Theor Appl Genet (2014) 127:297–307
DOI 10.1007/s00122-013-2218-2

ORIGINAL PAPER

Highly efficient genotyping of rice biparental populations by GoldenGate assays based on parental resequencing

Wei Chen · Haodong Chen · Tianqing Zheng · Renbo Yu · William Bryan Terzaghi · Zhikang Li · Xing Wang Deng · Jianlong Xu · Hang He

Received: 12 May 2013 / Accepted: 14 October 2013 / Published online: 5 November 2013

Abstract
Key message **A new time- and cost-effective strategy was developed for medium-density SNP genotyping of rice biparental populations, using GoldenGate assays based on parental resequencing.**
Abstract Since the advent of molecular markers, crop researchers and breeders have dedicated huge amounts of effort to detecting quantitative trait loci (QTL) in biparental populations for genetic analysis and marker-assisted selection (MAS). In this study, we developed a new time- and cost-effective strategy for genotyping a population of progeny from a rice cross using medium-density single nucleotide polymorphisms (SNPs). Using this strategy, 728,362 "high quality" SNPs were identified by resequencing Teqing and Lemont, the parents of the population. We selected 384 informative SNPs that were evenly distributed across the genome for genotyping the biparental population using the Illumina GoldenGate assay. 335 (87.2 %) validated SNPs were used for further genetic analyses. After removing segregation distortion markers, 321 SNPs were used for linkage map construction and QTL mapping. This strategy generated SNP markers distributed more evenly across the genome than previous SSR assays. Taking the *GW5* gene that controls grain shape as an example, our strategy provided higher accuracy (0.8 Mb) and significance (LOD 5.5 and 10.1) in QTL mapping than SSR analysis. Our study thus provides a rapid and efficient strategy for genetic studies and QTL mapping using SNP genotyping assays.

W. Chen, H. Chen and T. Zheng equally contributed to this work.

Communicated by T. Sasaki.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2218-2) contains supplementary material, which is available to authorized users.

W. Chen · H. Chen · R. Yu · X. W. Deng · H. He (✉)
Peking-Yale Joint Center for Plant Molecular Genetics and Agro-Biotechnology, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, China
e-mail: hehang@pku.edu.cn

H. Chen · R. Yu · X. W. Deng · H. He
Shenzhen Institute of Crop Molecular Design, Shenzhen 518107, China

T. Zheng · Z. Li · J. Xu (✉)
Institute of Crop Sciences, National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: xujlcaas@126.com

R. Yu · X. W. Deng
Frontier Laboratories of Systems Crop Design Co. Ltd., Beijing 100085, China

W. B. Terzaghi
Department of Biology, Wilkes University, Wilkes-Barre, PA 18766, USA

Z. Li · J. Xu
Shenzhen Institute of Breeding and Innovation, Chinese Academy of Agricultural Sciences, Shenzhen 518083, China

Introduction

Most important agronomic traits are complex quantitative traits controlled by multiple quantitative trait loci (QTL).

The application of molecular genetic markers enabled crop researchers and breeders to detect QTL in biparental populations, in which the trait typically shows continuous variation. Huge amounts of effort have been invested in the detection of QTL in many crop plants for implementation of marker-assisted selection (MAS) breeding or molecular breeding (Würschum 2012). DNA polymorphisms have been the basis for the development of various traditional molecular markers, from restriction fragment length polymorphisms (RFLPs) to simple sequence repeats (SSRs) (McCouch et al. 1988; Wu et al. 1993). Traditional molecular markers have been prevalently used in crop genetic research. However, traditional molecular marker assays are labor-intensive and time consuming (Xie et al. 2010).

New sequencing technology has enabled the discovery of numerous single nucleotide polymorphisms (SNPs) by mapping the entire genomic sequences of individuals to the reference genome. SNPs have become more popular than other traditional molecular markers because they are the most abundant form of genetic variation across genomes, and automated technologies have been developed for high-throughput analysis based on microarray or next-generation sequencing. Various genotyping platforms for SNP markers have been developed and successfully used for the construction of genetic maps (Xie et al. 2010), investigation of genome-wide linkage disequilibrium (Cao et al. 2011), detection of QTL in cross-derived populations (Wang et al. 2011), understanding of population evolutionary history (Chen et al. 2011; Huang et al. 2012a), and discovery of marker trait associations in association mapping studies (Huang et al. 2010, 2012b; Parida et al. 2012).

