Extensive de novo Genomic Variation in Rice Induced by
Introgression from Wild Rice (*Zizania latifolia* Griseb.)

Yong-Ming Wang, Zhen-Ying Dong, Zhong-Juan Zhang, Xiu-Yun Lin, Ye Shen,
Daowei Zhou and Bao Liu*

Laboratory of Molecular Epigenetics and The State Key Laboratory of Grassland &
Ecology, Northeast Normal University, Changchun 130024, China
Running title: Introgression and genomic instability

Key words: Introgression, genomic variation, AFLP, rice, Zizania

Corresponding author: Bao Liu,
Laboratory of Molecular Epigenetics,
Institute of Genetics and Cytology,
Northeast Normal University,
Changchun 130024, China

E-mail: baoliu6677@yahoo.com.cn
Phone: 86-431-5099367
Fax: 86-431-5099285
ABSTRACT

To study the possible impact of alien introgression on a recipient plant genome, we examined over 6,000 unbiased genomic loci of three stable rice recombinant inbred lines (RILs) derived from intergeneric hybridization between rice (cv. Matsumae) and a wild relative (Zizania latifolia Griseb.) followed by successive selfing. Results from amplified fragment length polymorphism (AFLP) analysis showed that, whereas the introgressed Zizania DNA comprised less than 0.1% of genome content in the RILs, extensive and genome-wide de novo variations occurred in up to 30% of the analyzed loci for all three lines studied. The AFLP-detected changes were validated by DNA gel blot hybridization and/or sequence analysis of genomic loci corresponding to a subset of the differentiating AFLP fragments. A BLAST analysis revealed that the genomic variations occurred in diverse sequences including protein-coding genes, transposable elements and sequences of unknown functions. Pairwise sequence comparison of selected loci between a RIL and its rice parent showed that the variations represented either base substitutions or small insertion/deletions. Genome variations were detected in all 12 rice chromosomes, although their distribution was uneven both among and within chromosomes. Taken together, our results implicate that even cryptic alien introgression can be highly mutagenic to a recipient plant genome.
HYBRIDIZATION between genetically differentiated populations plays an important role in plant genome evolution, and can lead to speciation (Anderson and Stebbins 1954; Stebbins 1959; Grant 1981; Rieseberg 1995; Arnold 1997; Wendel 2000; Rieseberg et al. 2003; Seehausen 2004). Recent studies in several plant systems have demonstrated that plant allopolyploidization, or interspecific/intergeneric hybridization followed by genome doubling, is often accompanied by unorthodoxical genetic and epigenetic changes that transgress Mendelian principles (reviewed in Matzke et al. 1999; Wendel 2000; Rieseberg 2001; Pikaard 2001; Liu and Wendel 2002; Finnegan 2002; Levy and Feldman 2002; Comai et al. 2003; Paterson et al. 2003; Soltis and Soltis 2004; Levy and Feldman 2004; Adams and Wendel 2004; Madlung and Comai 2004). It has been suggested that these genomic changes may constitute a stabilizing mechanism essential for the establishment of the incipient allopolyploid as successful new species (Feldman and Levy 2005; Ozkan et al. 2001; Adams et al. 2003; Adams et al. 2004; Adams and Wendel 2005). Studies on newly synthesized F1 hybrids and their allopolyploid derivatives have enabled dissection of the roles played by hybridization or genome doubling in these genetic and epigenetic changes, and it has been found that in most cases hybridization is the main elicitor (Comai et al. 2000; Shaked et al. 2001; Ozkan et al. 2001; Han et al. 2003; Ma et al. 2004; Wang et al. 2004; Pontes et al. 2004; Madlung et al. 2005). It is therefore expected that wide hybridization may induce similar genomic changes at the diploid (or homoploid) level, although the extent of changes to which a diploid can tolerate may be much smaller than an allopolyploid owing to the genome-wide redundancy of the latter. Indeed, rapid genomic remodeling
was detected in diploid sunflower interspecific hybrids, which has enabled rapid ecological speciation (Rieseberg et al. 1995; Burke et al. 2004).

Introgressive hybridization (Anderson 1949), namely, wide hybridization followed by repeated backcrossing with one of the parental species, usually does not result in instant speciation, but has been an important means for transfer, or for de novo origination, of traits related to adaptation and other genetic diversity in natural plant populations (Arnold 2004). In plant breeding, introgression of uncharacterized DNA segments from a wild species into a cultivated variety is a commonly used approach. In such practices, emphasis is usually given to the transfer of desired genes (traits) from the donor species into the cultivar being improved, while the potential consequences of alien DNA integration on the structural and/or functional features of the recipient genome are largely ignored.

