs was very low. A total of 120 normal embryo sacs were found, with a frequency of 18.32% (Table 10). There were 530 abnormal embryo sacs, and various types of abnormalities were observed in mature embryo sacs including whole embryo sac degeneration (Figure 6A), absence of female reproductive unit (Figure

6B), abnormal position, and a number of polar nuclei (Figure 6C). Abnormal number of egg cells (Figure 6D) and egg apparatus degeneration (Figure 6E) were also found in the mature embryo sac. There were different types of abnormalities of antipodal cells in the mature embryo sacs (Figure 6F). Four polar nuclei and two egg apparatus were found in a mature embryo sac (Figure 6G), and no antipodal cell or abnormal development of antipodal cells in the embryo sac (Figure 6H). A mature embryo sac contains a group of antipodal cells, an egg cell, synergids, and two polar nuclei above the egg apparatus (Figure 6I).

4. Discussion

4.1. Agronomic trait variations in autotetraploid rice line

Autotetraploid rice exhibited significant differences in morphology and agronomic traits after genome duplication. After doubling, the autotetraploid rice was morphologically characterized by shorter plants, fewer tillers, thicker and harder leaves, and longer and wider grains than diploid rice. Resistance against lodging and

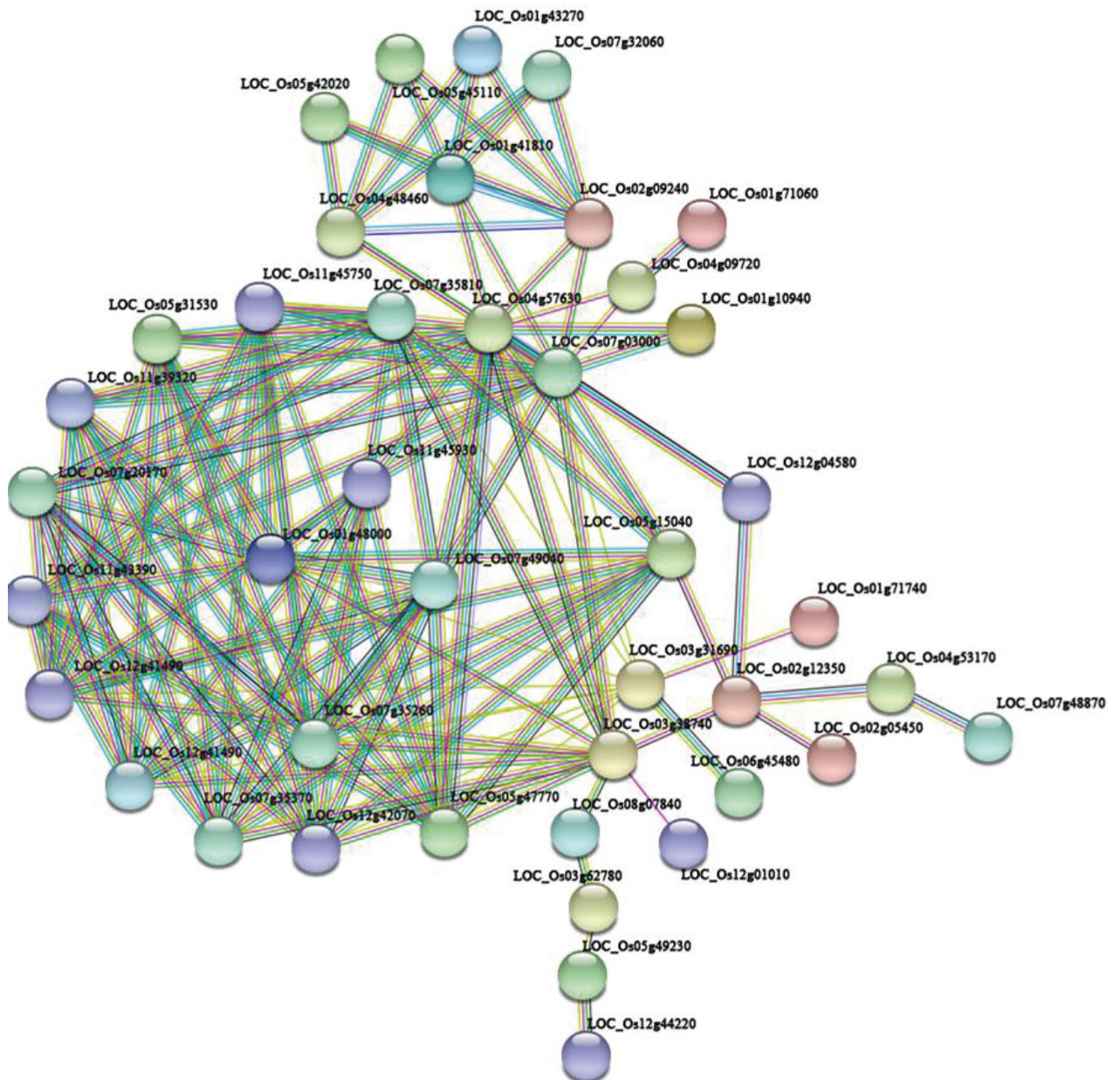


Figure 4. Predicted protein-protein interaction network of 140 genes associated with embryo sac fertility.

increased grain weight are favorable characters of polyploid rice that increase rice yield (Shahid et al., 2012, 2013b; Wu et al., 2013; Guo et al., 2017). However, autotetraploid rice has awns, and the presence of awns affects the propagation characteristics and edible characteristics of rice seeds (Song et al., 1992). The number of total grains and seed set rate decreased considerably, and the number of effective panicles is also small in autotetraploid rice. These factors

are important reasons for the low yield of autotetraploid rice (Shahid et al., 2013b). Compared to diploid, the grain length and grain width are significantly increased in autotetraploid rice, which is beneficial for production increase (Shahid et al., 2011; Wu et al., 2013). Here, the autotetraploid rice line exhibited greater panicle length and greater grain length and width, but low seed set.

4.2. The whole-genome polymorphisms in autotetraploid rice line

Whole-genome resequencing has become a reliable method to detect genetic variations in rice. Next-generation sequencing technologies have enabled the

resequencing of massive genomes, exploitation of polymorphic molecular markers, and identification of DNA variations (Huang et al., 2012). Previous studies have revealed variations in genomic sequences of different diploid rice lines and detected important genes (Huang

