Genomewide Variation in an Introgression Line of Rice-*Zizania* Revealed by Whole-Genome re-Sequencing

Zhen-Hui Wang[1,2◯], Di Zhang[1◯], Yan Bai[1], Yun-Hong Zhang[1], Ying Liu[1], Ying Wu[1], Xiu-Yun Lin[3], Jia-Wei Wen[3], Chun-Ming Xu[1], Lin-Feng Li[1*], Bao Liu[1*]

1 Key Laboratory of Molecular Epigenetics of Ministry of Education (MOE) and Institute of Genetics and Cytology, Northeast Normal University, Changchun, China, **2** Faculty of Agronomy, Jilin Agricultural University, Changchun, China, **3** Jilin Academy of Agricultural Sciences, Changchun, China

Abstract

Background: Hybridization between genetically diverged organisms is known as an important avenue that drives plant genome evolution. The possible outcomes of hybridization would be the occurrences of genetic instabilities in the resultant hybrids. It remained under-investigated however whether pollination by alien pollens of a closely related but sexually "incompatible" species could evoke genomic changes and to what extent it may result in phenotypic novelties in the derived progenies.

Methodology/Principal Findings: In this study, we have re-sequenced the genomes of *Oryza sativa* ssp. *japonica* cv. Matsumae and one of its derived introgressant RZ35 that was obtained from an introgressive hybridization between Matsumae and *Zizania latifolia* Griseb. in general, 131 millions 90 base pair (bp) paired-end reads were generated which covered 13.2 and 21.9 folds of the Matsumae and RZ35 genomes, respectively. Relative to Matsumae, a total of 41,724 homozygous single nucleotide polymorphisms (SNPs) and 17,839 homozygous insertions/deletions (indels) were identified in RZ35, of which 3,797 SNPs were nonsynonymous mutations. Furthermore, rampant mobilization of transposable elements (TEs) was found in the RZ35 genome. The results of pathogen inoculation revealed that RZ35 exhibited enhanced resistance to blast relative to Matsumae. Notably, one nonsynonymous mutation was found in the known blast resistance gene *Pid3/Pi25* and real-time quantitative (q) RT-PCR analysis revealed constitutive up-regulation of its expression, suggesting both altered function and expression of *Pid3/Pi25* may be responsible for the enhanced resistance to rice blast by RZ35.

Conclusions/Significance: Our results demonstrate that introgressive hybridization by *Zizania* has provoked genomewide, extensive genomic changes in the rice genome, and some of which have resulted in important phenotypic novelties. These findings suggest that introgressive hybridization by alien pollens of even a sexually incompatible species may represent a potent means to generate novel genetic diversities, and which may have played relevant roles in plant evolution and can be manipulated for crop improvements.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: lilf241@nenu.edu.cn (LL); baoliu@nenu.edu.cn (BL)

◯ These authors contributed equally to this work.

Introduction

It has been established that hybridization between genetically differentiated species or ecotypes may induce structural genomic changes (e.g., rearrangement) in the hybrids [1-7]. To address the genetic consequences of inter-specific hybridization, a larger number of studies have investigated the genetic variation patterns and molecular modes of the resultant hybrids. For example, Zou et al. [8] constructed a population of recombinant inbred lines of the new introgressed type of *Brassica napus* (A^r^A^r^C^n^C^n^) through two rounds of inter-specific hybridization between *B. napus* (A^n^A^n^C^n^C^n^) and *B. rapa* (A^r^A^r^). As expected, a remarkable range of novel genomic alterations occurred immediately after inter-specific hybridization and some of these genetic changes have profound effects on the yield-related traits in the introgressed *B. napus*. In addition, hybridization may also constitute a "genome shock" and lead to widespread activation of otherwise quiescent transposable elements (TEs) [9]. For instance, Yasuda et al. [10] have examined the mobilizations of

mPing in 17 introgression rice lines that were derived from the introgressive hybridization between *Oryza sativa* and *O. glaberrima*, and illustrated that introgression of genomic DNA from the wild species provoked transpositional activation of *mPing*.

To date, accumulated evidence showing that extensive genetic variations in resultant hybrids has led to the proposition that hybridization can rapidly generate novel phenotypes and hence contribute to the rapid origin of complex adaptations and evolutionary novelties, and ultimately increases overall biodiversity [11-14]. However, diploid (homoploid) hybrids of different species often face a seemingly intangible enigma, that is, the different parental chromosomes are unable to pair correctly during meiosis, and hence lead to sterility [15,16]. This problem can be resolved by whole genome doubling leading to allopolyploid formation, and this may explain why hybridization and polyploidy are often associated in plants [17,18]. However, a great majority of the allopolyploidization events would be met with extinction or an evolutionary dead-end [19] due to "minority cytotype disadvantage" among other problems [20,21].

Nevertheless, the pollination by alien species' pollens may occur frequently in nature because congeneric species usually display parapatric or sympatric distributions and share common or similar pollinators [22-26]. Apart from *bona fide* hybridization, the germinating alien pollens may constitute a "biological stress" to the recipient stigma genome due to its mechanistic resemblance to pathogen attack wherein the foreign DNA usually does not integrate into the host genome, yet its interaction with the host could provoke genetic and epigenetic instabilities of the host genome [27-30]. Therefore, in contrast to the hybridization and polyploidy, alien pollen-induced variations may represent an underlying mechanism that can also contribute to plant genome evolution.

To address the possible genomic impact of alien pollens on recipient plant genomes, three introgressant lines (RZ1, RZ2 and RZ35) were generated via a "repeated pollination" method [31]. In brief, the panicles of cv. Matsumae were emasculated and pollinated with the pollens of *Z. latifolia*. Forty to fifty hours later, the panicles of Matsumae were pollinated for the second time by its own pollens. All of the introgression lines were derived from a single "F₁" plant and then propagated by strict selfing. In a series of previous studies, we have shown that although a small amount of chromatin of *Z. latifolia* (< 0.1%) have been integrated into the rice ssp. *japonica* cv. Matsumae genome by the aforementioned method, these introgressants not only showed extensive *de novo* genetic changes but also exhibited morphological and physiological novelties [30-36]. These studies have provided compelling albeit circumstantial evidence showing that alien pollens can indeed provoke genetic and epigenetic instabilities of a recipient plant genome. Nonetheless, all previous studies are based on traditional molecular marker analysis, and hence, the analyzed loci are limited and potentially biased.

To gain further insights into the genomic effects of alien introgression from an unbiased, genomewide perspective, we have re-sequenced the genomes of a typical introgressant RZ35 along with its rice parental cultivar (Matsumae) using the next generation sequencing technology (Illumina HiSeq 2000). The specific aims of this investigation were: (1) to unravel the types and spectra of the induced genomic changes in the introgression line RZ35; (2) to investigate whether the induced genomic changes occurred randomly across the genome or in a locus-specific fashion in RZ35; and (3) to address to what extent the genomic variations might be linked to phenotypic novelties manifested in RZ35.