With the rapid progress of next-generation sequencing (NGS), genotyping by sequencing techniques (GBS) are being developed for genotyping studies (Hamilton and Buell 2012). For different genome size and funding, sequencing is carried out either on genomic DNA or focused on complexity-reduced DNA (Saintenac et al. 2013). For whole-genome sequencing method, high-throughput genotyping and estimation of recombination points based on resequencing of recombinant inbred lines were recently described, and even though the sequencing coverage was insufficient (only 0.02× coverage per line), the method provided plenty of useful information (1.2 M SNPs) (Huang et al. 2009; Xie et al. 2010). Another strategy includes approaches using complexity-reduced DNA generated based on digestion of genomic DNA by restriction enzymes, including CRoPS (van Orsouw et al. 2007), RAD-seq (Baird et al. 2008), Cornell genotyping by sequencing (Elshire et al. 2011; Poland et al. 2012), ddRADseq (Peterson et al. 2012), GR-RSC (Maughan et al. 2009, 2010), MSG (Andolfatto et al. 2011), DArT-Seq (Schouten et al. 2012). These approaches offer benefits including lower cost in the future, low ascertainment bias, detection of dynamic range of variants (SNPs, small insertions, deletions, and microsatellites), and even ability to analyze samples in the absence of a reference genome (Baird et al. 2008; Elshire et al. 2011; Poland et al. 2012). The rapid evolution of sequencing technologies is driving the development of low-cost, fast and accurate GBS methods for agrigenomics researches.

As a trusty and flexible genotyping platform, the Illumina GoldenGate assay can multiplex from 96 to 1, 536 SNPs in a single reaction over a 3-day period (Fan et al. 2003). This assay has been widely used to validate large numbers of SNPs in many crops, including rice (Yamamoto et al. 2010; Zhao et al. 2010; Thomson et al. 2012), maize (Lu et al. 2009), soybean (Hyten et al. 2008), wheat (Akhunov et al. 2009), and barley (Close et al. 2009). Among these versatile SNP platforms, 89 % of the SNPs generated reliable allelic discrimination and about 40 % of the SNPs had polymorphisms in each population. Recently, Hyten et al. (2008) genotyped recombinant inbred lines of soybean using a custom 384-plex GoldenGate assay, and Thomson et al. (2012) developed several 384-SNP arrays for different germplasm pools in rice: *Indica/Indica, Indica/Japonica, Japonica/Japonica, Indica/O. rufipogon, and Japonica/O. rufipogon*. However, only about 150–200 polymorphic SNPs could be obtained for biparental crosses. Therefore, designing specific SNP arrays for particular populations may be a way to increase the numbers of informative polymorphic SNPs.

Introgression lines (ILs), also termed 'chromosome segment substitution lines, or CSSL', can be generated by introgressing small chromosomal segments from the donor parent into the recurrent parent by backcrossing and selfing. Introgression libraries genetically dissecting both qualitative and quantitative traits have been very useful for plant breeding (Zamir 2001; Li et al. 2005). Accordingly, a great deal of effort has been invested over decades in detecting QTL in ILs (Doi et al. 1997, 2003; Ghesquiere et al. 1997; Sobrizal et al. 1999; Ahn et al. 2002; Li et al. 2005; Xu et al. 2005; Tian et al. 2006; Zheng et al. 2007).

In this study, we developed a new strategy for SNP genotyping of the population derived from the cross between Teqing and Lemont. First, we obtained "high quality" SNPs by resequencing the parents. Next, we genotyped the population with 384 specific informative SNPs using the Illumina GoldenGate assay based on the Illumina BeadXpress system. Finally, we performed QTL analysis of grain shape and milling quality traits, with better accuracy and significance than previous results obtained using SSR markers (Zheng et al. 2007). This study thus indicates that medium-density SNP genotyping using specific informative SNPs was time- and cost-effective for QTL mapping in biparental population studies.

Materials and methods

Plant materials and DNA extraction

The high-yielding *indica* rice variety Teqing was used as the recurrent parent to cross with the commercial semi-dwarf *japonica* rice variety Lemont. The F_1 plants were simultaneously backcrossed to Teqing, resulting in 100 BC_1F_1 plants. Consecutive backcrossing was carried out to the BC_2F_1, BC_3F_1 or BC_4F_1 generations. Stable lines were developed after more than two rounds of selfing, resulting in a set of 143 introgression lines (ILs) in the Teqing background, including 78 BC_2F_5, 50 BC_3F_4, and 15 BC_4F_3 (Xu et al. 2005; Zheng et al. 2007). We extracted mixed genomic DNA from approximately ten seedlings of each IL and of the two parents using plant DNA extraction kits according to the manufacturer's instructions (Qiagen).