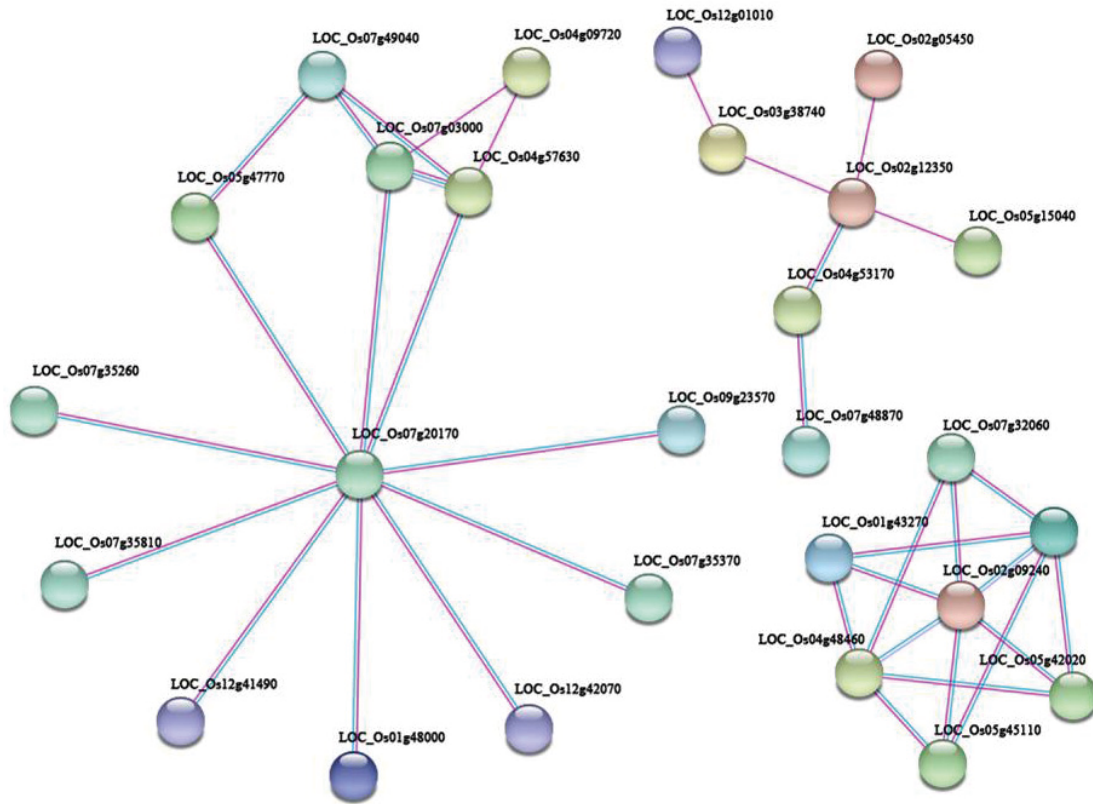


Figure 5. Known protein interaction network of genes associated with embryo sac fertility.

Table 9. Functional annotation of genes associated with embryo sac fertility.

Gene_ID		Annotation	Function
Os01g0880100	SNP	NON_SYNONYMOUS_CODING	Unknown
Os02g0292600	SNP	NON_SYNONYMOUS_CODING	GDSL-like lipase/acylhydrolase, putative, expressed
Os05g0331800	SNP	NON_SYNONYMOUS_CODING	PROLM17 - Prolamin precursor, expressed
Os05g0332000	SNP	NON_SYNONYMOUS_CODING	PROLM18 - Prolamin precursor, expressed
Os06g0565200	SNP	NON_SYNONYMOUS_CODING	HSF-type DNA-binding domain containing protein; transcriptional regulator that specifically binds DNA of HSE
Os07g0219400	SNP	STOP_GAINED	PROLM22 - Prolamin precursor, expressed
Os07g0216600	SNP	NON_SYNONYMOUS_CODING	LTPL163 - Protease inhibitor; LTP family protein precursor, expressed
Os08g0127900	SNP	NON_SYNONYMOUS_CODING	Glutelin, putative, expressed
Os11g0512200	SNP	NON_SYNONYMOUS_CODING	No apical meristem protein, putative
Os11g0644900	SNP	NON_SYNONYMOUS_CODING	DEFL34 - Defensin and defensin-like DEFL family

Table 10. Frequency of abnormal embryo sacs in autotetraploid rice.

Type	Number of cells	Frequency	Polar nuclei abnormalities			Abnormal egg apparatus			Antipodal cell abnormalities		Other abnormalities
			Abnormal number	Abnormal position	Degeneration	Multiple egg cells	Abnormal position	Degeneration	Located on the micropylar end	Degeneration	
Embryo sac degeneration	57	8.70%									
Normal	120	18.32%									
			287	57	4	39	3	27	34	19	8
			43.82%	8.70%	0.61%	5.95%	0.46%	4.12%	5.19%	2.90%	1.22%

et al., 2013; Jain et al., 2014; Fu et al., 2016). Recently, 29 million SNPs, 2.4 million small InDels, and over 90,000 SVs that lead to within- and between-population variation were detected in 3010 Asian cultivated rice lines (Wang et al., 2018). The polymorphisms between *Oryza sativa* (WAB56-104) and *Oryza glaberrima* (CG14) were detected using genome resequencing (Yamamoto et al., 2018). A total of 28 ancestral chromosomal blocks shared by all the high-yield cultivars were identified by whole-genome sequencing and pedigree analysis (Huang et al., 2018). Here, 2,535,390 SNPs and 359,805 InDels were detected in the whole genome of autotetraploid rice as compared to the Nipponbare reference genome. According to the statistical analysis of whole-genome variations between diploid and autotetraploid rice, 87,229 SNPs and 11,022 InDels were detected in autotetraploid compared to diploid rice, and there were 23.37 SNPs and 2.95 InDels per 100 kb. A total of 17,375 and 17,171 SVs were identified in the autotetraploid and its diploid counterpart and 2330, 3012, and 18070 SVs were detected in the exon, intron, and intergenic regions of both types of rice, respectively. Moreover, we detected 131 and 128 CNVs in autotetraploid and diploid rice, respectively. We inferred that mutation of the chromosome number may induce the changes in the genomic DNA nucleotide sequence, which then produces specific insertion and deletion mutations in polyploid rice.