Materials and Methods

Plant materials, DNA extraction and whole genome re-sequencing

A stabilized introgression line RZ35 (selfed for > 20 generations) derived from introgressive hybridization between rice (*Oryza sativa* ssp. *japonica* cv. Matsumae) and a sexually incompatible wild species of the tribe *Oryzeae*, *Zizania latifolia*, was produced by us [31,33]. Genomic DNAs were extracted from leaf tissue of RZ35 and its recipient rice parent cv. Matsumae using a modified CTAB method [37], and purified by phenol extractions. Quality of the DNA samples was determined by a ND-1000 NanoDrop spectrophotometer (Eppendorf, Germany). Approximately 350 µg of genomic DNA of each line was used for Illumina resequencing by the HiSeq 2000 sequencer.

Pathogen inoculation

One-month-old seedling-plants of Matsumae and RZ35 (at the 4th to 5th leaf stage) were inoculated with blast fungus (*Magnaporthe grisea* (Hebert) Barr) according to previously reported procedures [38]. The inoculated seedlings were placed in an auto-controlled incubator for 24 hours in darkness under the conditions of 28°C and >95% relative humidity.

Identification of genomic changes in RZ35

For all data sets, only reads ≥ 90 base pair (bp) in length were evaluated. Alignment of Matsumae raw data were initially screened against the Nipponbare reference genome (MSU7.0 http://rice.plantbiology.msu.edu/index.shtml) using Burrows-Wheeler Aligner [39]. Those reads that showed either multiple mapping to the reference genome of Nipponbare or have mapping quality scores less than 30 (Phred scale) were removed from the assembled data. Thereafter, we employed Perl scripts to produce a theoretical "Matsumae reference" genome by including the sequences containing variations in comparison with Nipponbare. The remaining reads of RZ35 were then aligned to assembled Matsumae reference genome to detect homozygous SNPs, INDELs using SAMtools (v0.1.5c). In order to minimize the false-positive rates of SNPs and INDELs, the filter method was applied with the following parameters: Phred score ≥ 30, coverage ≥ 10 and ≤ 100.

Distribution patterns of genomic changes and identification of mutated genes

To infer the genome-wide pattern of variations and identify the HMRs in the RZ35 genome relative to that of Matsumae, a sliding window with 100 kilo base pair (kb) interval was applied

to calculate polymorphisms for each chromosome using Perl scripts and the figures were generated by ggplot2 package in software R. Consequently, the locations of HMRs on each chromosome of RZ35 were compared with the domestication-related regions of rice that were identified previously [40]. In addition, to estimate the base substitution rates, we calculated the ratio of transition (Ts) and transversion (Tv) for each chromosome of RZ35. Furthermore, we also performed a genome scan to identify the positions of these genomic changes. Specifically, the Generic Feature Format Version 3 (GFF3) files of annotated Nipponbare reference were summarized and formatted to a two-dimensional array file that includes exons, introns and ATG locations of each gene. Then, the locations of each genomic variation were determined by calculating the distance between variations and ATG start codon. Finally, the nonsynonymous mutations were identified through the comparisons of protein sequences between RZ35 and Matsumae. The protein sequences of these mutated genes were retrieved and subsequently subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) to assign their positions in each pathway [41]

Identification and characterization of transposable elements (TEs) mobilization

Our previous studies have identified transpositional activation or mobilization of several TEs in the rice-*Zizania* introgression lines including RZ35 [30,33,34]. To further investigate if mobilization of additional TEs or mobilization sites of previously identified TEs in the RZ35 genome, we applied a PEM detection strategy (Figure S1) as employed previously by Sabot et al. [42]. In brief, the raw data of Matsumae and RZ35 were aligned against the Nipponbare reference genome (MSU7.0 http://rice.plantbiology.msu.edu/ index. shtml) using bowtie software [43]. Then, all of the non-matching reads were searched against the known sequences of TEs (such as *Tos17*, *Tos19* and *mPing* etc.) using software bowtie again. These reads that showed high similarities to TEs were selected, and the remaining reads were subsequently BLAST search against the Nipponbare reference genome. The chromosome positions of these non-matching reads were subjected to generate a list that contains the chromosomic regions, inserted element and paired ends of candidate TEs. Those candidates that showed a clear pattern of paired-end mapping on the chromosomal region and transposable element were regarded as mobilization of TEs. To further verify authenticity of the identified TE mobilizations in the RZ35 genome, PCR primers were designed targeting junctions of a set of selected TE insertions using Primer 5 (http://www.premierb iosoft.com) to perform locus-specific amplifications. The primer sequences used in this study were listed in the Table S1. The PCR reaction conditions are 2 min at 94°C; 35 cycles of 45 sec at 94°C, 45 sec at 58°C, 1 min at 72°C; a final extension for 8 min at 55°C. The PCR products were visualized by ethidium bromide staining after electrophoresis through 2% agarose gels.

Table 1. Number of GA reads and coverage of the genome.

	Matsumae	RZ35
FASTQ	24646673	40697827
Coverage	13.2	21.9
Unique map	34584615	57860050
Mutltiple map	11356645	18892937
Unmapped	3352086	4642667

doi: 10.1371/journal.pone.0074479.t001

RNA isolation and real-time qRT-PCR analysis for blast disease-resistance genes

Total RNA was isolated from young seedling at the same developmental stage (at the 3th-4th leaf-stage) from RZ35 and Matsumae with Trizol Reagent (Invitrogen) according to the manufacturer's protocol. The RNA was treated with DNase I (Invitrogen) to eliminate possible genomic DNA contamination before being reverse transcribed with the SuperScript Rnase H-Reverse Transcriptase (Invitrogen). The results of pathogen inoculation experiment revealed that RZ35 exhibited enhanced resistance to rice blast. Therefore, to analyze possible alteration in constitutive expression of 10 previously identified blast disease-resistance genes, gene-specific qRT-PCR primers were designed using Primer 5 (Table S2). All of these identified genes were retrieved from the China Rice Data Center (http://www.ricedata.cn/index.htm). The qRT-PCR reactions were conducted with an initial step at 95°C for 1 min followed by 45 cycles of 5 sec at 95°C, 10 sec at 60°C and 30 sec at 72°C. The relative expression of the analyzed genes was calculated using the Δ Ct method.

Results

Mapping of Illumina reads and detection of genomewide genetic changes in RZ35

In this study, a survey of the genome re-sequencing data for both Matsumae and RZ35 yielded 131 million 90 base pair (bp) paired-end reads in total, and of which 123 million reads were successfully mapped to the Nipponbare reference genome (MSU7.0 http://rice.plantbiology.msu.edu/index.shtml). Notably, 92 million reads aligned uniquely to the reference genome, whereas about 30 million reads showed multiple locations (Table 1). All reads together yielded 13.2x and 21.9x coverage of the Matsumae and RZ35 genomes, respectively. Next, through alignment to the Matsumae genome as a reference, a total of 59,563 homozygous genetic changes were identified in RZ35, which included 41,724 single nucleotide polymorphisms (SNPs), 8,418 insertions and 9,421 deletions (Table 2). All of these DNA sequences have been submitted to GenBank under the accession numbers of SRA072822.

Nucleotide substitution rates and variation patterns of genomic changes in RZ35

Estimation of nucleotide substitution rates was obtained through the comparison between assembled genomes of Matsumae and RZ35. In general, a high nucleotide substitution

Table 2. Polymorphism in genomic DNA observed in the RZ35 compared with Matsumae.