During the course of our effort to introduce useful genes from a sexually incompatible wild rice species, *Zizania latifolia* Griseb., into rice (*Oryza sativa* L.) by a novel sexual hybridization approach (Liu et al. 1999b; Shan et al. 2005), we observed that the resulting recombinant inbred lines (RILs) exhibited a wide range of heritable phenotypic variations such as changes in overall plant statue, growth periods, yield components, and disease resistance/susceptibilities (Shan et al. 2005; data not shown). These phenotypic variations were not readily accountable by cryptic introgression of *Zizania* genes per se, because these rice RILs derived from an atypical “F1” plant that had normal diploid chromosomes as indicated by the absence of additional chromosomes, chromosomal segments and gross aberrations (Liu et al. 1999b).
therefore suspected that genomic and/or epigenomic instability might have been elicited as a result of hybridization and cryptic introgression of the alien *Zizania* DNA. Indeed, we detected heritable and extensive alterations in DNA methylation patterns in the RILs in comparison with their rice parental cultivar, and these changes occurred to both protein-coding genes and transposon-related sequences (Liu *et al.* 2004). Similar disturbance of DNA methylation patterns induced by foreign DNA integration has also been observed earlier in animals (Heller *et al.* 1995; Remus *et al.* 1999; Muller *et al.* 2001). In addition to alterations in DNA methylation, at least two kinds of transposable elements, a *copia*-like retrotransposon *Tos17* (Hirochika *et al.* 1996) and a MITE transposon family *mPing* (Jiang *et al.* 2003; Kikuchi *et al.* 2003; Nakazaki *et al.* 2003) were apparently mobilized in the RILs (Liu and Wendel 2000; Shan *et al.* 2005). Taken together, what we observed in these rice RILs is reminiscent of the well-known phenomenon of hybrid dysgenesis in *Drosophila* and some other animals, which is characterized by multiple, interrelated genomic instabilities including mobilization of an array of transposons and various genomic rearrangements, which often lead to sterility and other morphological anomalies in the hybrids (Capy *et al.* 1990; Petrov *et al.* 1995; Evgen’ev *et al.* 1997; Capy *et al.* 2000; O’Nill *et al.* 1998; Brown *et al.* 2002).

To gain further insight into the effect of alien introgression on genome stability, we examined additional genome variations in the RILs and investigated the nature of these variations, using a genome-wide DNA fingerprinting approach based on amplified fragment length polymorphism (AFLP) analysis (Vos *et al.* 1995). Our results revealed
that introgression of a small amount of *Zizania* DNA had resulted in extensive sequence modifications of the recipient rice genome. This finding implicates additional potential role of alien introgression in genome evolution, and in its application in plant breeding.

**MATERIALS AND METHODS**

**Plant material:** The three rice recombinant inbred lines (RILs) used in this study, RZ1, RZ2 and RZ35, were derived from intergeneric hybridization between rice (cv. Matsumae) and a local accession of wild rice (*Zizania latifolia* Griseb.) by a novel sexual hybridization approach called “repeated pollination” (Liu *et al.* 1999b; Shan *et al.* 2005), and followed by successive selfing for 9-11 times (S9-S11). Each of the RILs is phenotypically homogeneous and exhibits heritable and novel morphological characteristics in multiple traits in comparison with their parental rice cultivar, Matsumae (Liu *et al.* 1999b), and hence represents a stabilized introgressant. The original crossing manipulations by “repeated pollination” to make the “F1” hybrid was performed in an isolated greenhouse where only plants of the recipient cultivar Matsumae and the wild rice (*Z. latifolia*) were grown. Successive selfing to construct the RILs, and their maintenance along with their exact parents and two sibling lines RZ36 and RZ60 were conducted by strictly controlled selfing under normal growing conditions (Liu *et al.* 1999b; Shan *et al.* 2005). These plant materials are available upon request for research purposes.

**AFLP analysis:** To explore the genomic composition of the RILs relative to their
rice and wild rice parents, we performed standard AFLP (amplified fragment length polymorphism) analysis (Vos et al. 1995), with minor modifications as previously described (Liu et al. 2001). Briefly, 400 ng of genomic DNA was digested with 1 U of MseI and 6 U of EcoRI (or HindIII) and simultaneously ligated to 50 pmol of MseI adaptors and 5 pmol of EcoRI (or HindIII) adaptors with 0.06 U T4 DNA ligase in 1 x T4 ligase buffer plus 50 mM NaCl and 50 ng BSA/L, in a total volume of 25 µL. The reaction was performed at 37°C for 3 h. The restriction–ligation reactions were diluted with 175 µL H2O prior to preselective polymerase chain reaction (PCR) amplification. PCR was performed using a set of MseI and EcoRI (or HindIII) primers (sequences available on request), with each primer having a single selective base at the 3’ end. Each PCR reaction (20 µL) contained 0.3 mM [MseI + 1] and 0.3 mM [EcoRI (or PstI) + 1] primers, 1.0 U of Taq DNA polymerase (Takara, Japan), 0.2 mM of dNTPs, and 4 µL of the diluted restriction–ligation reactions in 1 x PCR reaction buffer. The amplification profile was one cycle of 72°C for 2 min, followed by 20 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 2 min, plus one final extension at 60°C for 30 min. Ten µl of the PCR reaction was electrophoresed in 1.5% agarose gels and stained with ethidium bromide to verify preselective amplification, which typically gave a relatively uniform smear of amplification products in the size range of 100–1500 bp. The remaining 10 µL of PCR product was diluted with 150–200 µL of H2O prior to the selective PCR amplification using various primer combinations (information available on request). Profile for the selective amplification was 1 cycle of 94°C for 2 min, 1 cycle of 94°C for 30 s, 65°C for 30 s, and 72°C for 2 min, followed by nine cycles with
a 1.0°C decrease in annealing temperature per cycle, 35 cycles of 94°C for 30 min, 56°C for 30 s, and 72°C for 2 min, and a final extension at 60°C for 30 min. PCR products were separated in agarose gels, and the AFLP fragments were visualized by silver staining. Only those clear and reproducible bands that appeared in two independent PCR amplifications (starting from the digestion-ligation step, i.e., the first step of AFLP) were scored.

