Planta
DOI 10.1007/s00425-014-2159-2

ORIGINAL ARTICLE

Genome re-sequencing suggested a weedy rice origin from domesticated *indica-japonica* hybridization: a case study from southern China

Jie Qiu · Jinwen Zhu · Fei Fu · Chu-Yu Ye · Weidi Wang · Linfeng Mao · Zhangxiang Lin · Li Chen · Haiqiang Zhang · Longbiao Guo · Shen Qiang · Yongliang Lu · Longjiang Fan

Received: 6 June 2014 / Accepted: 16 August 2014

Abstract

Main conclusion **Whole-genome re-sequencing of weedy rice from southern China reveals that weedy rice can originate from hybridization of domesticated *indica* and *japonica* rice.**

Abstract Weedy rice (*Oryza sativa* f. *spontanea* Rosh.), which harbors phenotypes of both wild and domesticated rice, has become one of the most notorious weeds in rice fields worldwide. While its formation is poorly understood, massive amounts of rice genomic data may provide new insights into this issue. In this study, we determined genomes of three weedy rice samples from the lower Yangtze region, China, and investigated their phylogenetics, population structure and chromosomal admixture patterns. The phylogenetic tree and principle component analysis based on 46,005 SNPs with 126 other *Oryza* accessions suggested that the three weedy rice accessions were intermediate between *japonica* and *indica* rice. An ancestry inference study further demonstrated that weedy rice had two dominant genomic components (temperate *japonica* and *indica*). This strongly suggests that weedy rice originated from *indica-japonica* hybridization. Furthermore, 22,443 novel fixed single nucleotide polymorphisms were detected in the weedy genomes and could have been generated after *indica-japonica* hybridization for environmental adaptation.

Keywords Weedy rice (*Oryza sativa* f. *spontanea* Rosh.) · Whole-genome re-sequencing · Population structure · Hybridization origin

Introduction

Weedy rice (*Oryza sativa* f. *spontanea* Rosh.), which is also called red rice, is a variety of rice (*Oryza*) that produces much fewer grains per plant and competes aggressively with crops. It is widely present in rice-planting areas all over the world, and is particularly abundant in South and North America, southern Europe, and southern and southeast Asia (Ferrero et al. 1999; Noldin et al. 1999; Mortimer et al. 2000; Sun et al. 2013). Weedy rice emerged as an agricultural issue for Asian countries over the past 20 years, and had an impact for a much longer time in parts of the world that have long relied on direct seeding, like the United States. For example, weedy rice has been shown to cause an annual loss of over $50 million in southern parts of the United States (Gealy et al. 2002). The area affected

---

Electronic supplementary material The online version of this article (doi:10.1007/s00425-014-2159-2) contains supplementary material, which is available to authorized users.

J. Qiu · F. Fu · C.-Y. Ye · W. Wang · L. Mao · Z. Lin ·
L. Chen · H. Zhang · L. Fan (✉)
Department of Agronomy, College of Agriculture and
Biotechnology, Zhejiang University, Hangzhou 310058, China
e-mail: fanlj@zju.edu.cn

J. Qiu
e-mail: lefroyqiu@foxmail.com

J. Zhu
Department of Plant Protection, College of Agriculture and
Biotechnology, Zhejiang University, Hangzhou 310058, China

L. Guo · Y. Lu (✉)
China National Rice Research Institute, Chinese Academy of
Agricultural Sciences, Hangzhou 310006, China
e-mail: luyongliang@caas.cn

S. Qiang
Weed Research Laboratory, Nanjing Agricultural University,
Nanjing 210095, China

by weedy rice in China now exceeds three million ha, and the total crop yield has been reduced by 3.4 billion kilograms (Liang and Qiang 2011). Moreover, other crops including jute, maize and soybean, are also harmed by weedy rice (Baki et al. 2000). The invasive success of this weed is mainly due to its mixture of domesticated and undomesticated traits. It harbors characteristics of undomesticated *Oryza* species, including seed dispersal mechanisms and seed dormancy. On the other hand, it also possesses traits of domesticated rice, such as rapid growth, and resembles domesticated rice during the seedling stage, which promotes its invasiveness in the agroecosystem (Sun et al. 2013).

The earliest descriptions of weedy rice are found in ancient Chinese books (Ding 1983; Xiuling You, personal communication) and in American scientific papers from around 150 years ago (Allston 1846), which has led to an intensive debate regarding the origin of weedy rice. Some early studies supported the notion that weedy rice was formed from the ongoing selection and adaptation of wild rice (De Wet and Harlan 1975; Harlan 1992). The de-domestication hypothesis suggested that the reversion of crop species to a wild or feral form would probably occur when domesticated rice was abandoned (Bres-Patry et al. 2001). A third hypothesis suggested that weedy rice originated from hybridization between cultivated rice and its progenitor type (Tang and Morishima 1996). In addition, weedy rice may also originate from ongoing and multidirectional hybridization between weedy rice and cultivated types as well as hybridization among weedy types (Londo and Schaal 2007). Recently, some researchers have noted that conspecific weedy rice could also unexpectedly evolve from hybridization of its cultivated relatives (Ishikawa et al. 2005; Reagon et al. 2010; Xiong et al. 2012). For example, Xiong et al. (2012) demonstrated that weedy-type offspring could emerge after inter-subspecies and inter-varietal hybridization in rice. Reagon et al. (2010) suggested that US weedy rice evolved directly from within domesticated lineages. After the formation of weedy rice, the continual gene flow from the cultivated crop into neighboring weedy populations in combination with natural selection has been proposed to play a critical role in the adaptive evolution of weedy rice (Jiang et al. 2012; Sun et al. 2013).

Diverse molecular markers have been used to infer the population structure, genetic diversity, and adaptive loci of weedy rice. On the basis of RFLP and SSR analysis, red rice in the United States has been confirmed to have originated from Asian *O. sativa aus* and *indica*, neither of which has been cultivated in the US (Vaughan et al. 2001; Londo and Schaal 2007; Gross et al. 2010; Thurber et al. 2010). The overall SSR based genetic diversity of weedy rice populations from China suggested that the weedy rice found there most likely originated from local *indica* or *japonica* varieties (Cao et al. 2006; Shao et al. 2011; Wang et al. 2012; Sun et al. 2013). Reagon et al. (2010) used the SNP patterns from 111 sequenced tagged loci to infer the origin and demographic processes of US weedy rice evolution and found no evidence for the contribution of wild rice to US weedy rice. As a model system for crop genomics studies, huge amounts of rice genomic data, including wild and cultivated accessions, have been available for the rice research community, especially after the advent of next-generation high-throughput sequencing (NGS) technologies (e.g., Huang et al. 2012; Xu et al. 2012). Such advanced strategies and massive amounts of data have driven progress in rice genomics studies, such as on the characterization of genomic variation and genetic structure in rice populations, the relations between genomic variations and traits, and the origin of cultivated rice and its domestication history (e.g., Huang et al. 2013). Since weedy rice is a member of the *Oryza* genus, it also provides a great opportunity to understand the origin and evolution of problematic weedy species through dissection of the underlying genomic information.