4.3. Possible causes of poor seed set in autotetraploid rice
Embryo abortion is one of the most important reasons for the low fertility of autotetraploid rice. The major abnormalities of the embryo sac mainly include the deterioration of the embryo sac, an abnormal number of synergid cells, abnormal number and location of the eggs, and abnormal number of polar nuclei (Guo et al., 2006; Hu et al., 2009). During the embryo sac development of autotetraploid rice, the types of aborted mature embryo sacs were mainly degeneration of the embryo sac, abnormal number and location of polar nuclei, multiple egg cells, abnormal position of antipodal cells, and degeneration of the egg apparatus and antipodal cells, which is consistent with previous studies (Shahid et al., 2010; Li et al., 2017).

There are four sets of homologous chromosomes in autotetraploid rice that can undergo abnormal pairing during meiosis. Therefore, abnormal chromosome behavior was thought to be an important reason for low pollen fertility in autotetraploid rice (He et al., 2011a, 2011b; Wu et al., 2015). There are many studies about the pollen mother cell meiosis in autotetraploid rice (He et al., 2011a, 2011b; Wu et al., 2014, 2015; Li et al., 2017; Chen et al., 2018). Here, low pollen fertility (25.38%) was detected in an autotetraploid rice line, and different types of abortive pollens were detected, such as typical abortive pollens, stained abortive pollens, and empty abortive pollens. Similarly, various types of abortive pollens were observed in *indica* and *japonica* lines of autotetraploid rice