Chromosome	Chr. length (bp)	No. SNPs	SNP/100kb	No. insert.	Insert. /100kb	No. del.	Del. / 100kb	Ts:Tv
1	43268879	4217	9.75	1046	2.42	1215	2.81	2.33
2	35930321	2534	7.05	697	1.94	794	2.21	2.40
3	36406629	1485	4.08	478	1.31	490	1.35	2.87
4	35278165	3595	10.19	834	2.36	872	2.47	1.84
5	29894729	1259	4.21	340	1.14	373	1.25	2.84
6	31246729	6265	20.05	1119	3.58	1246	3.99	2.49
7	29696569	2000	6.73	591	1.99	669	2.25	2.32
8	28439248	1708	6.01	462	1.62	537	1.89	2.42
9	23011178	2729	11.86	510	2.22	594	2.58	2.56
10	23134699	9478	40.97	1070	4.63	1220	5.27	2.38
11	28512606	3844	13.48	737	2.58	821	2.88	2.47
12	27497154	2610	9.49	534	1.94	590	2.15	2.33
Total/Ave.	372316906	41724	11.99	8418	2.26	9421	2.53	2.38

doi: 10.1371/journal.pone.0074479.t002

rate (5.6×10^{-6} per nucleotide site) was found in RZ35 and these genomic variations showed an uneven distribution across each of the 12 rice chromosomes (Figure 1). The average densities of nucleotide polymorphisms were determined on a per 100 kb basis across each chromosome, and this gave rise to an average of 11.99 SNPs, 2.26 insertions and 2.53 deletions if all chromosomes were considered together in RZ35 relative to Matsumae (Table 2). In addition, according to the analysis of base substitution model, two most frequent types of substitution (G/A and C/T) were observed in the RZ35 genome, leading to a relatively higher substitution rate of Ts than Tv (Ts:Tv = 2.38) (Figure S2 and Table 2). Notably, several hyper-mutation regions (HMRs) were identified in the genome of RZ35 (Figure 1), and most of these HMRs occurred in the euchromatin regions of each chromosome, except in chromosome 10 where HMRs were located near the centromere. In addition, chromosome 10 also exhibited the highest SNP and insertions/deletions (INDELs) mutations among all 12 chromosomes (Table 2). Interestingly, it is noted that 11 of these HMRs were associated with domestication (Figure 1). Through search against the available database of *O. sativa*, we found that a small amount of these genomic changes were shared with other rice cultivars (data not shown). It indicates that the introgressive hybridization has triggered novel genetic changes in the RZ35 genome.

Classification of SNPs within mutated genes in RZ35

According to the analyses of global patterns of genomic variation in RZ35, a total of 16,571 SNPs were identified in genes, of which 6,433 were located in coding regions and 8,596 were found in introns (Figure 2). Importantly, 3,797 nonsynonymous mutations were detected in the RZ35 genome and which occurred in 2,250 genes (Additional File S1).

Notably, a large proportion of these nonsynonymous mutations were found in the transpose-coding genes of transposons and retrotransposons, and NBS-LRR protein-coding genes, suggesting these amino acid mutations may lead to the previously characterized phenotypic novelties of RZ35 [31,44].

As it is known from our previous studies that RZ35 has a significantly enhanced resistance to the blast disease relative to its rice parental cultivar Matsumae (e.g., Figure 3) [31], it is of apparent interest to investigate whether the known and putative blast-resistance genes identified in the rice genome are mutated in RZ5 relative to Matsumae. Indeed, we found that a total of 94 nonsynonymous mutations occurred in 40 NBS-LRR disease resistance genes, and one nonsynonymous mutation occurred in a known blast-resistance gene *Pid3/Pi25* (LOC_Os06g22460) (Figure S3). Therefore, we performed qRT-PCR assay for expression of *Pid3/Pi25* along with other nine non-mutated blast-resistance genes to test whether there was differential constitutive expression between RZ35 and Matsumae for these genes. We found that five of these nine genes showed up-regulation in RZ35 relative to Matsumae, and *Pid3/Pi25* showed the highest up-regulation (Figure 4). These findings implied that *Pid3/Pi25* might be responsible for the blast-resistance phenotype of RZ35. Furthermore, identification of the positions of these mutated genes revealed that a total of 1,244 genes were located in 54 pathways and some mutated genes participated in the glycolysis, fatty acid, phenylpropanoid biosynthesis and fatty acid, purine, pyrimidine metabolism pathways (Figure S4-9). Importantly, most of these mutated genes within these pathways were located in the HMRs, suggesting they may confer other phenotypic changes manifested in RZ35 [31,33,45,46].
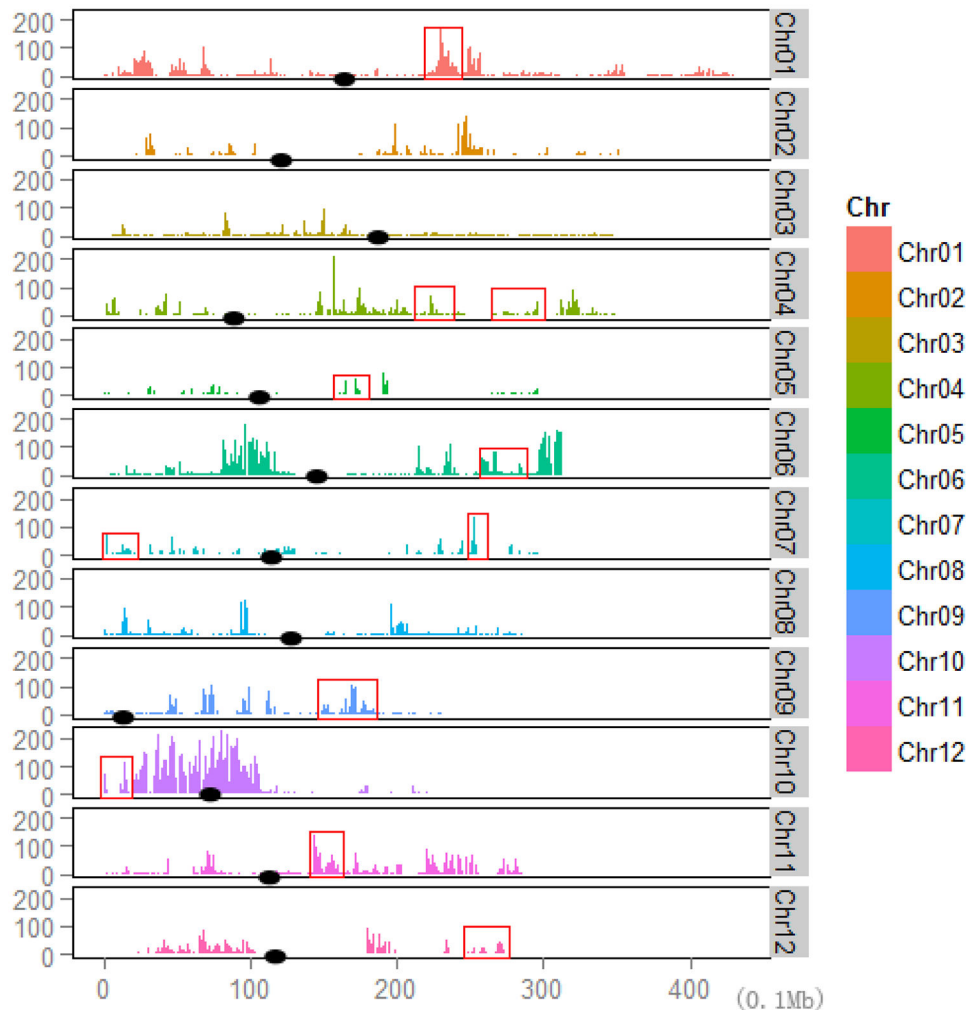
Figure 1. Distribution of SNPs within 100 kb windows across each chromosome (chr). For a change of color denotes different chromosomes. The x axis stands for each chromosome divided by 0.1 Mb. The y axis stands for the number of SNPs. The black solid ellipse stands for the centromere. The red frames stand for the domestication related regions which were found in rice ssp. *japonica*.