**DNA gel blot analysis:** Genomic DNA was isolated from expanded leaves of individual plants by a modified CTAB method (Kidwell and Osborn, 1992) and purified by phenol extractions. Genomic DNA (~3 µg per lane) was digested with *Eco*RI or *Hind*III (New England Biolabs Inc., Beverly, Massachusetts), separated in 1% agarose gels, and transferred onto Hybond N+ nylon membranes (Amersham Pharmacia Biotech., Piscataway, New Jersey) using the alkaline transfer buffer as recommended by the supplier. DNA fragments excised from the AFLP-gels were reamplified with appropriate pairs of primers used in the original AFLP amplifications, gel-purified, and used as probes. The DNA probes were labeled with fluorescein-11-dUTP using the Gene Images random prime-labeling module (Amersham Pharmacia Biotech). Hybridized membranes were washed in 0.2 x SSC, 0.1% SDS for 2 x 50 min, and exposed to x-ray films for 1-3 hrs depending on signal intensities. Hybridization signals were analyzed using the Gene Images CDP-Star detection module (Amersham Pharmacia Biotech).

**Sequencing:** Clear and reproducible AFLP fragments that showed deviation from additivity in the RIL(s) were eluted and reamplified using the same pair of
primers as in the original selective AFLP amplifications. The PCR products were
gel-purified and ligated to the TA cloning vector (Takara BioInc, Shiga, Japan), and
sequenced with universal vector primers. To examine whether the changed AFLP
patterns in a set of low-copy fragments (those could not be validated by gel-blot
analysis) were resultant from modifications at the restriction sites and/or adaptor
regions, locus-specific primers were designed, based on the available rice (cv.
Nipponbare) whole genome sequence, which bracketed the regions; to investigate the
nature of genomic changes at the primary DNA sequence level, sequence-specific
primers were designed against a set of selected AFLP fragments (within the sequenced
regions) using the Primer 3 program available at
http://biocore.unl.edu/cgi-bin/primer3/primer3_www.cgi, and used to amplify the
genomic sequences. The resulting PCR fragments were cloned and sequenced as
described above, but for both strands. Pairwise sequence comparisons were conducted
at the NCBI website using the BLAST2 program and confirmed by visual inspection.

RESULTS

The rice RILs contain less than 0.1% of Zizania DNA: The details for the
production of a series of rice lines with introgressed traits from the wild rice, Zizania
latifolia Griseb., by a novel approach called “repeated pollination”, and molecular
authentication of representative introgression lines (or RILs) by Zizania-specific DNA
repeats as probes have been described previously (Liu et al. 1999b; Shan et al. 2005). It
is important to note that rice and *Z. latifolia* are sexually incompatible, and therefore the so called “F1” plant is not a conventional hybrid; rather, it is a plant with a normal diploid rice genome plus genomically integrated DNA segments from *Z. latifolia*. The underlying mechanism for introgression remains unknown, but it may have similarities to that of gene transfer from irradiation-killed pollens observed in tobacco (Pandey 1975), and to that of introgression in sunflower from sexually incompatible wild relatives (Faure *et al*. 2002).

The RILs exhibited wide-ranging phenotypic variations (Liu *et al*. 1999b; Shan *et al*. 2005; unpublished data) in comparison with the rice parent (cv. Matsumae). However, AFLP analysis performed on ~2,000 loci for each of the three RILs (Table 1) indicated that the RILs contained *ca.* 0.1% DNA fragments putatively derived from *Z. latifolia* (Table 1). Only a fraction of these AFLP-derived fragments were unequivocally assigned as *Zizania*-specific, i.e., their absence in rice verses their presence in the RILs and their *Zizania* origin could be verified by DNA gel blot analysis (Shan *et al*. 2005). Thus, the amount of integrated *Zizania* DNA in the RILs was probably even less than 0.1%. This result suggests that the high degrees of phenotypic variations in the RILs were not a direct result of the integration of the *Zizania* DNA, instead, they were likely due to secondary genome modifications triggered by *Zizania DNA introgression.

**Extensive genomic variations occurred in the RILs as assessed by AFLP analysis:** Given the nature of the RILs (which have a complete set of diploid rice genome with a minute amount of genomically integrated *Zizania* DNA) and the
co-dominancy of AFLP markers it was expected that the RILs should contain all or most AFLP bands of the rice parent plus additional bands representing introgressed DNA segments from *Zizania*, with rare deviations (from additivity) occurring at and/or adjacent to the integration sites. It was therefore surprising that, of the ca. 2,000 distinct AFLP bands scored for each RIL, only ~70% were shared with the parental rice line. The remaining ~30% were either absent in, or unique to, one or more of the RILs, with loss of parental bands being more frequent than the gain of novel bands for all three RILs (Figure 1 and Table 1). The variations (loss or gain of bands) could be grouped into three different types: those being shared by all three RILs, those by two lines, and those unique to only one line (Figure 1).

**Validation of the AFLP-detected genomic variations by DNA gel blot analysis:**

To confirm that the changing patterns in the AFLP gels represent bona fide genomic variations rather than PCR amplification artifacts, we performed DNA gel blot analysis on the RILs. Thirty-one cloned AFLP fragments, which showed deviations from additivity in the RILs, were randomly selected and used to hybridize with *Eco*RI- or *Hin*dIII-digested genomic DNA. These 31 clones were subsequently characterized by sequence analysis (see below) as representing 10 genic sequences, nine transposon- or retrotransposon-related sequences and 12 sequences of unknown functions (data not shown).