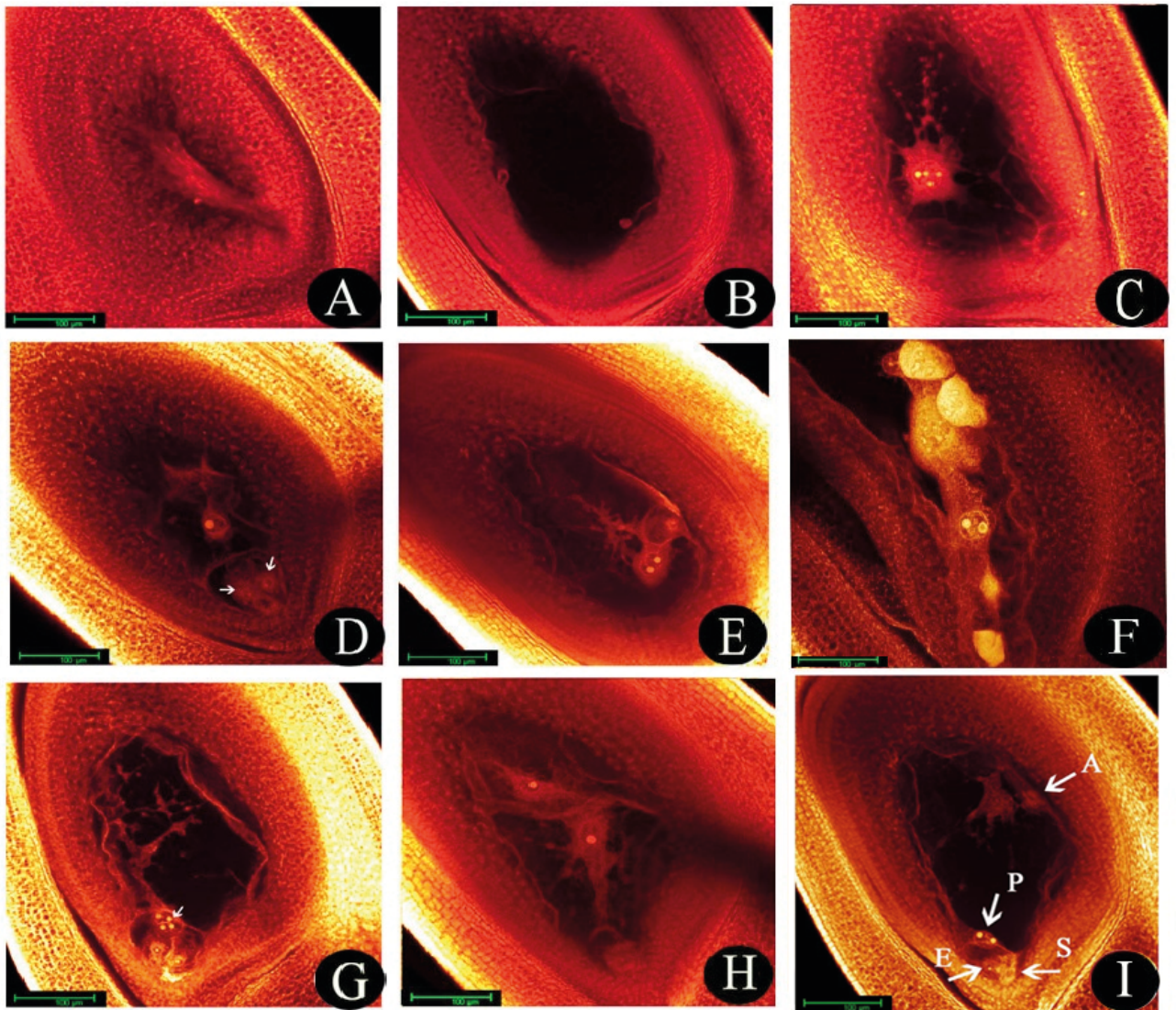


Figure 6. Mature embryo sac abnormalities in autotetraploid rice line: A) embryo sac degeneration; B) embryo sac without female germ unit; C) four polar nuclei near the center of embryo sac; D) two egg cells; E) egg apparatus degeneration, two polar nuclei are present; F) antipodal cells scattered in the whole embryo sac; G) four polar nuclei and two egg apparatus; H) no antipodal cells, two poles, and abnormal position of polar nuclei; I) normal mature embryo sac, where A, P, E, and S represent antipodal cells, polar nuclei, egg apparatus, and synergids, respectively.

(Shahid et al., 2010). Hence, low pollen fertility and embryo sac fertility resulted in low seed set in autotetraploid rice. This is consistent with the results of previous studies that concluded that low embryo sac and pollen fertilities cause low seed set in autotetraploid rice (Shahid et al., 2010; Wu et al., 2015; Li et al., 2017, 2018).

The low fertility and seed setting in autotetraploid rice may be caused by factors such as genomic variations, gene expression changes, and epigenetic recombination. The variation of the genes involved in embryo sac development and the sequence variation of the regulatory region

directly affect the structure and function of genes and their expression products (Zhang et al., 2014). Here, SNPs and InDel mutations in 140 genes or gene regulatory regions related to the development of the embryo sac were detected, including 10 genes that may be closely associated with the development of the embryo sac. Among the 369 genes that had a large effect on mutations, none were associated with embryo sac abortion in previous studies. Previously detected embryo sac abortion-related genes, including *OsRPA1a*, *ZIP4*, *OsRH36*, *OsDEES1*, *ORF3*, *OsMSH4*, *HSA1a*, and *OsIG1*, were compared with the diploid rice,

and three genes (*ORF3*, *OsMSH4*, and *HSA1a*) exhibited mutations in autotetraploid rice. The mutations of *ORF3* and *HSA1a* were located in the downstream region of the gene, and *HSA1a* was involved at the same site of diploid and autotetraploid mutation as a heterozygous locus. The mutation of *OsMSH4* was detected in the upstream region of the gene, and all mutations were homozygous.