doi: 10.1371/journal.pone.0074479.g001

Identification and verification of mobilizations of transposable elements (TEs) in RZ35

Through the paired-end mapping (PEM) approach, genome-wide transpositional reactivations of several TEs were detected in the RZ35 genome (Figure 5). Specifically, a total of 63 mobilization events by 11 TE families were identified in RZ35, of which 28 and 34 new TE insertions were caused by *mPing* and LTR retrotransposons, respectively (Table 3). It should be noted that most of the mobilization events appeared to occur in gene-rich regions and 18 new insertions were located in the intron and upstream of transcription start (< 1 kb) (Table S3). Of these mobilization events of TEs in RZ35, 11 insertion events representing the 11 TE families were selected and independently tested by locus-specific PCR amplifications anchoring the insertion junctions, and all these mobilization

events were validated by this independent analysis (Figure S10).

Discussion

Distribution patterns of the introgressive hybridization-induced genomic changes

It has been documented by an array of studies that introgressive hybridization can indeed evoke extensive and genomewide genetic variations in the resultant hybrids [4,47-55]. In addition to the *bona fide* hybridization, our previous studies have demonstrated that pollination by alien pollens can also induce *de novo* genomic changes in the rice-*Zizania* introgression lines [30,31,33-35]. Nevertheless, because these studies are based on molecular markers sampling limited loci, it remained unclear the extent of the
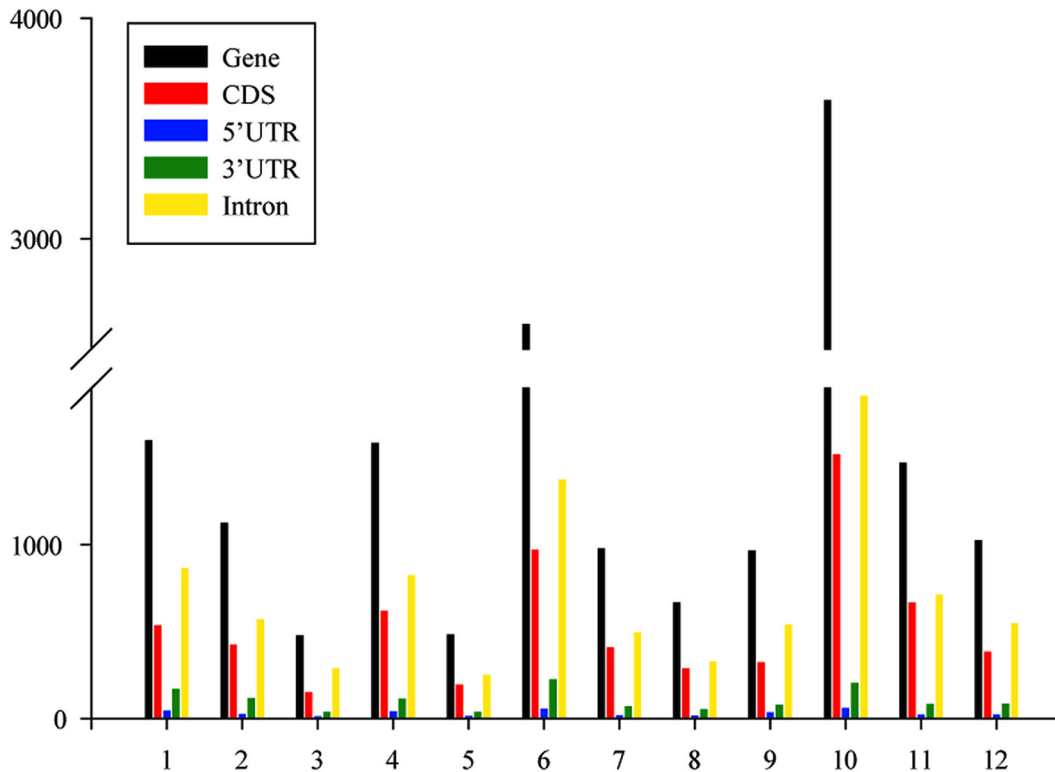
Figure 2. Diagrams showing distribution of homozygous single nucleotide polymorphisms (SNPs) in genes according to each chromosome. The x axis numbers stand for the each chromosome. The y axis numbers stand for the numbers of the homozygous SNPs. The different colors stand for the different components of the genes.

induced genomic changes and their distribution patterns on a genomewide scale. Here, based on deep-coverage whole genome re-sequencing, we show that genomewide genomic changes occurred in a typical introgression line of rice-*Zizania*, RZ35. Analysis of base substitutions indicated that the frequencies of nucleotide transition are significantly higher than those of transversion. This nucleotide substitution bias is in broad consistence with previous studies that also showed higher frequencies of transition than transversion among natural species [56-59], thus corroborating that the genomic variations in RZ35 were actually triggered by introgression of *Z. latifolia* rather than some unknown stochastic factors during the experimental process.

Notably, our results have demonstrated that the nucleotide variations tended to localize in clusters in a given chromosome rather than being randomly distributed. Although such non-randomness with respect to the accumulation of SNPs were also found previously among rice cultivars [60], we note that only a very small amount of the HMRs uncovered in RZ35 is shared by those among the existing rice cultivars, suggesting that novel genetic variations have been induced in rice by the introgressive hybridization with *Zizania*. Rates of nucleotide substitution are known to vary widely across the genome of natural plant species due to the variable efficacy of selection and genetic drift against mutations [61-64]. In fact, previous

studies have implicated that nucleotide substitution rates were expected to be accelerated at those genomic regions under natural selection, but will otherwise be determined by genetic drift only in the neutral regions [65]. Consequently, the genome of natural species is a mosaic of discrete segments and each segment may have its own unique history and contributes differentially to the evolution of a species. This hypothesis is also confirmed by our study because several HMRs were identified across each of the 12 chromosomes of RZ35. Specifically, a number of the mutated genes within HMRs were participated in some biologically important pathways. For example, the phenylpropanoid biosynthesis has a range of important functions in plants, including the synthesis of pigments, growth regulators as well as plant toxins, and all of which can enhance plant's tolerance to biotic and abiotic stresses [45,46]. Similarly, the biosynthesis of glycolysis, fatty acid as well as the metabolism of fatty acid, purine and pyrimidine pathway have also contributed to the abilities of plants in response to environmental stresses [66-68]. Additionally, the pathogen inoculation experiment demonstrated that RZ35 showed enhanced resistance to rice blast and the constitutively unregulated *Pid3/Pi25* may be responsible for the blast-resistance phenotype of RZ35. Although an alternative possibility remains that the blast resistance of RZ35 was introgressed from *Z. Latifolia*, we are