Since direct seeding and related technologies have gradually become more popular over the past 20 years, weedy rice has become one of the most serious weed issues in rice-planting areas in China, particularly in the northeast region and in Jiangsu, Guangdong and Hainan provinces (Cao et al. 2006; Liang and Qiang 2011; Shao et al. 2011; Wang et al. 2012; Sun et al. 2013). In Jiangsu province, where both *indica* and *japonica* rice are grown, direct seeding is still preferred by farmers to save labor and costs, even though a change from direct-seeded rice to transplanted rice is an effective approach to controlling weedy rice. Most farmers did nothing to remove weedy rice, and their associated yield losses were initially ignored. As a consequence, the weedy rice seed bank in dry direct-seeded rice fields is likely to be gradually increasing, which will be followed by devastating invasions of weedy rice (Zhang et al. 2014).

It is important that we elucidate the origin of weedy rice, not only so that we can understand the adaptive evolution of a plant taxon under human influences, but also for the effective management of this weed. In this study, we re-sequenced three weedy rice accessions from the lower Yangtze region in Jiangsu province, where more than two-thirds of the rice in China is grown, with the aim of understanding the origin of the weedy rice invading this district. The origin and ancestry proportions of weedy rice were inferred using the genomic data for these three weedy rice accessions along with those for 126 other members of the *Oryza* genus. All of our results support the idea that this weedy rice originated from hybridization between *indica* and temperate *japonica* rice.

Materials and methods

Plant materials and DNA sequencing

Three weedy rice plants (YG02, YJ06 and YJ07) were collected from the Guocun region in Yangzhou, Jiangsu province, China (E119°79′, N32°53′) in 2013. Jiangsu province, located in the Yangtze basin, is an important area of *japonica* rice production and is also one of the regions in China that is seriously affected by weedy rice (Shao et al. 2011; Wang et al. 2012). Genomic data from 126 *Oryza* accessions (including 36 and 59 domesticated *indica* and *japonica* varieties, respectively; 15 *O. rufipogon*, 5 *O. nivara* and 11 admixture accessions; Table S1) (Huang et al. 2012; Xu et al. 2012) were downloaded from ftp://public.genomics.org.cn/BGI/rice/rice_resequencing/02.SNP/50_stringent/ and http://202.127.18.221/RiceHap3/Genotype.php. The geographical locations of the 126 accessions and three weedy rice lines used in this study are listed in Table S1. For weedy rice genome sequencing, their entire DNAs were extracted from green leaves using a routine protocol (CTAB). A 400 bp insertion size lib for each weedy rice accession was generated by the Illumina Hiseq 2000 sequencing platform. The short-read sequence data obtained in this study have been deposited into GenBank under the SRA study accession number SRP041867.

Reads mapping

Raw paired-end reads were first filtered into clean data using NGSQCtoolkit v2.3.3 (Patel and Jain 2012). The cut-off value for the PHRED quality score was set to Q20 and 70 % of the read lengths met the given quality. The clean reads of each weedy rice were aligned to the *japonica* rice reference genome IRGSP build 4 (IRGSP, 2005), Os-Nipponbare-Reference-IRGSP-1.0 (Phytozome v9.1) and an *indica* rice reference genome (ASM465v1.21) using BWA v0.5.9 (Li and Durbin 2009). Reads were uniquely mapped to the reference by allowing only one alignment output, and singletons were further excluded by the setting ‘sampe –n 1 –N 1′.

SNP detection and genotyping

SNPs detection was performed using the UnifiedGenotyper tool in GATK v2.3 (McKenna et al. 2010). The standard call confidence for SNP calling was set to 30. To achieve a low missing rate and a non-heterozygous SNP dataset for a further population study, we first excluded heterozygous SNPs in both datasets and chose SNPs with a missing rate of <50 % for the 76 Chinese rice accessions in the 126 *Oryza* lines. The two filtered SNPs datasets were intersected into a final SNP dataset. We integrated genotypes of the three weedy rice samples into the intersected genotype table by genotyping based on the known SNP locations and then choosing alleles with at least five reads supported for each sample.

Domestication-related allele identification and PCR validation

Three domestication-related genes in weedy rice were examined using our genomic data and a PCR-based approach: *sh4*, which controls seed shattering; *Phr1*, which acts as a classic diagnostic trait for subspecies differentiation; and one seed color-related gene (*Rc*) (Sweeney et al. 2006; Yu et al. 2008; Zhang et al. 2009). The primers for PCR amplification of *sh4, Phr1*, and *Rc* adopted by Sun et al. (2013) were used in this study.

Population structure and admixture pattern analysis

We applied a maximum likelihood method to construct the phylogenetic tree based on the SNP-based genetic distance using MEGA v5.1 software (Tamura et al. 2011) with 1000 replicates for a bootstrap confidence analysis. FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) was used to plot the constructed tree. The population structure was generated by Frappe v1.1 (Tang et al. 2005), which is based on a maximum likelihood method for estimating individual ancestry and admixture proportions using high-density SNP data. We predefined the number of genetic clusters *K* ranging from 2 to 8, and the converged admixture proportion result for each *K* calculated by Frappe v1.1 was further drawn by distruct v1.1 (Rosenberg 2004). Principle component analysis was performed using SNPRelate v0.9.19 (Zheng et al. 2012).

Admixture pattern detection across chromosomes

Admixture patterns along chromosomes were estimated using StepPCO software (Pugach et al. 2011). As indicated by the Frappe result, two sets, including 27 temperate *japonica* and 12 *indica* accessions with no admixture, were selected as the two parental populations for weedy rice individuals. SNPs with a missing rate of 30 % or heterozygous SNPs for the 39 accessions were initially excluded. The *indica* and *japonica* datasets with filtered SNPs were then intersected with the weedy rice genotypes. We used the default setting (sigma = 3; Nbins = 1,024) to determine the window size, as suggested by Pugach et al. (2011), so that ancestral populations are sufficiently well separated by the properties of the collection of SNPs in this window size.