5. Conclusions

In this study, an autotetraploid rice line with low set was investigated at morphological, cytological, and genomic levels. Abrupt changes in the expression patterns of genes associated with embryo sac development laid a foundation for the molecular mechanism associated with low embryo sac fertility and seed set in autotetraploid rice. We detected various types of abortive embryo sacs in autotetraploid rice. The mutation in the genes related to the development of the embryo sac might be the cause of abrupt gene expression patterns and the changes in

structure and function of the expression products. A total of 140 large-effect SNPs and InDel variants related to embryo sac development were detected within the gene or gene regulatory region in autotetraploid rice. The differentiation in SVs and CNVs may have also influenced the development of the embryo sac by gain/loss in the genes. Ten genes were likely associated with embryo sac fertility, and none of these genes had been reported for embryo sac fertility in previous studies. Further studies are required to functionally validate these genes.

Acknowledgments

The authors thank Professor Weiming Li for donating E246 (APIV) and Shuhong Yu for developing T446. The work of MQ Shahid was supported by the NSFC (31850410472) and the Natural Science Foundation of Guangdong Province of China (2017A030313142), and the work of XD Liu was supported by the Guangzhou Science and Technology Key Program (201707020015).

References

- Alp S, Ercisli S, Dogan H, Temim E, Leto A et al. (2016). Chemical composition and antioxidant activity *Ziziphora clinopodioides* ecotypes from Turkey. *Romanian Biotechnological Letters* 21: 11298-11303.
- Boeva V, Popova T, Bleakley K, Chiche P, Cappo J et al. (2012). Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28 (3): 423-425.
- Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM et al. (2009). BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* 6 (9): 677-681.
- Chen L, Shahid MQ, Wu J, Chen Z, Wang L et al. (2018). Cytological and transcriptome analyses reveal abrupt gene expression for meiosis and saccharide metabolisms that associated with pollen abortion in autotetraploid rice. *Molecular Genetics and Genomics* 293: 1407. doi: 10.1007/s00438-018-1471-0
- Chen L, Yuan Y, Wu J, Chen Z, Wang L et al. (2019). Carbohydrate metabolism and fertility related genes high expression levels promote heterosis in autotetraploid rice harboring double neutral genes. *Rice* 12: 34. doi: 10.1186/s12284-019-0294-x
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Genetics* 42: 443-461.
- Fu CY, Liu WG, Liu DL, Li JH, Zhu MS et al. (2016). Genome-wide DNA polymorphism in the *indica* rice varieties RGD-7S and Taifeng B as revealed by whole genome re-sequencing. *Genome* 59 (3): 197-207.
- Galiana-Belaguer L, Ibanez G, Cebolla-Cornejo J, Rosello S (2018). Evaluation of germplasm in Solanum section Lycopersicon for tomato taste improvement. *Turkish Journal of Agriculture and Forestry* 42: 309-321. doi: 10.3906/tar-1712-61
- Guo H, Mendrikahy JN, Xie L, Deng JF, Lu ZJ et al. (2017). Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Scientific Reports* 10: 40139. doi: 10.1038/srep40139
- Guo HB, Lu YG, Feng JH, Yang B, Liu XD (2006). Further observation on the formation and development of autotetraploid rice embryo sac using laser scanning confocal microscopy. *Acta Laser Biology Sinica* 15: 111-117 (in Chinese with English abstract).
- He JH, Shahid MQ, Chen ZX, Cheng XA, Liu XD et al. (2011a). Abnormal PMC microtubule distribution pattern and chromosome behavior resulted in low pollen fertility of an intersubspecific autotetraploid rice hybrid. *Plant Systematics and Evolution* 291: 257-265. doi: 10.1007/s00606-010-0386-y
- He JH, Shahid MQ, Li YJ, Guo HB, Cheng XA et al. (2011b). Allelic interaction of F₁ pollen sterility loci and abnormal chromosome behaviour caused pollen sterility in intersubspecific autotetraploid rice hybrids. *Journal of Experimental Botany* 62: 4433-4445. doi: 10.1093/jxb/err098
- Hu CY, Zeng YX, Lu YG, Li JQ, Liu XD (2009). High embryo sac fertility and diversity of abnormal embryo sacs detected in autotetraploid *indica*/*japonica* hybrids in rice by whole-mount eosin B-staining confocal laser scanning microscopy. *Plant Breeding* 128: 187-192.
- Huang J, Li J, Zhou J, Wang L, Yang S et al. (2018). Identifying a large number of high-yield genes in rice by pedigree analysis, whole-genome sequencing, and CRISPR-Cas9 gene knockout. *Proceedings of the National Academy of Sciences of the United States of America* 115 (32): E7559-E7567.
- Huang X, Lu T, Han B (2013). Resequencing rice genomes: an emerging new era of rice genomics. *Trends in Genetics* 29 (4): 225-232.