Next-generation sequencing

Paired-end sequencing libraries were constructed using about 10 μg genomic DNA following the manufacturer's instructions (Illumina). 100 bp short reads at each end were generated following the manufacturer's base calling pipeline (SolexaPipeline-1.3, Illumina).

SNP calling for genotyping

High-throughput reads of Teqing and Lemont were mapped to the reference Nippon bare genome (MSU build 6.1) using bwa version 0.6 (Li and Durbin 2010). Then, the SNPs from Teqing and Lemont with high-quality reads of more than eight were extracted using Perl scripts. Further, only SNPs without other SNPs/InDels in 100 bp of flanking sequence were kept. After that, we submitted all the SNPs to Illumina to determine their design ability rank scores that indicate the success rates for the GoldenGate assay. Finally, a total of 384 SNPs distributed evenly across the genome (about one marker per 1 Mb) with a relatively high success rate according to the design ability rank score were further selected for Oligo Pool Assay (OPA) synthesis.

SNP genotyping using the Illumina GoldenGate assay

We carried out SNP genotyping using the Illumina GoldenGate assay. We used 250 ng of DNA for each IL. A total of 384 SNP genotypes were generated from the BeadXpress system using the GoldenGate assay and vericode technology, which can genotype 384 markers in a single plate. The intensity data for each SNP were normalized and cluster positions were assigned using Illumina GenomeStudio Genotyping Analysis Module version 1.9.4. All SNP loci were manually checked to remove errors in calling the homozygous and heterozygous clusters. SNPs with GenTrain scores less than 0.25 or call rates lower than 90 % were removed. Finally, 335 high-quality SNPs were kept for further analysis.

Segregation distortion calculation

The Chi square test was used to determine whether the marker ratio was consistent with normal Mendelian segregation in three different generations of ILs derived from Teqing and Lemont (77 BC_2F_5 lines, 49 BC_3F_4 lines, and 17 BC_4F_3 lines). χ^2 values used to estimate the P value were calculated using the following formula:

$$\chi^2 = \sum_{i=1}^{2} \left(|O_i - T_i| - 0.5\right)^2 / T_i,$$

where O_i is the number of markers for an individual parent, and T_i is the theoretical value of the two parents for three different generations of ILs (BC_2F_5 for 1:7, BC_3F_4 for 1:15, and BC_4F_3 for 1:31). The P value was adjusted using the Benjamini and Hochberg method (Benjamini and Hochberg 1995) to control for the false discovery rate.

Linkage map construction and QTL mapping

The genetic linkage map was constructed using ICIMapping version 3.1 (Li et al. 2010). After removing 14 SNP segregation distortion markers, 321 evenly distributed and successfully genotyped polymorphic SNP markers were used to construct the linkage map.

We performed QTL analysis of rice agronomic traits using these SNP markers in comparison with SSR markers to investigate the efficiency of this strategy for detecting QTL of complex traits. The phenotypic data including grain length (GL), grain width (GW), grain thickness (GT), brown rice percentage (BR), milled rice percentage (MR), head rice percentage (HR), and the two derived traits, ratios of GL/GT (RGLT) and GL/GW (RGLW) of the 143 TQ-ILs were previously obtained (Zheng et al. 2007). QTL analysis was performed using the interval mapping (IM) method by QTL ICIMapping (Li et al. 2010). The permutation method (Churchill and Doerge 1994) was used to obtain empirical thresholds for claiming QTL based on 1,000 runs of randomly shuffling the trait values, which ranged from 4.2 for GT by 160 SSR markers to 11.3 for GL by 321 SNP markers.

Results

A stepwise strategy (Fig. 1) was followed to design the markers with resequencing data and characterize QTL