Identification of novel SNPs

For the identification of novel SNPs in weedy rice, all available rice SNPs data were downloaded from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/, 2014-04-05) (Sherry et al. 2001). The FASTA sequences with SNP information were mapped to the *japonica* reference Os-Nipponbare-Reference-IRGSP-1.0 by Bowtie2 v2.0.5 (Langmead and Salzberg 2012) to locate the SNPs in the reference genome. Based on the SNP calling result, we detected SNPs supported by at least five reads from weedy rice, which were absent from the available rice dbSNP (Sherry et al. 2001) for each weedy rice sample. We then further searched for potentially fixed novel SNPs across the whole genome and guaranteed that the SNPs were shared by the three weedy rice samples. SNPEff v3.4 (Cingolani et al. 2012) was used to predict the effects of the fixed novel SNPs.

GO and metabolic pathway enrichment analysis

Genes with novel SNPs were extracted from the rice genome annotation file (www.phytozome.net). GO enrichment analysis was performed with AgriGO with the '*Oryza sativa*' set as a species background (http://bioinfo.cau.edu.cn/agriGO/) (Du et al. 2010). The *P* value and FDR (Yekutieli) criteria for the considered enrichment GO terms were <0.0001 and <0.05, respectively. Plant metabolic pathway enrichment analysis was performed online at http://structuralbiology.cau.edu.cn/PlantGSEA/analysis.php (Yi et al. 2013).

Results

Phenotype characterization

The three weedy rice lines were collected from the Yangtze basin, which is one of the regions in China that is seriously affected by weedy rice. Weedy rice plants grow vigorously and have an erect plant architecture, which generally resembles *indica* rice in shape and is different from *japonica* rice in the field (Fig. 1a). The seeds of the three weedy rice samples are shorter than those of wild (*O. rufipogon*) and *indica* rice, but similar to that of *japonica* rice. The hull color of weedy rice is generally the same as that of cultivated rice, but with black dots (Fig. 1b), resembling wild rice. Significant seed shattering was observed for weedy rice. A smooth abscission layer was observed, similar to that in wild rice, which is consistent with their shattering phenotype. The red-pigmented pericarp of weedy rice is similar to that of wild rice, which explains why it is also called 'red rice' (Fig. 1b). In brief, the three weedy rice plants sampled by this study harbor a mixed phenotype of domesticated *indica*, *japonica*, and wild rice.

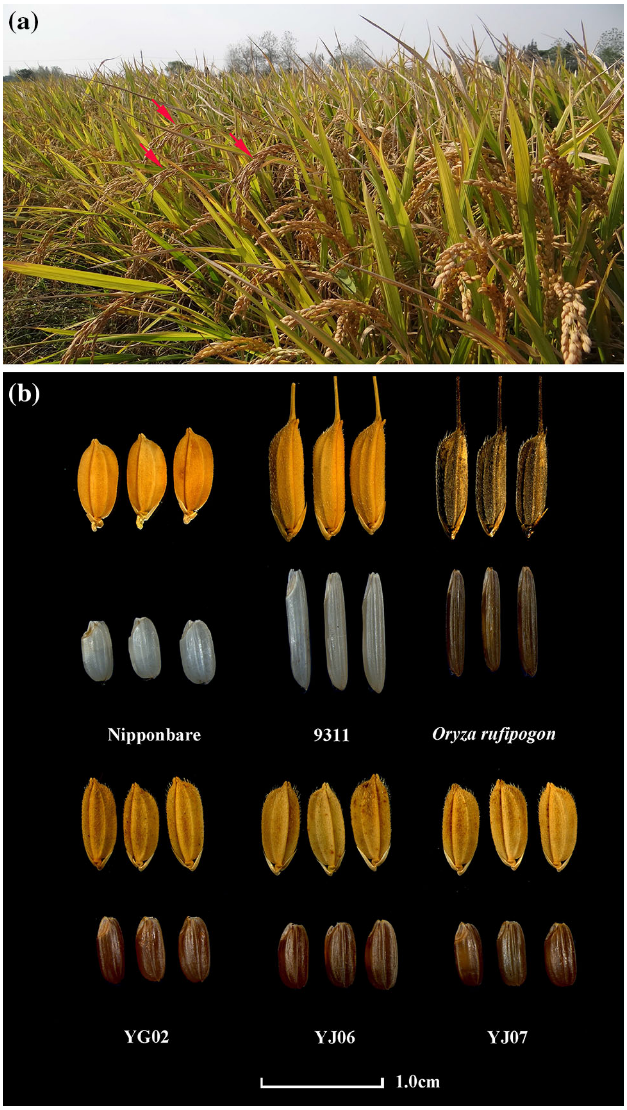

Fig. 1 Phenotypes of the three weedy rice lines in this study. **a** A rice field where weedy rice samples were collected in this study. The *japonica* rice field has been severely invaded by weedy rice (*red arrows*). **b** Comparison of the seed phenotypes of weedy rice and domesticated (*indica* 9311 and *japonica* Nipponbare) and wild rice

Genome re-sequencing and mapping

A total of 43.4 million reads (14.6 Gb) or ~5× genome coverage (raw data) were generated for the three weedy rice plants using HiSeq 2000 (Table 1). After low-quality reads were removed, about 89 % of the reads (2.65–4.73 genome coverage) were retained as clean data and used for further investigation. All of the clean reads of each weedy rice sample were mapped to the *indica* and *japonica*

Table 1 Summary of sequencing and mapping results of three weedy rice lines

Reference	*Indica* (va. 9311)			*Japonica* (va. nipponbare)		
Accession	YG02	YJ06	YJ07	YG02	YJ06	YJ07
Total raw bases	2,058,511,200	2,297,730,900	2,147,688,900	2,058,511,200	2,297,730,900	2,147,688,900
Total clean bases	1,838,435,850	2,052,857,100	1,890,765,450	1,838,435,850	2,052,857,100	1,890,765,450
Clean reads percentage (%)	89.31	89.34	88.04	89.31	89.34	88.04
Clean reads mapped percentage (%)	85.31	86.81	61.66	72.58	74.32	52.93
Mapping depth (x)	4.17	4.73	3.10	3.54	4.05	2.65
Reference coverage (%)	84.57	85.88	80.89	74.63	76.07	70.97

Table 2 Identification of domestication allele in weedy rice

Locus	Domesticated type	By mapping read numbers			By PCR verification		
		YG02	YJ06	YJ07	YG02	YJ06	YJ07
sh4	T	T(7)	0	0	T	T	T
Rc	A 14 bp deletion	0	NF(3)	NF(1)	NF	NF	NF
Phr1	An 18 bp deletion (only in *japonica*)	NF(18)	NF(9)	NF(8)	NF	NF	NF

NF not found

reference genomes, respectively. More reads (~10 %) of weedy rice could be mapped to *indica* rice than to the *japonica* rice reference genome, which resulted in an average mapping depth to 3.1–4.7-fold in the *indica* and 2.7–4.0-fold in the *japonica* reference genome, and a higher average mapping coverage for *indica* (80.9–84.6 %) than for *japonica* (71.0–76.1 %). The results indicated that the sequence of weedy rice was more similar to that of *indica* rice.