Figure 3. Comparisons of representative phenotype of Matsumae and RZ35. This phenotype was generated at a 7-day period after *Magnaporthe grisea* infection and the scale bar indicates 1cm.
doi: 10.1371/journal.pone.0074479.g003

more in favor of the scenario that the resistance was generated *de novo*, as a consequence of the genomic changes provoked by the introgressive hybridization based on the following arguments: (1) it is known that less than 0.1% of *Zizania* DNA was integrated into the genome of RZ35 [35], and (2) the blast resistance gene *Pid3/Pi25* of RZ35 showed high homology (99%) to rice. These observations allow us to speculate that the induced genomic variations may be at least partly responsible for the observed morphological and physiological novelties of RZ35 in comparison with Matsumae. Interestingly, we also noted that 11 HMRs were located within the domestication-related regions in rice. A previous study showed that these genomic regions may control domestication-related traits and have been selected during the domestication process

of rice [40]. Although the exact functions of these domestication-related regions and how they affect the phenotypes of rice remained unclear, our findings implicated that these regions are likely more fragile and hence labile to mutations as a result of introgressive hybridization. Taken together, our findings suggest that the occurring of genetic variations in RZ35 might have been shaped by some underlying modes during or immediately following the introgression process and these induced genomic changes confer beneficial phenotypes to RZ35.

In contrast to the nonrandom distribution pattern of nucleotide changes discussed above, mobilization of TEs appeared to have occurred randomly across the RZ35 genome. Apart from insertional mutagenesis, transpositionally
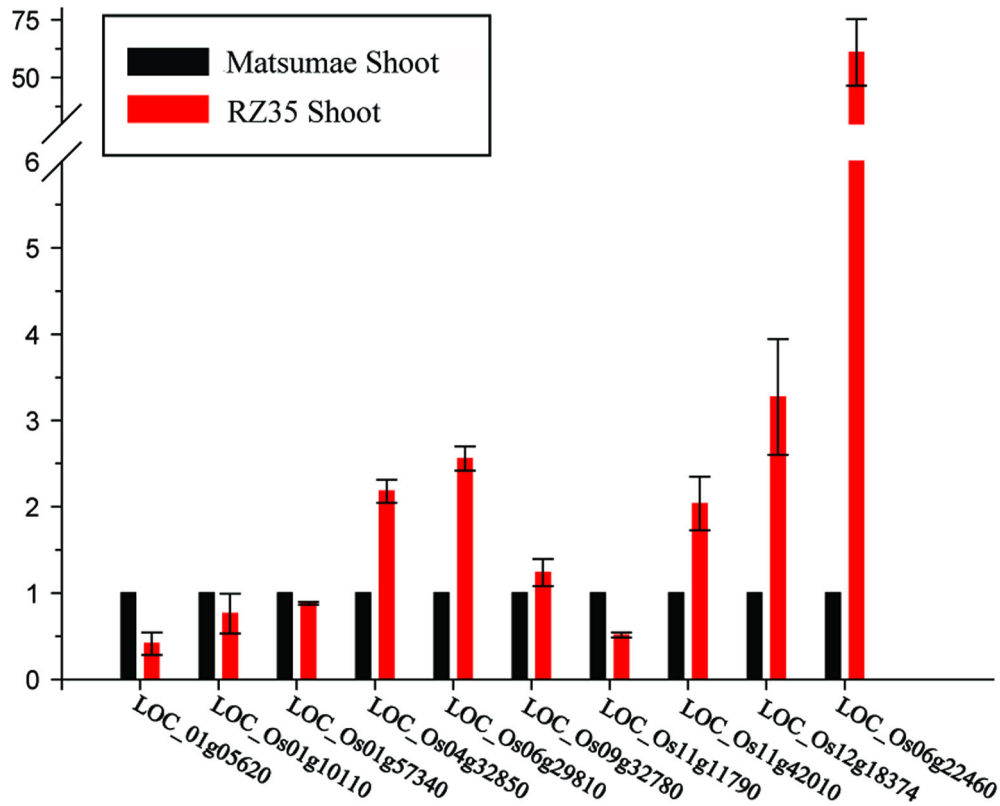
Figure 4. Measurements of expression of the 10 major blast-resistance genes in the leaf tissue taken from of Matsumae and RZ35, measured by real time qRT-PCR.

doi: 10.1371/journal.pone.0074479.g004
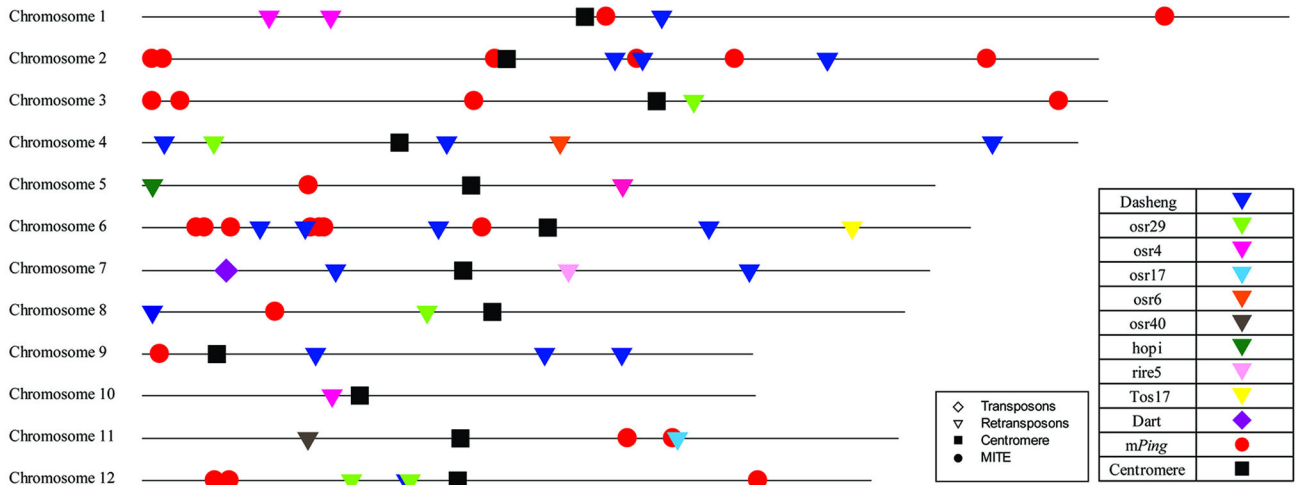
Figure 5. *In silico* mapping of the 63 *de novo* TE insertions detected using PEM method on each chromosome. Arrows represent LTR retrotransposons, diamonds and red dots represent Dart and *mPing*, respectively.

doi: 10.1371/journal.pone.0074479.g005

activated TEs are known to have potent effects on genome structure by inducing both ectopic rearrangements and nucleotide changes at or adjacent to TE excision and insertions

sites [69]. Indeed, several studies have revealed that a significant proportion of plant genes have been assembled or amplified via the action of TEs, and most genes contain

Table 3. The analyzed TEs and their New TE insertions identified in RZ35 and their original copy numbers in Matsumae based on PEM mapping.