- Huang X, Zhao Y, Wei X, Li C, Wang A et al. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44: 32-39.
- Jain M, Moharana KC, Shankar R, Kumari R, Garg R (2014). Genome wide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant Biotechnology Journal* 12: 253-264.
- Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.
- Li X, Shahid MQ, Wu J, Wang L, Liu X et al. (2016). Comparative small RNA analysis of pollen development in autotetraploid and diploid rice. *International Journal of Molecular Science* 17: 499. doi: 10.3390/ijms17040499
- Li X, Shahid MQ, Xia J, Lu ZJ, Fang N et al. (2017). Analysis of small RNAs revealed differential expressions during pollen and embryo sac development in autotetraploid rice. *BMC Genomics* 18: 129. doi: 10.1186/s12864-017-3526-8
- Li X, Yu H, Jiao YM, Shahid MQ, Wu JW et al. (2018). Genome-wide analysis of DNA polymorphisms, the methylome and transcriptome revealed that multiple factors are associated with low pollen fertility in autotetraploid rice. *PLoS One* 13 (8): e0201854. doi: 10.1371/journal.pone.0201854
- Liu W, Ghouri F, Yu H, Li X, Yu SH et al. (2017). Genome wide re-sequencing of newly developed rice lines from common wild rice (*Oryza rufipogon* Griff.) for the identification of NBS-LRR genes. *PLoS One* 12: e180662. doi: 10.1371/journal.pone.0180662
- Liu W, Shahid MQ, Bai L, Lu Z, Chen Y et al. (2015). Evaluation of genetic diversity and development of a core collection of wild rice (*Oryza rufipogon* Griff.) populations in China. *PLoS One* 10 (12): e0145990. doi: 10.1371/journal.pone.0145990
- Luan L, Wang X, Long WB, Liu YH, Tu SB et al. (2008). Microsatellite analysis of genetic variation and population genetic differentiation in autotetraploid and diploid rice. *Biochemical Genetics* 46: 248-266.
- Masterson J (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264: 421-424.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K et al. (2010). The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297-1303.
- Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment* 32: 261-285. doi: 10.1080/13102818.2017.1400401
- Parisod C, Holderegger R, Brochmann C (2010). Evolutionary consequences of autopolyploidy. *New Phytologist* 186: 5-17.
- Shahid MQ, Chen FY, Li HY, Wang SZ, Chen FF et al. (2013a). Double-neutral genes, *Sa-n* and *Sb-n*, for pollen fertility in rice to overcome *indica* × *japonica* hybrid sterility. *Crop Science* 53: 164-176. doi: 10.2135/cropsci2012.07.0451
- Shahid MQ, Li YJ, Saleem MF, Wei CM, Naeem M et al. (2013b). Yield and yield components in autotetraploid and diploid rice genotypes (*indica* and *japonica*) sown in early and late seasons. *Australian Journal of Crop Science* 7: 632-641.
- Shahid MQ, Liu GF, Li JQ, Naeem M, Liu XD (2011). Heterosis and gene action study of agronomic traits in diploid and autotetraploid rice. *Acta Agriculturae Scandinavica Section B - Soil & Plant Science* 61: 23-32.
- Shahid MQ, Sun JF, Wei CM, Zhang P, Liu XD (2010). Studies on the abnormality of embryo sac and pollen fertility in autotetraploid rice during different growing seasons. *Pakistan Journal of Botany* 42: 7-19.
- Shahid MQ, Xu HM, Lin SQ, Chen ZX, Naeem M et al. (2012). Genetic analysis and hybrid vigor study of grain yield and other quantitative traits in autotetraploid rice. *Pakistan Journal of Botany* 44: 237-246.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH et al. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336-348.
- Song WC, Zhang YH (1992). Rice tetraploidy and its effect on agronomic traits and nutritional constituents. *Acta Agronomica Sinica* 18: 137-144.
- Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, Vaddadi S et al. (2012). Genome-wide DNA polymorphisms in elite *indica* rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal* 10: 623-634.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D et al. (2015). String v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43 (D1): D447-D452.
- Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557: 43-49.
- Wang Y, Shahid MQ, Ghouri F, Ercişli S, Baloch FS et al. (2019). Transcriptome analysis and annotation: SNPs identified from single copy annotated unigenes of three polyploid blueberry crops. *PLoS One* 14 (4): e0216299.
- Wu J, Chen L, Shahid MQ, Chen M, Dong Q et al. (2017). Pervasive interactions of *Sa* and *Sb* loci cause high pollen sterility and abrupt changes in gene expression during meiosis that could be overcome by double neutral genes in autotetraploid rice. *Rice* 10: 49. doi: 10.1186/s12284-017-0188-8
- Wu JW, Hu CY, Shahid MQ, Guo HB, Zeng YX et al. (2013). Analysis on genetic diversification and heterosis in autotetraploid rice. *Springer Plus* 2: 1-12.