Hybridization with 17 of the clones (seven genic, four TEs and seven with unknown function) showed clear genomic variations between the parental rice line and one or more of the three RILs. Examples of the hybridization patterns with the 17
clones are shown in Figure 2. Nine of the 31 clones did not give differentiating hybridization patterns in either EcoRI or HindIII digests. However, sequence analysis of the nine clones [locus-specific primers used to obtain these segments were designed based on the Nipponbare genomic sequence published at the Gramene (Ware et al. 2002) website (http://www.gramene.org), such that the enzyme restrictions sites and the adaptor regions were included] confirmed that they all contained changes either at the MseI restriction sites or in the selective nucleotides of the AFLP primers (see below), and therefore represented genuine genomic variations. The remaining five clones gave a smearing (faint or strong) hybridization patterns in both enzyme digests, hence precluding further analysis. Thus, at least 26 out of 31 differentiating AFLP bands were confirmed by gel blot hybridization or sequencing to represent bona fide genomic variations in the RILs.

**Genomic variations were not caused by parental heterozygosity, cross manipulation, or pollen contamination:** First, parental heterozygosity (though unlikely due to the self-pollinating property of rice and strict precautions taken during the crossing and maintenance of materials) could also result in pattern variations in the introgression lines. To address this possibility, we randomly selected eight individual plants of Matsumae and performed DNA gel blot analysis using all the 17 AFLP clones as probes. We observed perfect monomorphic hybridization patterns among all plants with each of the probes (Figure 3a and data not shown). This confirmed the homozygous nature of the parental rice line and ruled out the possibility that the observed genomic variations in the RILs were caused by parental heterozygosity.
Second, it was possible, though also unlikely, that the genomic variations could be induced by the cross manipulation per se (i.e. pollination of the emasculated rice [cv. Matsumae] stigma by pollens of *Zizania* followed by pollination again with Matsumae’s own pollens; Liu *et al.* 1999b). To address this concern, we analyzed two lines (RZ36 and RZ60) that were sibling to the RILs (derived from the same cross) but contained no evidence of *Zizania* introgression (Shan *et al.* 2005). Gel blot hybridization with the 17 AFLP clones mentioned above gave identical hybridization patterns between the 7 randomly selected individual plants of the two lines and the parental rice cultivar Matsumae (e.g. Figure 7). This result indicates that the cross manipulation per se was not responsible for the genomic variations.

A final concern was accidental pollen contamination from an unknown source. Nonetheless, this was deemed extremely unlikely based on the following grounds: First, because the original crossing was performed in an isolated greenhouse where only the recipient rice parent Matsumae and the *Zizania* plants were grown (see *Materials and Methods*), it was not possible for pollen contamination to occur at the “F1” generation. Second, because genomic variations occurred in all three studied RILs that were derived from independent “F2” plants, contamination would have to occur independently in all three RILs to cause the observed variations, which was apparently impossible given the precautions taken (see *Materials and methods*).

**Types of sequences involved in the genomic variations:** To characterize the sequences involved in the genomic variations in the RILs, we cloned and sequenced 500 AFLP fragments that showed deviation from additivity in one or more of the three
RILs, and obtained high-quality sequence reads for 466 clones (sequences available upon request). Based on a BlastX analysis performed at the NCBI website, these clones could be classified into four categories: 64 clones have homology to known-function genes, 68 are homologous to putative protein-coding genes, 97 represent transposon or retrotransposons, and the remaining 237 clones show homology to sequences of unknown functions (Table 2 and data not shown). This result indicates that nearly one-third of the genomic variations occurred in the coding regions of the RILs. This, together with the estimated 30% overall genomic variations (Table 1), suggests that about 9% of the genic regions in the RILs have undergone genomic changes.

**Nature of the genomic variations at the primary DNA sequence level:** To gain further insight into the nature of the genomic variations in the RILs, we designed locus-specific primers based on the sequences of 30 selected AFLP clones (16 of Matsumae origin and 14 of RILs origin), and used the primers to amplify the corresponding genome regions from both the specific RILs (that showed the changes) and the parental rice Matsumae. The resulting PCR fragments were cloned and sequenced for both strands, and 27 of the 30 clones gave unequivocal sequences that were then used in pairwise comparisons between the RILs and the parental line. The result, summarized in Table 4, indicates that nucleotide substitutions occurred in all analyzed sequences, whereas indels (insertion/deletion) occurred in 19 of the 27 sequences. There are 236 base substitutions and 260 deleted or inserted nucleotides, making the total number of changed nucleotides as 496, or ~6.4% of the 7,802 bp analyzed sequence. For the 236 base substitutions, there are apparently more transitions
than transversion (176 vs. 60). It should be noted that the overall frequency of nucleotide changes in the entire genome is likely to be lower than 6.4% because the analyzed sequences were from genome regions known to have changes in the RILs, based on the AFLP profiles. Nevertheless, the result clearly indicates that cryptic Zizania DNA introgression has been highly mutagenic to the recipient rice genome.

To estimate the overall genomic divergence of the RILs from their rice parent, cluster analysis was performed to measure the genetic distances using the AFLP markers, and a dendrogram was generated as shown in Figure 5. It is estimated that the genetic distances between Matsumae and the RILs (RZ1, RZ2 and RZ5) are 0.43, 0.45 and 0.46, respectively. This indicates that cryptic Zizania DNA introgression has resulted in extensive genomic variations that culminated to substantial genome-wide differentiation of the RILs from their parental genotype.