Genotypes of functional polymorphic sites of domestication genes in weedy rice

The genotypes of the functional polymorphic sites of three main domestication-related genes (*sh4*, *Rc* and *Phr1*) were further determined with PCR verification based on our genomic data (Table 2). Most of the domesticated rice accessions carry a non-shattering *sh4* allele (Zhang et al. 2009). All of our weedy rice accessions, similar to cultivated rice, possess the derived-type allele at the *sh4* locus, which is contrary to their shattering phenotype. It has been reported that the rice pericarp color could turn from red to white due to a 14 bp deletion in the *Rc* gene during domestication (Sweeney et al. 2006). In our case, both the read mapping result and PCR verification suggested that the three weedy rice accessions held a functional *Rc* allele that coordinated with their red pericarp phenotype (Table 2; Fig. 1b).

Most *japonica* rice exhibits a negative phenol reaction due to an 18 bp deletion at the *Phr1* locus, while *indica* *Oryza* species without this deletion mostly show a positive reaction. Therefore, this has been recognized as a significant marker to distinguish between *indica* and *japonica* rice (Yu et al. 2008). In our study, we mapped the reads of all three weedy rice accessions to the *indica* reference, and the 18 bp functional allele region could be mapped by at least 8 reads for each of the three weedy rice samples (Table 2). The results indicated that our weedy rice hosts an *indica* genotype for the *Phr1* gene. Thus, the verification result regarding the domestication-related loci in weedy rice genomes implies that they might harbor the cultivated germplasm and are closer to *indica* rice.

Population structure of weedy rice with *Oryza* members

To investigate the phylogenetic relationships of weedy rice and to examine the genetic population structure of weedy rice and other *Oryza* species, we first performed SNP calling to determine the overall SNP distribution of weedy rice as compared to the *indica* and *japonica* reference genomes (Table S2). Approximately 1 million (0.96) SNPs were identified in weedy rice when *indica* (cultivar 9311) was used as the reference genome, which was less than half of the SNPs (2.10 millions) when *japonica* rice was used (cultivar Nipponbare) as the reference (Table S2). This result indicated that the general genetic background of the three weedy rice in this study is more like that of *indica* rice.

One hundred and twenty-six rice accessions from widespread geographical locations (Table S1) and the *Oryza* genus (Table S1) were selected to construct a three weedy rice maximum likelihood tree. As a result, a total of 46,005 SNPs were genotyped and used for tree construction (Fig. 2a). The overall topology of our phylogenetic tree was consistent with that reported by Huang et al. (2012), which suggested that the selected SNP datasets