TE name	TE type	No. new TE insertions in RZ35	No. original TE copies in Matsumae
Dasheng	LTR retrotransposon	19	289
osr29	LTR retrotransposon	5	74
osr4	LTR retrotransposon	4	29
osr17	LTR retrotransposon	1	13
osr6	LTR retrotransposon	1	5
osr40	LTR retrotransposon	1	120
Hopi	LTR retrotransposon	1	270
Rire5	LTR retrotransposon	1	18
Tos17	LTR retrotransposon	1	4
Dart	DNA Transposon	1	54
mPing	MITE	28	55
Total		63	

doi: 10.1371/journal.pone.0074479.t003

legacies of multiple TE insertions into their promoter regions [70-72]. Our study showing that most of the *de novo* insertion events by the mobilized TEs occurred in gene-rich regions is consistent with the known propensity of these TEs.

Implications of introgressive hybridization-induced genomic changes in plant genome evolution and crop improvement

Although the roles of hybridization in plant genome evolution is still under debate with opposing views of hybridization as either evolutionary noise [73] or an engine for generating biodiversity [74], it has been generally accepted that introgressive hybridization provides an efficient means for transfer of important genes from the donor to the recipient genome [75]. However, the question as to whether the introgressive hybridization process *per se* may generate *de novo* genetic variations remain poorly studied [76]. Our previous, molecular marker-based studies in several rice–*Zizania* introgression lines have shown that the introgressive hybridization process, even between these two sexually incompatible combination of rice and *Zizania*, can generate extensive *de novo* genetic changes [33,35], which has been further confirmed in this study at the whole genome scale. Another pertinent question is whether the genetic variations generated by this means might be relevant to genome evolution or can be explored for crop improvement. Given that closely related species usually exhibit parapatric or sympatric distributions and often share similar pollinators [22-26], we believe that pollination by alien species' pollens may represent a frequent occurrence. Therefore, it is possible that genetic variations produced by this kind of introgressive hybridization

process may have played a role in genome evolution under natural conditions. In this aspect, the variations produced by this novel means likely share similar biological consequences as those produced by *bona fide* hybridizations [75,77]. With respect to the possible utility of the genomic variations generated by introgressive hybridization, the introgression lines of rice-*Zizania* have provided a compelling positive answer: several of these lines have been actively used in rice breeding programs, and the line studied here, RZ35, alone has been involved in the release of 18 new rice cultivars.

In conclusion, we have shown by whole genome re-sequencing that introgressive hybridization between rice (cv. Matsumae) and a sexually incompatible wild species, *Z. latifolia*, has generated genomewide genomic variations, which included SNPs, INDELS and mobilization of TEs. These extensive genomic variations likely underlie genesis of an array of phenotypic and physiological changes in the introgression line. These genomic and phenotypic variations might occur in nature to facilitate genome evolution and can be readily utilized in crop improvements.

Supporting Information

File S1. Details of total homozygous SNPs annotation information. (XLSX)

Figure S1. Schematic view of the procedure of PEM that is the same as Sabot et al. [42].
(TIF)

Figure S2. The matrix of base substitutions. The SNPs detected were classified as transitions (C/T and G/A) or transversions (C/G, T/A, A/C and G/T) based on nucleotide substitutions. The numbers on the line refer to the SNP amount for this type. The degree of thickness of the lines was calculated by the times of the minimum number.
(TIF)

Figure S3. The mutation model of the anti-blast genes. The red vertical lines stand for the SNP mutation site. The red amino acid sequences represent the amino acid site which has a nonanonymous SNP.
(TIF)

Figure S4. The locations of mutated genes in the Glycolysis/Gluconeogenesis pathway. These frames that marketed with green color indicate the positions of each mutated gene.
(TIF)

Figure S5. The locations of mutated genes in the fatty acid Biosynthesis pathway. These frames that marketed with green color indicate the positions of each mutated gene.
(TIF)

Figure S6. The locations of mutated genes in the phenylpropanoid Biosynthesis pathway. These frames that

marketed with green color indicate the positions of each mutated gene.
(TIF)

Figure S7. The locations of mutated genes in the fatty acid Metabolism pathway. These frames that marketed with green color indicate the positions of each mutated gene.
(TIF)

Figure S8. The locations of mutated genes in the purine Metabolism pathway. These frames that marketed with green color indicate the positions of each mutated gene.
(TIF)

Figure S9. The locations of mutated genes in the pyrimidine Metabolism pathway. These frames that marketed with green color indicate the positions of each mutated gene.
(TIF)

Figure S10. Identifications of transposable element (TE) insertions. The Illumina amplicons mapped are indicated above the genome sequence. (A) And (B) indicate the strategies of primer designing that is the same as Sabot et al.

[42]. (C) Ma: represents Marker, M and R donate Matsumae and RZ35, respectively.
(TIF)

Table S1. The primer sequences for the identification of transposon elements insertion.
(DOC)

Table S2. The primers sequences of anti-blast genes used in this study.
(DOC)

Table S3. Information of these genes that new TE insertion was found within 1 kb.
(DOC)

Author Contributions

Conceived and designed the experiments: LFL BL. Performed the experiments: ZHW DZ YB YHZ YL YW JWW XYL. Analyzed the data: ZHW CMX LFL BL. Contributed reagents/materials/analysis tools: ZHW CMX LFL BL. Wrote the manuscript: ZHW LFL BL.

References

1. Brennan AC, Bridle JR, Wang A-L, Hiscock SJ, Abbott RJ (2009) Adaptation and selection in the Senecio (Asteraceae) hybrid zone on Mount Etna, Sicily. New Phytol 183: 702-717. doi:10.1111/j.1469-8137.2009.02944.x. PubMed: 19594693.

2. Chen ZJ (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annu Rev Plant Biol 58: 377-406. doi:10.1146/annurev.arplant.58.032806.103835. PubMed: 17280525.

3. Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic Changes in Resynthesized Brassica napus and Their Effect on Gene Expression and Phenotype. Plant Cell Available: 19: 3403-3417 PubMed: 18024568.

4. Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J et al. (2002) Remodeling of DNA Methylation and Phenotypic and Transcriptional Changes in Synthetic Arabidopsis Allotetraploids. Plant Physiol 129: 733-746. doi:10.1104/pp.003095. PubMed: 12068115.

5. Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-Induced Rapid Genome Evolution in the Wheat (Aegilops–Triticum) Group. Plant Cell Available: 13: 1735-1747

6. Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Mol Ecol 14: 1163-1175. doi:10.1111/j.1365-294X.2005.02488.x. PubMed: 15773943.

7. Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence Elimination and Cytosine Methylation Are Rapid and Reproducible Responses of the Genome to Wide Hybridization and Allopolyploidy in Wheat. Plant Cell Available: 13: 1749-1759 PubMed: 11487690.

8. Zou J, Fu D, Gong H, Qian W, Xia W et al. (2011) De novo genetic variation associated with retrotransposon activation, genomic rearrangements and trait variation in a recombinant inbred line population of Brassica napus derived from interspecific hybridization with Brassica rapa. Plant J 68: 212-224. doi:10.1111/j.1365-313X.2011.04679.x. PubMed: 21689170.

9. McClintock B (1984) The significance of responses of the genome to challenge. Science 226: 792-801. doi:10.1126/science.15739260. PubMed: 15739260.

10. Yasuda K, Tsukiyama T, Karki S, Okumoto Y, Teraishi M et al. (2013) Mobilization of the active transposon mPing in interspecific hybrid rice between Oryza sativa and O. glaberrima. Euphytica 192: 17-24. doi:10.1007/s10681-012-0810-1.

11. Fritz RS, Moulia C, Newcombe G (1999) RESISTANCE OF HYBRID PLANTS AND ANIMALS TO: HERBIVORES, PATHOGENS, AND Annual: PARASITES Review of Ecology and Systematics 30: 565-591

12. Leitch AR, Leitch IJ (2008) Genomic Plasticity and the Diversity of Polyploid Plants. Science 320: 481-483. doi:10.1126/science.1153585. PubMed: 18436776.

13. Soltis PS (2013) Hybridization, speciation and novelty. J Evol Biol 26: 291-293. doi:10.1111/jeb.12095. PubMed: 23324011.

14. Whitham TG, Martinsen GD, Keim P, Floate KD, Dungey HS et al. (1999) PLANT HYBRID ZONES AFFECT BIODIVERSITY: TOOLS FOR A GENETIC-BASED UNDERSTANDING OF COMMUNITY STRUCTURE. Ecology 80: 416-428. doi:10.1890/0012-9658(1999)080[0416:PHZABT]2.0.CO;2.

15. Hegarty MJ, Hiscock SJ (2008) Genomic Clues to the Evolutionary Success of Polyploid Plants. Curr Biol CB 18: R435-R444. doi:10.1016/j.cub.2008.02.034. PubMed: 18492478.

16. Rieseberg LH (1997) HYBRID ORIGINS OF PLANT SPECIES. Annu Rev Ecol Syst 28: 359-389. doi:10.1146/annurev.ecolsys.28.1.359.

17. Otto SP, Whitton J (2000) POLYPLOID INCIDENCE AND EVOLUTION. Annu Rev Genet 34: 401-437. doi:10.1146/annurev.genet.34.1.401. PubMed: 11092833.

18. Rieseberg LH, Willis JH (2007) Plant Speciation. Science 317: 910-914. doi:10.1126/science.1137729. PubMed: 17702935.

19. Stebbins GL (1971) Chromosomal evolution in higher plants. Chromosomal evolution in higher plants.

20. Husband BC (2000) Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. Proc R Soc Lond B Biol Sci 267: 217-223. doi:10.1098/rspb.2000.0990. PubMed: 10714875.

21. Levin DA (1975) Minority cytotype exclusion in local plant populations. Taxon: 35-43.

22. Anderson RP (1999) Preliminary review of the systematics and biogeography of the spiny pocket mice (Heteromys) of Colombia. Rev Academia Colombiana Ciencias Exactas Físicas Nat 23.

23. Genoways HHB, James H (1993) Biology of the Heteromyidae: Spec. Publ. Am. Soc. Mammal

24. Sakai S, Kawakita A, Ooi K, Inoue T (2013) Variation in the strength of association among pollination systems and floral traits: Evolutionary changes in the floral traits of Bornean gingers (Zingiberaceae). Am J Bot 100: 546-555. doi:10.3732/ajb.1200359. PubMed: 23425560.

25. Schiestl FP, Johnson SD (2013) Pollinator-mediated evolution of floral signals. Trends Ecol Evol, 28: 307–15. PubMed: 23480953.

26. Wilson DE, Reeder DM (2005) Mammal species of the world: a taxonomic and geographic reference. Johns Hopkins University Press.

27. Boyko A, Kovalchuk I (2011) Genetic and Epigenetic Effects of Plant–Pathogen Interactions: An Evolutionary Perspective. Molecular Plant.

28. Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM et al. (2012) Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci USA 109: E2183-E2191. doi:10.1073/pnas.1209329109. PubMed: 22733782.

29. Holeski LM, Jander G, Agrawal AA (2012) Transgenerational defense induction and epigenetic inheritance in plants. Trends Ecol Evol 27: 618-626. doi:10.1016/j.tree.2012.07.011. PubMed: 22940222.

30. Wang H, Chai Y, Chu X, Zhao Y, Wu Y et al. (2009) Molecular characterization of a rice mutator-phenotype derived from an incompatible cross-pollination reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic instability. BMC Plant Biol 9.

31. Liu B, Zhao R, Piao H, Zhao F, Zhao J (1999) Production and molecular characterization of rice lines with introgressed traits from a wild species Zizania latifolia (Griseb.)[Oryza sativa. p. L.]. Journal of Genetics & Breeding 53

32. Liu Z, Wang Y, Shen Y, Guo W, Hao S et al. (2004) Extensive Alterations in DNA Methylation and Transcription in Rice Caused by Introgression from Zizania Latifolia. Plant Mol Biol 54: 571-582. doi:10.1023/B:PLAN.0000038270.48326.7a. PubMed: 15316290.

33. Shan X, Liu Z, Dong Z, Wang Y, Chen Y et al. (2005) Mobilization of the active MITE transposons mPing and Pong in rice by introgression from wild rice (Zizania latifolia Griseb.). Mol Biol Evol 22: 976-990. doi:10.1093/molbev/msi082. PubMed: 15647520.

34. Wang N, Wang H, Wang H, Zhang D, Wu Y et al. (2010) Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with Zizania latifolia. BMC Plant Biol 10: 190. doi:10.1186/1471-2229-10-190. PubMed: 20796287.

35. Wang Y-M, Dong Z-Y, Zhang Z-J, Lin X-Y, Shen Y et al. (2005) Extensive de novo genomic variation in rice induced by introgression from wild rice (Zizania latifolia Griseb.). Genetics 170: 1945-1956. doi:10.1534/genetics.105.040964. PubMed: 15937131.

36. Dong Z, Wang H, Dong Y, Wang Y, Liu W et al. (2013) Extensive Microsatellite Variation in Rice Induced by Introgression from Wild Rice. PLOS ONE 8: e62317. doi:10.1371/journal.pone.0062317. PubMed: 23638037.

37. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11-15.

38. Wang Z-X, Yano M, Yamanouchi U, Iwamoto M, Monna L et al. (1999) The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19: 55-64. doi:10.1046/j.1365-313X.1999.00498.x. PubMed: 10417726.

39. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754-1760. doi:10.1093/bioinformatics/btp324. PubMed: 19451168.

40. Huang X, Kurata N, Wei X, Wang Z-X, Wang A et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490: 497-501. doi:10.1038/nature11532. PubMed: 23034647.

41. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35: W182-W185. doi:10.1093/nar/gkm321. PubMed: 17526522.

42. Sabot F, Picault N, El-Baidouri M, Llauro C, Chaparro C et al. (2011) Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. Plant J 66: 241-246. doi:10.1111/j.1365-313X.2011.04492.x. PubMed: 21219509.

43. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25. doi:10.1186/gb-2009-10-3-r25. PubMed: 19261174.

44. Wang YM, Dong ZY, Zhang ZJ, Lin XY, Shen Y et al. (2005) Extensive de Novo genomic variation in rice induced by introgression from wild rice (Zizania latifolia Griseb.). Genetics 170: 1945-1956. doi:10.1534/genetics.105.040964. PubMed: 15937131.

45. Weisshaar B, Jenkins GI (1998) Phenylpropanoid biosynthesis and its regulation. Curr Opin Plant Biol 1: 251-257. doi:10.1016/S1369-5266(98)80113-1. PubMed: 10066590.

46. Anzenbacher P, Anzenbacherová E (2001) Cytochromes P450 and metabolism of xenobiotics. Cell Mol Life Sci CMLS 58: 737-747. doi:10.1007/PL00000897. PubMed: 11437235.

47. Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH et al. (2000) Phenotypic instability and rapid gene silencing in newly formed Arabidopsis allotetraploids. Plant Cell Available: 12: 1551-1567 PubMed: 11006331.

48. Ge X-H, Wang J, Li Z-Y (2009) Different genome-specific chromosome stabilities in synthetic Brassica allohexaploids revealed by wide crosses with Orychophragmus. Ann Bot 104: 19-31. doi:10.1093/aob/mcp099. PubMed: 19403626.

49. He G, Zhu X, Elling AA, Chen L, Wang X et al. (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell Available: 22: 17-33 PubMed: 20086188.

50. Hegarty MJ, Batstone T, Barker GL, Edwards KJ, Abbott RJ et al. (2011) Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in Senecio (Asteraceae). Mol Ecol 20: 105-113. doi:10.1111/j.1365-294X.2010.04926.x. PubMed: 21073590.

51. Hegarty MJ, Jones JM, Wilson ID, Barker GL, Coghill JA et al. (2005) Development of anonymous cDNA microarrays to study changes to the Senecio floral transcriptome during hybrid speciation. Mol Ecol 14: 2493-2510. doi:10.1111/j.1365-294x.2005.02608.x. PubMed: 15969730.

52. Liu B, Vega J, Segal G, Abbo S, Rodova M et al. (1998) Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. I. Changes in low-copy noncoding DNA sequences. Genome 41: 272-277. doi:10.1139/gen-41-2-272.

53. Malinska H, Tate JA, Matyasek R, Leitch AR, Soltis DE et al. (2010) Similar patterns of rDNA evolution in synthetic and recently formed natural populations of Tragopogon (Asteraceae) allotetraploids. BMC Evol Biol 10: 291. doi:10.1186/1471-2148-10-291. PubMed: 20858289.

54. Petit M, Guidat C, Daniel J, Denis E, Montoriol E et al. (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186: 135-147. doi:10.1111/j.1469-8137.2009.03140.x. PubMed: 20074093.

55. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301: 1211-1216. doi:10.1126/science.1086949. PubMed: 12907807.

56. Cantu D, Vanzetti LS, Sumner A, Dubcovsky M, Matvienko M et al. (2010) Small RNAs, DNA methylation and transposable elements in wheat. BMC Genomics 11: 408. doi:10.1186/1471-2164-11-408. PubMed: 20584339.

57. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM et al. (2010) The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science 327: 92-94. doi:10.1126/science.1180677. PubMed: 20044577.

58. Park S, Yu H-J, Mun J-H, Lee S-C (2010) Genome-wide discovery of DNA polymorphism in Brassica rapa. Mol Genet Genomics 283: 135-145. doi:10.1007/s00438-009-0504-0. PubMed: 20024583.

59. Wang X-W, Zhao Q-Y, Luan J-B, Wang Y-J, Yan G-H et al. (2012) Analysis of a native whitefly transcriptome and its sequence divergence with two invasive whitefly species. BMC Genomics 13: 529. doi:10.1186/1471-2164-13-529. PubMed: 23036081.

60. Xu X, Liu X, Ge S, Jensen JD, Hu F et al. (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30: 105-111. doi:10.1038/nbt.2050. PubMed: 22158310.

61. Chen J-Q, Wu Y, Yang H, Bergelson J, Kreitman M et al. (2009) Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Mol Biol Evol 26: 1523-1531. doi:10.1093/molbev/msp063. PubMed: 19329651.

62. Ewens WJ (2004) Mathematical population genetics: I. Theoretical introduction. Springer Verlag.

63. Lang GI, Botstein D, Desai MM (2011) Genetic variation and the fate of beneficial mutations in asexual populations. Genetics 188: 647-661. doi:10.1534/genetics.111.128942. PubMed: 21546542.

64. Loewe L, Hill WG (2010) The population genetics of mutations: good, bad and indifferent. Philos Trans R Soc Lond B Biol Sci 365: 1153-1167. doi:10.1098/rstb.2009.0317. PubMed: 20308090.

65. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12: 1805-1814. doi:10.1101/gr.631202. PubMed: 12466284.

66. Farmer EE, Weber H, Vollenweider S (1998) Fatty acid signaling in Arabidopsis. Planta 206: 167-174. doi:10.1007/s004250050388. PubMed: 9736997.

67. Hu Z, Plaxton WC (1996) Purification and Characterization of Cytosolic Pyruvate Kinase from Leaves of the Castor Oil Plant. Arch Biochem Biophys 333: 298-307. doi:10.1006/abbi.1996.0394. PubMed: 8806784.

68. Mou Z, He Y, Dai Y, Liu X, Li J (2000) Deficiency in Fatty Acid Synthase Leads to Premature Cell Death and Dramatic Alterations in Plant Morphology. Plant Cell Available: 12: 405-417 PubMed: 10715326.

69. Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Molecular Evolution. Springer Verlag. pp. 251-269.

70. Wendel JF (2000) Genome evolution in polyploids. Plant Molecular Evolution. Springer Verlag. pp. 225-249.

71. Wendel JF, Wessler SR (2000) Retrotransposon-mediated genome evolution on a local ecological scale. Proc Natl Acad Sci USA 97: 6250-6252. doi:10.1073/pnas.97.12.6250. PubMed: 10841529.

72. White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91: 11792-11796. doi:10.1073/pnas.91.25.11792. PubMed: 7991537.

73. Wagner W Jr (1970) Biosystematics and evolutionary noise. Taxon: 146-151.

74. Arnold ML (1997) Natural hybridization and evolution. USA: Oxford University Press.

75. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ et al. (2013) Hybridization and speciation. J Evol Biol 26: 229-246. doi:10.1111/j. 1420-9101.2012.02599.x. PubMed: 23323997.

76. Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010) Homoploid hybrid speciation in action. Taxon 59: 1375-1386.

77. Sætre GP (2013) Hybridization is important in evolution, but is speciation? J Evol Biol 26: 256-258. doi:10.1111/jeb.12005. PubMed: 23324000.