**Chromosomal distribution of genomic variations:** To examine whether the genomic variations spread uniformly across the genome or whether they tended to exist as clusters in certain regions representing “hot spots”, we mapped all 466 sequenced AFLP clones to the 12 rice chromosomes based on the IRGSP complete genome sequence of the japonica rice Nipponbare published at Gramene (Ware et al. 2002) website (http://www.gramene.org). As shown in Figure 4, for all three RILs, the changed sequences distributed in all 12 chromosomes. However, the distribution is not uniform, both in terms of the number of clones per chromosome and the distribution pattern of clones within a given chromosome. When all three lines are considered, chromosome 4 contained the highest mean number of clones (26.3), followed by
chromosome 2 (19.7), with chromosome 11 having the least (7.7) (Table 3). Taking into account of the difference in chromosome length, we calculated the average density of genomic changes for each chromosome (number of changed fragments per Mb of chromosomal DNA). The result (Table 3) shows that chromosome 4 contained the highest density of changed fragments (0.73/Mb), followed by chromosome 10 (0.55/Mb), with chromosome 11 having the lowest density of changed fragments (0.26/Mb). For each of the 12 chromosomes, there is a marked differential distribution of the changed fragments in all the RILs (Figure 4). For instance, the region between 21.0 and 26.2 Mb of chromosome 9 contained no changed sequences, the first 13.0 Mb region had a density of 0.15/Mb, and in the region between 13.0 and 21.0 Mb, the density increases to 0.88/Mb. This suggests that there have been apparent “hot spots” in the chromosomes for introgression-induced genomic variations. In addition, there are differential distributions for the types of changed sequences between chromosomes or between regions of a given chromosome. For instance, none of the 7-11 changed fragments in chromosome 8 of the three RILs represents known-function genes. Similarly, none of the 5-14 changes in chromosome 9 of the three RILs corresponds to known-function genes or transposon/retrotransposons. There was a clear regional differential distribution for the types of changed fragments in chromosome 10; its short-arm (10S) contained 10 changed transposon/retrotransposon fragments, while the long-arm (10L) contained none (Figure 4). This regional differential distribution might reflect the structural characteristics of chromosome 10, as it has been reported that 10S is rich in repetitive sequences (largely transposons and retrotransposons).
whereas 10L mainly contains expressed genes (The Rice Chromosome 10 Sequencing Consortium 2003).

DISCUSSION

Genomic variations in the RILs are induced by cryptic *Zizania* DNA integration: We have previously reported that introgression of *Zizania* DNA resulted in significant changes in epigenetic states of the rice RIL genome such as alteration in DNA methylation patterns (Liu et al. 2004) and mobilization of transposable elements (Liu et al. 2000; Shan et al. 2005). Here we show that this alien DNA integration also induces more fundamental changes in the recipient rice genome, involving extensive and widespread sequence modifications that affected up to 30% of the genomic loci. This is unorthodoxical and was unexpected given the diploid nature of RILs and the small amount (<0.1%) of integrated *Zizania* DNA. Nonetheless, evidence from our experiments is compelling. We analyzed a total of over 6,000 loci in the three RILs using AFLP, and validated the detected variations using gel blot hybridization and sequence analysis. We have ruled out the possibility that parental heterozygosity and cross manipulations are responsible for the observed genome changes. In addition, precaution was taken during the initial production and subsequent maintenance of the RILs to avoid pollen contamination from other rice cultivars: no other rice cultivars were allowed to grow in the isolated greenhouse where the initial crossing was made. Furthermore, each RIL being derived from an independent “F2” plant all showed
genomic variations (Figure 1 and 2; Table 1), which argues against possible pollen contamination during the maintenance stages (see Results). That integration of *Zizania* DNA has been the inducer of genomic variations is further supported by AFLP analysis of an independently produced asymmetric somatic nuclear hybrid (SH6) between a different recipient rice genotype (Zhonghua 8) and *Z. latifolia*, which shows that the ‘hybrid’ line contains a small amount of *Z. latifolia* DNA (Liu *et al.* 1999a) but extensive genome-wide changes that can not be attributable to the protoplast isolation and tissue culture process (unpublished data).

**Current stability of the genomic variation in the RILs and timing of their occurrence:** Because the RILs are phenotypically homogeneous by the 9th selfed generation (Liu *et al.* 1999b), we suspect that all the genomic variations are currently stabilized and homogeneous. Indeed, DNA gel blot analysis of six randomly selected individuals from two more consecutive generations (S10 and S11) gave monomorphic hybridization patterns for each of the three RILs and with all 17 AFLP probes (Figure 3b and data not shown), confirming the stabilization of the variations.