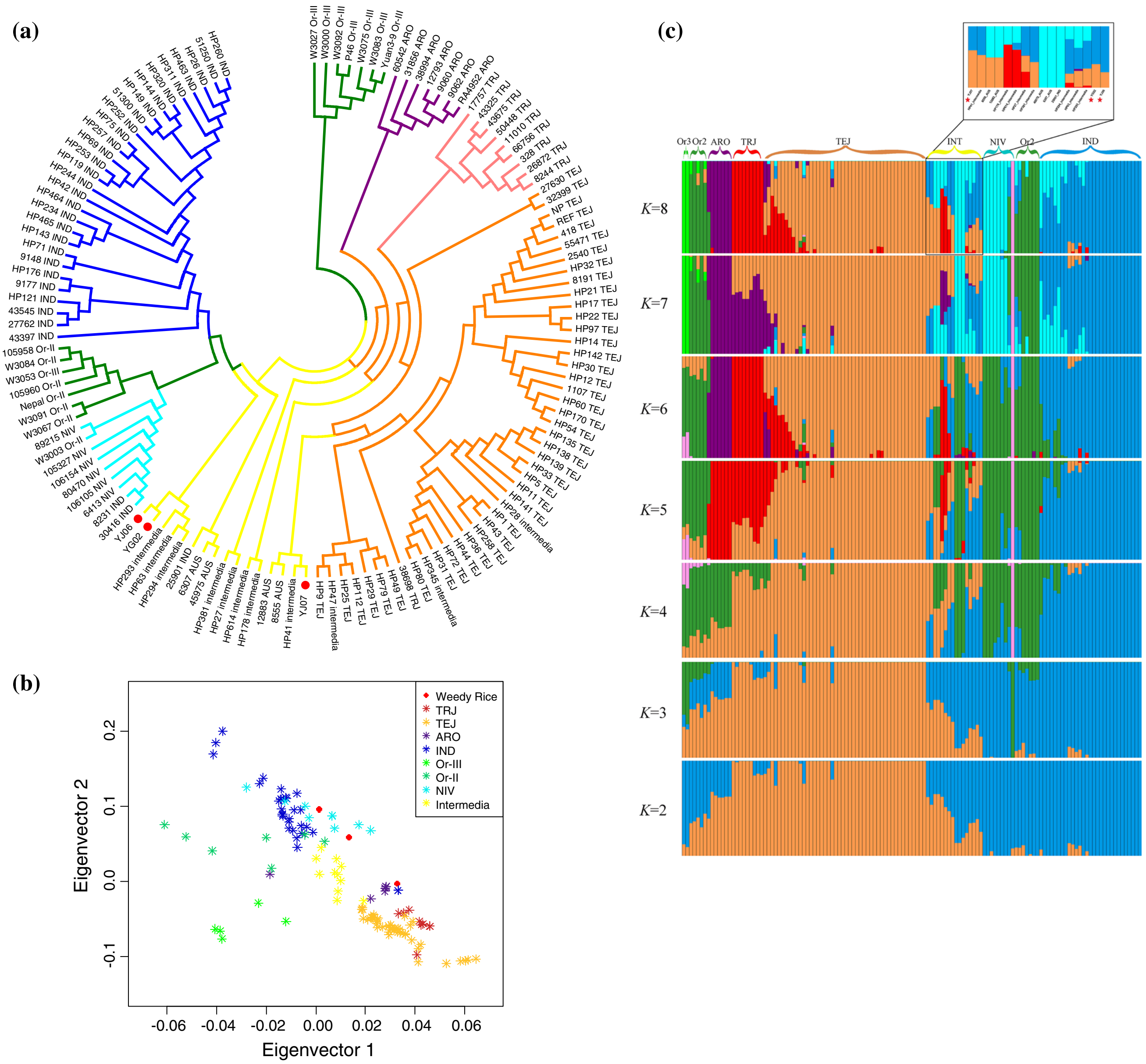

Fig. 2 Population structure analysis of weedy rice with other *Oryza* species. Species are color-coded: *O. rufipogon*: *dark green* for the Or-II subgroup and *light green* for the Or-III subgroup; *O. nivara*: *light blue*; *O. sativa* ssp. *japonica*: *purple* for aromatic *japonica*, *red* for tropical *japonica* and *orange* for temperate *japonica*; *O. sativa* ssp. *indica*: *blue*; intermediate type: *yellow*. The three weedy rice samples are highlighted with *red symbols*. A maximum likelihood phylogenetic tree (**a**), a PCA plot by the first and second eigenvectors (**b**) and maximum likelihood clustering with *K* ranging from 2 to 8 (**c**) were constructed based on 46,005 SNPs. The zoomed-in picture of intermediate accessions is plotted at the *top right corner* of **c**

were effective for inferring the phylogeny of weedy rice. Based on the phylogenetic tree, the three weedy rice samples did not fall into any wild or cultivated rice groups, but rather were in an intermediate state between the *indica* and *japonica* groups. Principle component analysis (PCA) was further conducted and most of the rice accessions were divided by the first and second eigenvectors into wild rice groups (*nivara*, Or-II and Or-III), two main cultivated rice groups including *japonica* (ARO, TRJ and TEJ) and *indica* (*indica* and AUS), and some intermediate accessions between the two cultivated groups (Fig. 2b). The three weedy rice accessions again did not converge into any particular subdivision, but rather were dispersed in the area between the *indica* and *japonica* subgroups. The results indicated that our weedy rice samples were in an intermediate status between *indica* and *japonica*.

We further investigated the genetic ancestry using Frappe, a maximum likelihood based clustering method

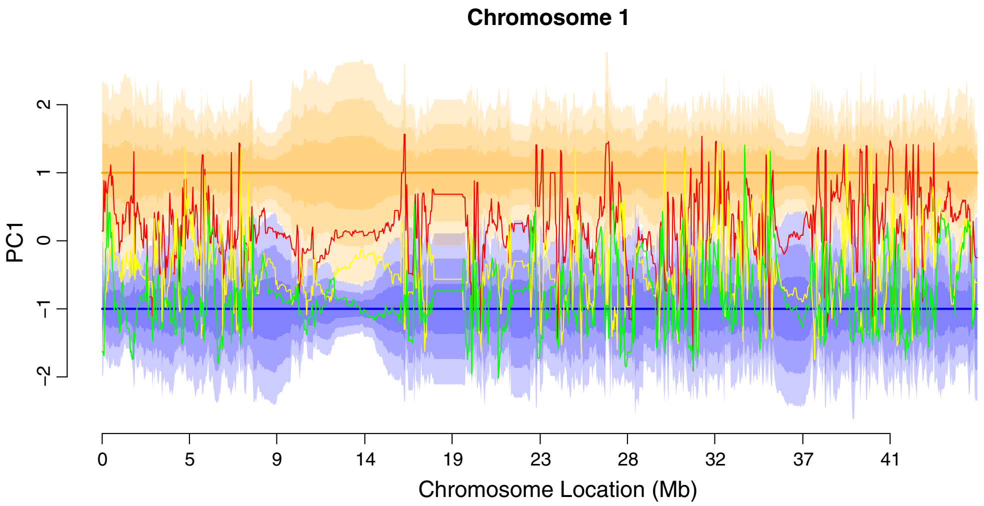

Fig. 3 StepPCO plot for chromosome 1 of the three weedy rice samples. *Indica* (*blue*) and *japonica* (*orange*) were treated as ancestral groups. Chromosomal regions of the admixed genome are attributed to either the *orange* or *blue* ancestry with admixed regions in the middle (YG02: *yellow*; YJ06: *green*; YJ07: *red*). The mean PC1 coordinates for each parental population are centered at 1 and −1, and the progressively lighter shading surrounding the mean of each parental group indicates ±1, ±2 or ±3 SDs from the mean

(Tang et al. 2005) that considers each individual as having originated from a specific number (*K*) of hypothetical ancestral populations, and gives the source proportion of each individual from ancestral populations. The *K*-value was tested from 2 to 8 (Fig. 2c), and each *K* will yield a final converged result as calculated by Frappe. The population structure result was generally consistent with the PCA and phylogenetic analysis. Surprisingly, all three of the weedy rice samples had only two genetic components, despite an increase in the *K*-value. The results suggested that the weedy rice samples in southern China had a mixed genetic background from temperate *japonica* and *indica* rice.

Parental admixture pattern across chromosomes

The chromosomes of a descendant genome usually contain different combinations of the parental ancestral genomic blocks, and we further used a wavelet transform analysis (Pugach et al. 2011) to confirm the above potential admixture pattern across the whole genome of the three weedy rice individuals. Stepwise principal component analysis was performed along 12 chromosomes of weedy rice and their respective ancestral populations, which gave insights into where the stretches of the admixed genome were inherited. After the genotypes of the three weedy rice and 39 parental accessions were intersected, a total of 27,557 SNPs were used for the parental admixture analysis (see details in the “Materials and methods”). The StepPCO plots for 12 chromosomes of the three weedy rice individuals (Fig. S1) showed a clear admixture pattern in which the chromosomal segments of the three weedy rice accessions switched between the two ancestral populations (chromosome 1 as an example is shown at Fig. 3; temperate *japonica* in orange and *indica* in blue). Moreover, the three weedy rice individuals did not share the same admixture pattern, which suggested that genetic divergence and ancestral inheritance were different among the three samples. We also found that, based on the StepPCO plots, higher proportions of the blocks belonged to the *indica* population, which agreed with the Frappe genetic ancestry results.

Novel genomic variations in weedy rice

Compared to the current dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), we detected unique or novel and common (potentially fixed) SNPs in the three weedy accessions. A total of 22,443 fixed novel SNPs in the weedy rice genome, which reside in 4492 rice genes, were found (data not shown). Among these 22,443 fixed novel SNPs, 3,075 SNPs had moderate effects on 2,201 genes due to non-synonymous codon mutations (‘NON_SYNONYMOUS_CODING’). We performed a GO enrichment study for the 2,201 genes with moderate impact, and observed that some of the enriched GO categories were related to environmental adaptation, such as ‘defense response’, ‘response to chemical stimulus’, and ‘response to stress’, while others were related to ‘regulation of nitrogen compound metabolic process’ (Table S3). We furthered the enrichment analysis with a plant metabolic pathway database (PMN/PlantCyc), which contains more than 800 pathways and their catalytic enzymes and genes, as well as compounds from over 350 plant species (http://www.

plantcyc.org). Most of the enriched pathways were related to hormone biosynthesis (such as IAA, cytokinins and brassinosteroid) (Table S3). The above results suggested that these novel fixed SNPs in the weedy rice genome should have mutated under positive selection and play a vital role in their environmental adaptation.