- Wu JW, Shahid MQ, Chen L, Chen M, Dong Q et al. (2015). Polyploidy enhances F_1 pollen sterility loci interactions that increase meiosis abnormalities and pollen sterility in autotetraploid rice. *Plant Physiology* 169: 2700-2717. doi: 10.1104/pp.15.00791
- Wu JW, Shahid MQ, Guo HB, Yin W, Chen Z et al. (2014). Comparative cytological and transcriptomic analysis of pollen development in autotetraploid and diploid rice. *Plant Reproduction* 27: 181-196. doi:10.1007/s00497-014-0250-2
- Yaldız G, Çamlıca M, Nadeem MA, Nawaz MA, Baloch FS (2018). Genetic diversity assessment in *Nicotiana tabacum* L. with iPBS-retrotransposons. *Turkish Journal of Agriculture and Forestry* 42: 154-164. doi:10.3906/tar-1708-32
- Yamamoto N, Garcia R, Suzuki T, Solis CA, Tada Y et al. (2018). Comparative whole genome re-sequencing analysis in upland New Rice for Africa: insights into the breeding history and respective genome compositions. *Rice* 11: 33.
- Yu H, Shahid MQ, Li R, Li W, Liu W et al. (2018). Genome-wide analysis of genetic variations and the detection of rich variants of NBS-LRR encoding genes in common wild rice lines. *Plant Molecular Biology Reporter* 36: 618-630.
- Zhang L, Cheng Z, Qin R, Qiu Y, Wang JL et al. (2012). Identification and characterization of an epi-allele of FIE1 reveals a regulatory linkage between two epigenetic marks in rice. *Plant Cell* 24: 4407-4421.
- Zhang LM, Luo H, Liu ZQ, Zhao Y, Luo JC et al. (2014). Genome-wide patterns of large-size presence/absence variants in sorghum. *Journal of Integrative Plant Biology* 56: 24-37.
- Zheng LY, Guo XS, He B, Sun LJ, Peng Y et al. (2011). Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biology* 12: R114.
- Zhang P, Zhong K, Shahid MQ, Tong H (2016). Association analysis in rice: from application to utilization. *Frontiers in Plant Science* 7: 1202. doi: 10.3389/fpls.2016.01202