The observation that a large fraction of the genomic variations are shared by two or all three RILs (Figures 1, 2, 4 and quantitatively diagramed in Figure 6) suggests that the changes occurred at the very early stages, probably during the first or few cell divisions of the “F1 hybrid” zygote, which thereafter remained largely static and were transmitted through Mendelian segregation upon selfing. We theoretically tested this possibility using the following calculations: Assuming that all the genomic variations (including those not shared by the RILs) occurred at early developmental stages of the
“F1” hybrid and were independently inherited, the probability for a progeny or RIL to receive a given variation (a heterozygous locus) should be 1/2, and by extension, the probability for all three RILs to have a particular variation is (1/2)³. Among the ca. 2,000 scored bands, we detected a total of 173 variations that are shared by all three RILs (Figure 6). Thus, the expected number of variations that occurred in the “F1” hybrid is 1384 (173 x 8), and hence, 692 (1384÷2) in each of the homologous RILs. Statistical analysis indicates that this value is not significantly different from the actual mean number of variations detected for each RIL, which is 641 [(583 + 654 + 686)÷3] (Table 2 and Figure 6), thus supporting the notion that genomic variations occurred at very early stages subsequent to introgression.

Possible causes for the genomic variation: The mechanism by which Zizania DNA introgression induced extensive genomic variations in the rice genome is presently unknown. It is unlikely that the variations were caused by conventional or unorthodoxical meiotic recombination of homeoalleles (e.g., Osborn et al. 2003; Udall et al. 2005) between rice and Zizania, because even at the “F1” hybrid stage, the plant was found to have normal diploid rice chromosomes with no additional chromosomes or chromosomal segments. The mechanism may bear resemblance to those responsible for the genomic changes detected in several nascent interspecific F1 hybrids of Triticeae (Shaked et al. 2001; Ozkan et al. 2001; Han et al. 2003; Ma et al. 2004).

However, we have noted that two features distinguish the genomic variations in the rice RILs from those that occurred in the Triticeae F1 hybrids. First, whereas the changes observed in the rice RILs are largely stochastic, the changes in Triticeae are often
nonrandom and reproducible among independently synthesized F1 hybrids (Shaked et al. 2001; Ozkan et al. 2001; Han et al. 2003; Ma et al. 2004). Second, the changes in the rice RILs are predominantly base substitutions and small indels, while those in Triticeae are mainly physical deletions of relatively large genome segments (Shaked et al. 2001; Ozkan et al. 2001). In fact both the nature and the extent of the genomic variations in the rice RILs relative to their parental cultivar Matsumae, at the nucleotide sequence level, are rather similar to those found between the two rice subspecies, japonica and indica, (Edwards et al. 2004; Feltus et al. 2004; Ma and Bennetzen 2004), thus further pointing to the extensiveness of genomic divergence of the RILs from their original parental genotype.

Given the characteristics of the genomic variations in the rice RILs (genome-wide occurrence, representing diverse kinds of sequences, characterized by base substitutions and small indels), it is possible that they are the consequence of certain stochastic “mutator” effects, either directly transferred from Zizania and/or being activated by Zizania DNA introgression. In this respect, possible enhanced activities of some mobile genetic elements (transposons and retrotransposons) could be one of the causes for the genomic changes. Although no direct transpositional events were found among the sequenced AFLP fragments, many of the affected sequences are transposon- or retrotransposon-related (Table 2). Moreover, our previous work has shown that both a copia-like retrotransposon (Tos17) and members of a MITE family (mPing/Pong) are mobilized in these RILs (Liu and Wendel 2000; Shan et al. 2005), thus strengthening the possibility that transpositional activation of quiescent mobile elements was a cause
for the introgression-induced genomic variations. Another possibility is that alien introgression might have temporally titrated the activities of DNA repair enzymes and/or compromised the fidelity of DNA replication and/or cellular repair systems, which would undoubtedly result in various types of genomic variations (Greig et al. 2003; Capy et al. 2000). A final consideration is that the genomic structural variations may be directly related to, or even coupled with, the alterations in DNA methylation patterns induced by Zizania introgression in these lines (Liu et al. 2004), and hence, changes in chromatin structures. Conceivably, the disruption of intrinsic chromatin states by alien DNA insertion could affect genomic stability in myriad ways (Comai 2000; Comai et al. 2003; Madlung and Comai 2004). For instance, it has been shown that the timing and fidelity of DNA replication are related to DNA methylation states (Selig et al. 1988; Knox et al. 2000; Ros and Kunze 2001; Ehrenhofer-Murray 2004). Thus, extensive alteration in DNA methylation patterns of the rice RILs (Liu et al. 2004) could compromise the precision of DNA replication and produce “errors” at the nucleotide sequence level in the daughter cells, leading to the observed genomic changes.