Discussion

Origin of weedy rice

Unlike alternative strategies using limited markers, whole-genome sequencing can detect millions of SNPs and offers a more confident phylogenetic relationship and admixture patterns within the weedy rice. As shown in this study, it is very convenient for us to deepen our studies of weedy rice at the genome level by incorporating the genomic data generated from weedy rice with the massive amount of genomic data available to the rice community (e.g., Huang et al. 2012; Xu et al. 2012).

In this study, three weedy rice samples from Jiangsu province were re-sequenced. An NGS genomic data analysis revealed that weedy rice in this region may have originated from *indica-japonica* hybridization. Several results support this speculation: (1) The phylogenetic tree and PCA study based on over 40 k SNPs with 126 other *Oryza* accessions indicated that the three weedy rice samples mainly cluster with admixture accessions between *japonica* and *indica*, (2) According to the ancestry study based on Frappe, the weedy genomes had only two genomic components (temperate *japonica* and *indica*), which also suggests local introgression into weedy rice from these two domesticated rice types that are widely cultivated in China, (3) The parental admixture pattern across all the weedy rice chromosomes illustrated that the chromosomal segments of the three weedy rice accessions alternate between the temperate *japonica* and *indica* populations, indicating introgression between the two groups, (4) All cultivated and weedy rice share similar *sh4* haplotypes and contain a single derived mutation at *sh4* that is associated with decreased seed shattering phenotype. The *Rc* genotype in weedy rice may have been inherited from some cultivars with the ancestral *Rc* alleles and fixed under natural selection (see next section). This hypothesis regarding the origin of Jiangsu weedy rice is consistent with several previous studies by researchers from local institutes who used SSR markers to explore the genetic diversity of weedy rice populations (e.g., Shao et al. 2011; Wang et al. 2012). In addition, the heterozygous rates of the three weedy rice lines (YG02, YJ06 and YJ07) were 5.0–6.2 % based on their genomic data, suggesting that the hybridization event did not occur recently.

A genome-wide picture of introgression with autosomal resolution could enhance our knowledge of evolution via hybridization, which could provide insights into how the genomic regions or genes of an admixture individual are shaped by recombination. In addition, the genetic architecture of adaptation to abiotic and biotic conditions could be inferred by assessing introgression with high resolution (Hufford et al. 2013). As mentioned above, several researchers have noted that weedy rice could have unexpectedly evolved from its cultivated relatives (Ishikawa et al. 2005; Reagon et al. 2010; Thurber et al. 2010; Xiong et al. 2012). For example, Ishikawa et al. (2005) performed isozyme and morpho-physiological analyses, and proposed that weedy rice originated from *indica-japonica* hybridization. Xiong et al. (2012) demonstrated that some percentage of weedy rice-like plants could have been segregated from the offspring of inter-subspecies and inter-varietal hybridization based on seven *indica*/*japonica* and four *indica*/*indica* combinations. In particular, the frequency of weedy rice-like plants derived from *indica*/*japonica* hybridization was greater than that of such plants derived from *indica*/*indica* hybridization. The weedy rice-like plants segregated from the progeny of *indica*/*japonica* hybridization are phenotypically more like *indica* plants. Our findings, which show that *indica*-type weedy rice was derived from *indica*/*japonica* hybridization, support Xiong et al.'s observations. We performed whole-genome re-sequencing and confirmed that weedy rice can originate from the hybridization of *indicia* and temperate *japonica* rice. This should be a critical concept in rice breeding, since many rice breeders have been trying to hybridize *indica* and *japonica* rice to improve rice production in China (Chen et al. 2007; Sun et al. 2012).

Weedy rice in southern China is more likely to be similar to *indica* (Zhang et al. 2008), while weedy rice in northern China is more likely to be similar to *japonica* (Sun et al. 2013). We believe that the environment, and especially the temperature and humidity, is responsible for the fact that *indica* is significantly more represented in these weedy rice genomes than *japonica*. For example, the *Phr1* allele was not selected in *indica* rice, which is cultivated in areas with higher levels of humidity and temperature, like southern China. However, in drier environments, like northern China, the *Phr1* allele was selected in *japonica* rice to prevent grain discoloration during storage (Sun et al., 2013). With regard to the seed dormancy phenotype, secondary dormancy is a critical characteristic in *indica* varieties to retain viability during hot and humid conditions (Vaughan et al. 2008). Therefore, the related genes of the three weedy rice individuals sampled from the lower Yangtze River (southern China) in this study may be selected under hot and humid environmental conditions, and therefore present phenotypes and genotypes that are more like those of *indica* rice.

Novel genomic variants in weedy rice

Weedy rice has become a serious problem in China over the past 20 years. Interestingly, a large set of novel fixed SNPs was found in weedy rice, which were absent from the currently available rice dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). These novel variants in weedy rice may have accumulated through mutation that occurred since divergence from the hybridization of *indica* and *japonica.* Another possible explanation for the existence of these novel variants is the undersampling of *Oryza* diversity (not included in dbSNP as of 2014-04-05). Thereafter, both favored natural and artificial selection could have fixed these novel SNPs for weedy rice. Based on the results regarding GO and PlantCYC pathway enrichment, the functions of the potentially fixed genes are mostly enriched in response to the environment, such as 'response to stress' or 'response to chemical stimulus', or are related to the 'regulation of nitrogen compound metabolic process' or 'hormone biosynthesis', which enables them to compete with crop plants. Weedy rice had to evolve at least two phenotypes: the seeds had to be able to overwinter under natural field conditions and it had to be able to match the growth rhythm of cultivated rice in the field. The mutations that favor environmental adaption for weedy rice may be fixed under natural selection. Furthermore, due to less-intensive management, alleles associated with ecological adaptability tend to reproduce at a high frequency or are selected in the weedy rice population (Sun et al. 2013). For domesticated genes, one piece of evidence that does not support our hypothesis that weedy rice originated from the hybridization of domesticated *indica-japonica* is the *Rc* gene, which controls the red pericarp phenotype. In this study, the three weedy rice samples had a functional *Rc* allele that was the same as that of wild rice. This result has also been observed in other weedy rice (e.g., Gross et al. 2010). Previous studies have indicated that *Rc* is a pleiotropic gene that controls both the pericarp color and seed dormancy (Gu et al. 2011). As mentioned above, seed dormancy must contribute to the ability of weedy rice seeds to overwinter. Natural selection most likely favored red rice because of such seed dormancy. In contrast, domestication favored a white pericarp, and the associated pleasant appearance and taste of grains (absence of proanthocyanidins in the pericarp tissue) (Sweeney et al. 2007), which is caused by a flavonoid biosynthetic pathway downstream of the pleiotropic gene (Gu et al. 2011). Since some cultivated varieties harbor the ancestral *Rc* alleles, the *Rc* allele of weedy rice may be inherited from these cultivars and fixed via natural selection.