**Implications of introgression-induced genomic variation on plant genome**

**Evolution and breeding:** It has long been recognized that hybridization and introgression occur widely among natural plant populations (Anderson 1949; Anderson and Stebbins 1954; Stebbins 1959; Grant 1981). Empirical studies on hybrid speciation with molecular tools conducted in the past decade have greatly strengthened the earlier conceptual hypothesis that hybridization plays a critical role in genome evolution and
speciation, at least in angiosperms (Rieseberg 1995; Arnold 1997; Wendel 2000; Rieseberg et al. 2003; Cronn and Wendel 2003; Doyle et al. 2004; Arnold 2004; Seehausen 2004). Although introgressive hybridization does not necessarily lead to immediate hybrid speciation, it is an important means for the transfer and/or de novo origination of traits related to ecological adaptations, and therefore plays an important role in facilitating speciation in changing niches (Arnold 2004). Our study shows that even cryptic alien introgression (< 0.1%) of *Zizania latifolia* DNA could function as a potent mutagen and induced extensive and genome-wide de novo variations affecting up to 30% of the genomic loci in the rice RILs. Although this specific introgression was accomplished artificially, it is conceivable that the simple process of “repeated pollination” (Liu et al. 1999b; Shan et al. 2005), used to construct the RILs, may occur under natural circumstances. Thus, our findings may have implications for introgression-facilitated genome evolution and speciation. It is possible that a key role of introgressive hybridization is to generate extensive stochastic genomic and epigenomic variations that can be translated into phenotypic novelties, and upon which natural selection may act. In this regard, we have noted that the rice RILs showed a wide range of phenotypic features including traits (e.g., changes in flowering time, fertility and disease tolerance) that are relevant to adaptation. Our current and previous (Liu et al. 2004; Shan et al. 2005) studies also have implications for the underlying mechanism(s) by which novel transgressive traits are expressed (Devicente and Tanksley 1993) in wide hybridization-derived breeding materials. It is likely that some of these novel traits are the result of introgression-induced genetic or epigenetic
variations. Further studies are needed to examine whether or not the introgression-induced changes are a widespread phenomenon and to understand the mechanisms by which these changes are induced.
ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (30430060), the National Science Award for Distinguished Young Investigators in China (30225003) and the State Key Basic Research and Development Plan of China (2001CB1088). We thank Drs. Ming-Bo Wang of CSIRO, Australia and Jun-Kang Rong of University of Georgia, USA, as well as an anonymous reviewer for critical reading and constructive comments to improve the manuscript.


BURKE, J. M., Z. LAI, M. SALMASO, T. NAKAZATO, S. TANG, A. HEESACKER,


EVGENE'V, M. B., H. ZELENTSOVA, N. SHOSTAK, M. KOZITSINA, V. BARSKYI,


HELLER, H., C. KAMMER, P. WILGENBUS, and W. DOERFLER, 1995. The


Plant Physiol. 130: 1606-1613.

TABLE 1
Genomic variation in the rice RILs based on AFLP analysis

<table>
<thead>
<tr>
<th>RIL</th>
<th>Total no. bands scored</th>
<th>No. &amp; frequency (%) of bands shared with the rice parent (cv. Matsumae)</th>
<th>No. &amp; frequency of bands showed deviation from the rice parent (cv. Matsumae)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Loss</td>
</tr>
<tr>
<td>RZ1</td>
<td>1,981</td>
<td>1,398 (70.6%)</td>
<td>453 (22.9%)</td>
</tr>
<tr>
<td>RZ2</td>
<td>2,100</td>
<td>1,446 (68.9%)</td>
<td>405 (19.3%)</td>
</tr>
<tr>
<td>RZ35</td>
<td>2,098</td>
<td>1,412 (67.3%)</td>
<td>439 (20.9%)</td>
</tr>
</tbody>
</table>
TABLE 2
Classification of DNA fragments showing genomic variation in the RILs based on AFLP analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of clones</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known-function gene</td>
<td>64</td>
<td>13.6</td>
</tr>
<tr>
<td>Putative protein-coding gene</td>
<td>68</td>
<td>14.5</td>
</tr>
<tr>
<td>Transposon &amp; retrotransposon</td>
<td>97</td>
<td>20.6</td>
</tr>
<tr>
<td>No similarity</td>
<td>237</td>
<td>51.3</td>
</tr>
<tr>
<td>Total</td>
<td>466</td>
<td>100</td>
</tr>
</tbody>
</table>
### TABLE 3
Chromosomal distribution of DNA fragments showing genomic variation in the RILs based on AFLP analysis

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Size (Mb)</th>
<th>No. changed fragment in</th>
<th>Total no. changed fragment</th>
<th>Average no. changed fragment</th>
<th>Variation density (No./Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RZ1</td>
<td>RZ2</td>
<td>RZ35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>45.7</td>
<td>11</td>
<td>18</td>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>38.8</td>
<td>23</td>
<td>23</td>
<td>13</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>12</td>
<td>25</td>
<td>18</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>24</td>
<td>36</td>
<td>19</td>
<td>79</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>11</td>
<td>13</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>33.4</td>
<td>10</td>
<td>17</td>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>30.9</td>
<td>14</td>
<td>13</td>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>29.9</td>
<td>9</td>
<td>11</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>9</td>
<td>26.2</td>
<td>8</td>
<td>14</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>23</td>
<td>12</td>
<td>17</td>
<td>9</td>
<td>38</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>12</td>
<td>31</td>
<td>15</td>
<td>18</td>
<td>10</td>
<td>43</td>
</tr>
</tbody>
</table>
TABLE 4
Nature of genomic variations in the RILs at the primary nucleotide sequence level

<table>
<thead>
<tr>
<th>Clone</th>
<th>Source of sequence</th>
<th>Location</th>
<th>Length sequenced (bp)</th>
<th>No. of nucleotide substitution (bp)</th>
<th>No. &amp; length of indel</th>
<th>Deletion</th>
<th>Insertion</th>
<th>No. bp</th>
<th>No. bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF 18</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 2</td>
<td>217</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 41</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 10</td>
<td>209</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 49</td>
<td>Matsumae vs. RZ1</td>
<td>Chr. 12</td>
<td>242</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 72</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 4</td>
<td>150</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 81</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 5</td>
<td>259</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 92</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 10</td>
<td>191</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>AF 101</td>
<td>Matsumae vs. RZ1</td>
<td>Chr. 6</td>
<td>165</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 103</td>
<td>Matsumae vs. RZ35</td>
<td>Chr. 4</td>
<td>212</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 133</td>
<td>Matsumae vs. RZ1</td>
<td>Chr. 3</td>
<td>218</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 135</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 11</td>
<td>193</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 156</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 3</td>
<td>398</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 158</td>
<td>Matsumae vs. R2</td>
<td>Chr. 2</td>
<td>340</td>
<td>43</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 161</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 10</td>
<td>263</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 197</td>
<td>Matsumae vs. RZ1</td>
<td>Chr. 7</td>
<td>274</td>
<td>3</td>
<td>1</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 207</td>
<td>Matsumae vs. R2</td>
<td>Chr. 7</td>
<td>269</td>
<td>27</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>AF 210</td>
<td>Matsumae vs. RZ2</td>
<td>Chr. 12</td>
<td>80</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 212</td>
<td>Matsumae vs. RZ1</td>
<td>Chr. 7</td>
<td>156</td>
<td>11</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 278</td>
<td>Matsumae vs. RZ1</td>
<td>Chr. 8</td>
<td>153</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 283</td>
<td>Matsumae</td>
<td>Chr. 11</td>
<td>592</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chr.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*AF 341 Matsumae vs. RZ2</td>
<td>3</td>
<td>412</td>
<td>2</td>
<td>1</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 352 Matsumae vs. RZ2</td>
<td>5</td>
<td>619</td>
<td>26</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 387 Matsumae vs. RZ35</td>
<td>7</td>
<td>555</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 418 Matsumae vs. RZ1</td>
<td>10</td>
<td>154</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 477 Matsumae vs. RZ35</td>
<td>11</td>
<td>206</td>
<td>14</td>
<td>2</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 480 Matsumae vs. RZ2</td>
<td>2</td>
<td>419</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 499 Matsumae vs. RZ1</td>
<td>6</td>
<td>301</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF 609 Matsumae vs. RZ1</td>
<td>7</td>
<td>555</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sum.</strong></td>
<td></td>
<td><strong>7,802</strong></td>
<td><strong>236</strong></td>
<td><strong>14</strong></td>
<td><strong>55</strong></td>
<td><strong>18</strong></td>
<td><strong>205</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* From *Z. latifolia*
Figure Legend

FIGURE 1 Examples of genomic variation in the three rice recombinant inbred lines or RILs (RZ1, RZ2 and RZ35) as compared with their rice parent (cv. Matsuame) detected by AFLP analysis with primer combinations (from left to right) EcoRI+AAG /MseI +CAA, EcoRI+AAG /MseI + CAG and EcoRI+AAG /MseI + CAT respectively. The arrows, arrowheads and circles respectively refer to parental bands disappeared in one or more RILs, novel bands appeared in one or more RILs and bands putatively from the donor species Zizania latifolia.

FIGURE 2 Examples of validation of the genomic variations in the RILs by DNA gel blot analysis. Hybridization of each of four isolated AFLP-derived bands to a blot containing EcoRI- or HindIII-digested genomic DNAs from the rice parental cultivar Matsumae (Lane 1), RILs RZ1 (Lane 2), RZ2 (Lane 3), RZ35 (Lane 4) and the wild species Zizania latifolia (Lane 5). Panels (a), (b), (c) and (d) are respectively AF206, AF34, AF480 and AF153. Arrows and circles respectively refer to rice parental bands disappeared in one or more of the RILs and novel bands appeared in one or more of the RILs.

FIGURE 3 Examples of homogeneity within the rice parent cultivar Matsumae (a) and the RILs (b). Hybridization of probe AF206 on HindIII-digested genomic DNA isolated from eight random individual plants respectively from Matsumae and RZ35. Similar monomorphic hybridization patterns were observed for Matsumae and the three RILs in all 17 selected AFLP fragments used as probes.

FIGURE 4 Examples of genomic stability in two rice lines (RZ36 and RZ60)
sibling to the three RILs (RZ1, RZ2 and RZ35) but with no introgressed DNA from Z. *latifolia*, as revealed by DNA gel blot analysis. Hybridization of probe AF206 on *Hind*III-digested genomic DNA isolated from seven random individual plants (lanes 2 through 8) from each of the RZ36 and RZ60 revealed monomorphic patterns as the rice parent Matsumae (lane 1). Similar results were obtained for all 17 selected AFLP fragments that showed conspicuous genomic variations in one or more of the three RILs (e.g., Figure 2).

**FIGURE 5** A dendrogram derived from UPGMA cluster analysis using the Jaccard’s coefficient of similarity calculated on the AFLP markers to show the overall genomic differentiation of the RILs from their rice parental cultivar Matsumae

**FIGURE 6** Distribution of the genomic variations affecting different types of sequences (depicted as different colors) within and among the 12 rice chromosomes in the RILs (from left to right RZ1, RZ2 and RZ35). Predicted functions of the mapped sequences are based on BLAST X analysis at the NCBI website (http://www.ncbi.nlm.nih.gov/), whereas chromosomal position of the variations is based on BLASTN analysis at the Gramene website (http://www.gramene.org/) (Ware *et al.* 2002).

**FIGURE 7** Diagrammatic illustration of the number of genomic variations that are shared by all three RILs (173), by two of the RILs, RZ1 and RZ2 (173 +107), RZ1 and RZ35 (173 + 77) and RZ2 and RZ35 (173 + 189), as well as unique to each RIL, RZ1 (226), RZ2 (185) and RZ35 (274).
Known-function genes
Predicted protein-encoding
Transposons and retrotransposons
Unknown

From *Z. latifolia*