In this study, we detected over 20 k fixed novel SNPs across the weedy genome. The newly identified SNPs will be useful as markers for pinpointing important genes and for the marker-assisted mapping of important weedy rice traits such as seed dormancy and shattering. With respect to the shattering gene *sh4* in our study, all three of the weedy rice samples held the single derived mutation associated with decreased seed shattering, similar to the cultivars. Meanwhile, similar haplotypes at the major shattering loci *sh4* have been observed in US weedy rice groups and cultivated lines. It has been suggested that weedy rice may have re-acquired the seed dispersal trait after divergence from its progenitors (Thurber et al. 2010). Therefore, while the molecular mechanism of seed shattering in weedy rice has not yet been determined, several QTLs have been mapped (Subudhi et al. 2014) and genes other than *sh4* could have played an important role in shattering (Zhu et al. 2012; Zhou et al. 2012).

Implications for weedy rice control

Breeding programs that involve inter-subspecies hybridization between *indica* and *japonica* to improve rice cultivars are becoming widespread, and the introduction of germplasm of *indica* subspecies into temperate *japonica* has been reported to be largely responsible for the increased rice yield in northern China (Chen et al. 2007; Sun et al. 2012). However, as mentioned above, field experiments by Xiong et al. (2012) revealed that inter-subspecies and inter-varietal hybridization had a high chance for the emergence of weedy rice-like progeny, particularly for *indica-japonica* hybridization. Our genomic data provide unambiguous evidence to support this possibility. Therefore, these results provide several implications for controlling weedy rice. First, more attention should be paid to the management of inter-subspecies hybrids during rice breeding. The control of breeding fields that are the origin of weedy rice and the prevention of weed dispersion from the area of origin are key steps to controlling these weeds. More manpower and proper farming practices should be used. Second, unexpected natural hybridization events might occur in rice fields, and therefore, we should expect that weedy rice-like plants will frequently be generated in our current rice production system and may always accompany rice production. As an evidence to support this, Jiangsu weedy rice is mentioned in ancient Chinese books (usually called "Ludao" or "Lvdao") (Ding 1983; Xiuling You, personal communication). This indicates that weedy rice has existed for a long time, maybe from the beginning of rice cultivation. Therefore, the use of proper farming practices such as transplantation rather than direct seeding is another important way to control weedy rice. As mentioned above, weedy rice has become an important issue in China over the past 20 years, which should show what happens when we ignore the rice ecosystem. Third, GM rice varieties/lines which possess genes that confer resistance to herbicides and diseases, and tolerance to environmental stress (Lu and Snow 2005) should be carefully controlled to prevent the transference of such genes into the weedy rice population, which could make them even more robust in terms of environmental adaptation.

Conclusions

In this study, we re-sequenced the genomes of weedy rice samples from the lower Yangtze region, China, and investigated their phylogenetics, population structure and chromosomal admixture patterns. The phylogenetic tree and principle component analysis based on 46,005 SNPs with 126 other *Oryza* species suggested that the three weedy rice accessions were intermediate between *japonica* and *indica* rice. An ancestry inference study further demonstrated that weedy rice had only two dominant genomic components (temperate *japonica* and *indica*) Thus, our results suggested that the weedy rice sampled in our study may have originated due to *indica-japonica* hybridization Furthermore, 22,443 novel fixed SNPs were detected in the weedy genomes and may have been generated after *indica-japonica* hybridization to promote environmental adaptation. Our research strongly suggests that weedy rice originated from hybridization between *indica* and *japonica* rice and provides important insights for controlling this notorious weed.

Author contribution J.Q., W.D.W., L.F.M., Z.X.L., L.C., and H.Q.Z. carried out the analysis. J.Q. C.Y.Y., and L.J.F. prepared the manuscript. F.F. performed the experiments. S.Q., Y.L.L., J.W.Z., and L.B.G. reviewed the manuscript and gave insightful suggestions.

Acknowledgments We thank Xunbin Guo and Junsen Jiao for collecting the weedy rice samples. This work was financially supported in part by grants from the China Agriculture Research System (CARS-01-02A) and the Natural Science Foundation of China (31171863).

Conflict of interest We declare that we have no financial or personal relationships with other people or organizations that may have inappropriately influenced our work, and there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in or the review of the manuscript.

Ethical standard This manuscript has not been published previously and there is no prior interaction with Planta. We confirm that there are no known conflicts of interest associated with the publication. The manuscript has been read and approved by all of the authors.

References

Allston RFW (1846) The rice plant. DeBow's Rev I:320–356 (Available online from U. Michigan)

Baki BB, Chin DV, Mortimer M (eds) (2000) Wild and weedy rice in rice ecosystems in Asia—a review, Los Baños. International Rice Research Institute, Philippines 118

Bres-Patry C, Bangratz M, Ghesquiere A (2001) Genetic diversity and population dynamics of weedy rice in Camargue area. Genet Sel Evol 33:S425–S440

Cao Q, Lu BR, Xia H, Rong J, Sala F, Spada A, Grassi F (2006) Genetic diversity and origin of weedy rice (*Oryza sativa f. spontanea*) populations found in North-eastern China revealed by simple sequence repeat (SSR) markers. Ann Bot 98:1241–1252

Chen W, Xu Z, Zhang L, Zhang W, Ma D (2007) Theories and practices of breeding *japonica* rice for super high yield. Sci Agric Sin 40:869–874 (in Chinese)

Cingolani P, Platts A, le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso–2; iso–3. Fly 6:80–92

De Wet JMJ, Harlan JR (1975) Weeds and domesticates: evolution in the man-made habitat. Econ Bot 29:99–107

Ding Y (1983) Classification of Chinese culture rice. Anthology of Ding Ying's paper in rice. Agricultural Publishing Company, Beijing (in Chinese)

Du Z, Zhou X, Ling Y, Zhang ZH, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38:W64–W70

Ferrero A, Vidotto F, Balsari P, Airoldi G (1999) Mechanical and chemical control of red rice (*Oryza sativa L-var. sylvatica*) in rice (*Oryza sativa L.*) pre-planting. Crop Prot 18:245–251

Gealy DR, Tai TH, Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci 50:333–339

Gross BL, Reagon M, Hsu SC, Caicedo AL, Jia Y, Olsen KM (2010) Seeing red: the origin of grain pigmentation in US weedy rice. Mol Ecol 19:3380–3393

Gu XY, Foley ME, Horvath DP, Anderson JV, Feng JH, Zhang LH, Mowry CR, Ye H, Suttle JC, Kadowaki K, Chen ZX (2011) Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. Genetics 189:1515–1524

Harlan J (1992) Crops and Man. American Society of Agronomy, Crop Science Society of America, Madison, pp 117–130

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501

Huang X, Lu T, Han B (2013) Resequencing rice genomes: an emerging new era of rice genomics. Trends Genet 29:225–232

Hufford MB, Lubinksy P, Pyhajarvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J (2013) The genomic signature of crop-wild introgression in maize. PLoS Genet 9:e1003477

Ishikawa R, Toki N, Imai K, Sato YI, Yamagishi H, Shimamoto Y, Ueno K, Morishima H, Sato T (2005) Origin of weedy rice grown in Bhutan and the force of genetic diversity. Genet Resour Crop Evol 52:395–403

Jiang ZX, Xia HB, Basso B, Lu BR (2012) Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. Theor Appl Genet 124: 309–322

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:U354–U357

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

Liang DY, Qiang S (2011) Current situation and control strategy of weedy rice in China. China PlantProt 31:21–24 (in Chinese)

Londo JP, Schaal BA (2007) Origins and population genetics of weedy red rice in the USA. Mol Ecol 16:4523–4535

Lu BR, Snow AA (2005) Gene flow from genetically modified rice and its environmental consequences. Bioscience 55:669–678

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Mortimer M, Pandey S, Piggin C (2000) Weedy rice: approaches to ecological appraisal and implications for research priorities. In: Baki BB, Chin DV, Mortimer M (eds.) Proceedings of wild and weedy rice in rice ecosystems in Asia—a review. Limited proceedings no. 2, International Rice Research Institute, Los Baños, pp 97–105

Noldin JA, Chandler JM, McCauley GN (1999) Red rice (*Oryza sativa*) biology. I. Characterization of red rice ecotypes. Weed Technol 13:12–18

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7:e30619

Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M (2011) Dating the age of admixture via wavelet transform analysis of genome-wide data. Genome Biol 12:R19

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL (2010) Genomic patterns of nucleotide diversity in divergent populations of US weedy rice. BMC Evol Biol 10:180

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Shao Q, Dai WM, Zhang LJ, Song XL, Qiang S (2011) Genetic diversity and origin of weedy rice in central region of Jiangsu Province, China. Acta Agronomica Sinica 37:1324–1332 (in Chinese with English abstract)

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–311

Subudhi PK, Singh PK, DeLeon T, Parco A, Karan R, Biradar H, Cohn MA, Sasaki T (2014) Mapping of seed shattering Loci provides insights into origin of weedy rice and rice domestication. J Hered 105:276–287

Sun J, Liu D, Wang JY, Ma DR, Tang L, Gao H, Xu ZJ, Chen WF (2012) The contribution of intersubspecific hybridization to the breeding of super-high-yielding japonica rice in northeast China. Theor Appl Genet 125:1149–1157

Sun J, Qian Q, Ma DR, Xu ZJ, Liu D, Du HB, Chen WF (2013) Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. New Phytol 197:290–299

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283–294

Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet 3:1418–1424

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tang LH, Morishima H (1996) Genetic characteristics and origin of weedy rice. In: Origin and differentiation of Chinese cultivated rice. China Agricultural University Press, Beijing, pp 211–2180

Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. Genet Epidemiol 28:289–301

Thurber CS, Reagon M, Gross BL, Olsen KM, Jia Y, Caicedo AL (2010) Molecular evolution of shattering loci in US weedy rice. Mol Ecol 19:3271–3284

Vaughan LK, Ottis BV, Prazak-Havey AM, Bormans CA, Sneller C, Chandler JM, Park WD (2001) Is all red rice found in commercial rice really Oryza sativa? Weed Sci 49:468–476

Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. Plant Sci 174:394–408

Wang J, Yang J, Yang JH, Chen ZD, Fan FJ, Zhu JY, Zhong WG (2012) Genetic similarity and origin of weedy rice in Yangzhong region of Jiangsu province. Jiangsu J Agric Sci 28:748–753 (in Chinese with English abstract)

Xiong HB, Xu HY, Xu Q, Zhu Q, Gan SX, Feng DD, Zhang XL, Xie XD, Zhang H, Li J, Han YL, Wen JC, Li DX, Shi YM, Wei XH, Chen LJ (2012) Origin and evolution of weedy rice revealed by inter-subspecific and inter-varietal hybridizations in rice. Mol Plant Breed 10:131–139 (in Chinese with English abstract)

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105–111

Yi X, Du Z, Su Z (2013) PlantGSEA: a gene set enrichment analysis toolkit for plant community. Nucleic Acids Res 41:W98–W103

Yu Y, Tang T, Qian Q, Wang Y, Yan M, Zeng D, Han B, Wu CI, Shi S, Li J (2008) Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. Plant Cell 20:2946–2959

Zhang J, Nilda RB, Ma K, Zhou YJ, Geng RM, Yu LQ (2008) Genetic diversity and relationship of weedy rice in Taizhou City, Jiangsu Province, China. Rice Sci 15:295–302

Zhang LB, Zhu QH, Wu ZQ, Ross-Ibarra J, Gaut BS, Ge S, Sang T (2009) Selection on grain shattering genes and rates of rice domestication. New Phytol 184:708–720

Zhang Z, Dai W, Song X, Qiang S (2014) A model of the relationship between weedy rice seed-bank dynamics and rice-crop infestation and damage in Jiangsu Province, China. Pest Manag Sci 70:716–724

Zheng XW, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328

Zhou Y, Lu D, Li C, Luo J, Zhu BF, Zhu J, Shangguan Y, Wang Z, Sang T, Zhou B, Han B (2012) Genetic control of seed shattering in rice by the APETALA2 transcription factor shattering abortion1. Plant Cell 24:1034–1048

Zhu Y, Ellstrand NC, Lu BR (2012) Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus sh4 played a minor role in Asian rice domestication. Ecol Evol 2:2106–2113