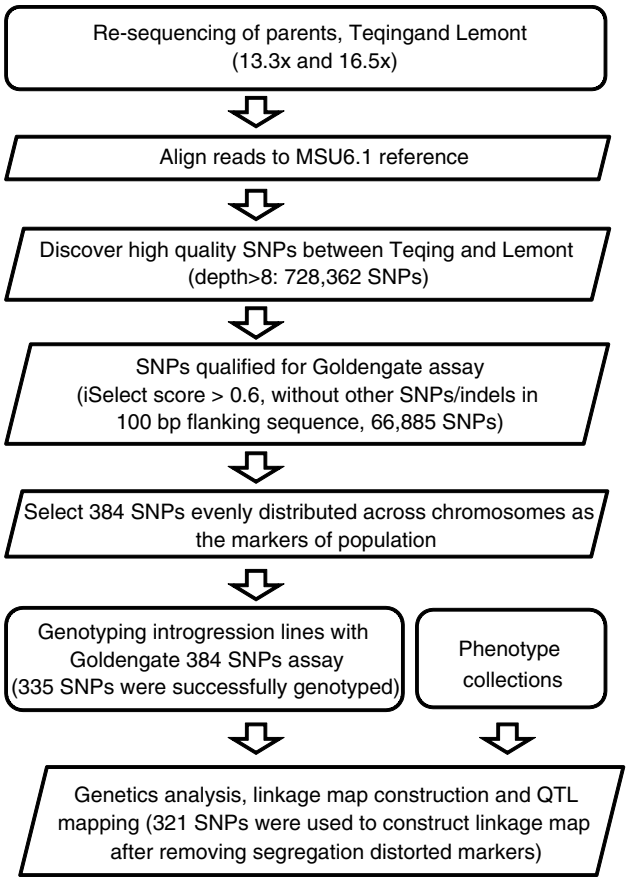

Fig. 1 Workflow for the high-efficiency genotyping of a biparental population derived from a cross between Teqing and Lemont

using the GoldenGate assay in 143 ILs derived from the cross between Teqing and Lemont.

SNP discovery between two parents

We resequenced the entire genomes of Teqing and Lemont, the parents of the ILs. Resequencing Lemont yielded 5.2G bases from 26.1 million 100-bp paired-end reads. All the reads were mapped to the reference Nipponbare genome version 6.1 (Ouyang et al. 2007) using bwa (Li and Durbin 2010), and covered approximately 92.4 % of the Nipponbare genome, with an average depth of 14x. The Teqing resequencing data generated by Zhou et al. (unpublished) contained 147.7 million 45-bp reads, covering approximately 93.6 % of the Nipponbare genome (Fig. S1).

The sequence data were analyzed for SNP discovery using the Samtools application (Li et al. 2009). 2,794,158 SNPs were detected between Teqing and Lemont with at least one read covered. Stricter criteria were required to filter and verify these SNP to reliably design the microarray. We, therefore, defined a "high quality SNP" as (a) the number of reads in both Teqing and Lemont were >8 at the locus and (b) minor allele frequencies were <10 % in Teqing and Lemont. Using these criteria, 728,362 "high quality" SNPs were identified (Fig. 2; Table S1). Of these, 56 "high quality" SNPs were randomly selected for PCR verification. The PCR products obtained using Teqing and Lemont as templates were sequenced, showing that 100 % of these SNPs between Teqing and Lemont were correct (Table S2).

The 728,362 "high quality" SNPs were distributed along all the chromosomes, except for some regions where polymorphisms between Teqing and Lemont were unavailable. The average density was 196.1 SNP per 100 kb. The density of "high quality" SNPs varied across chromosomes, with the highest density (233.3/100 kb) on chromosome 9 and the lowest density (147.4/100 kb) on chromosome 5. No "high quality" SNPs were detected in thirty-three 100 kb windows, including a 400-kb contiguous region between 37.0 and 37.4 Mb on chromosome 1 (Fig. 2). This sequence information identified many "high quality" SNP markers covering the entire rice genome that could be used for further analysis.

SNP genotyping with the GoldenGate assay

We selected the Illumina BeadXpress system and used the GoldenGate 384 SNP assay for genotyping. We picked 384 SNPs that were evenly distributed across the genome for the assay (Fig. S2a), of which 355 had design ability rank scores of 0.8 or higher while the remaining 29 SNPs had scores between 0.6 and 0.8 (the scores range from 0 to 1 where the higher score indicates higher synthesis success rate).

SNP genotyping of ILs was carried out using the GoldenGate assay and Vericode technology, which can genotype 96 samples at 384 loci in a single plate. All SNP genotyping data (143 ILs and controls at 384 SNP loci) were generated using the BeadXpress system and then processed from images to genotypes with the Illumina GenomeStudio Analysis Module software. Quality control was, therefore, crucial at this step. Of the 384 SNP loci, 371 (96.6 %) could be successfully genotyped with a GenTrain cut-off score of $\geq$0.25, which was recommended for obtaining a reliable genotype call. Additionally, the genotyping results were manually adjusted for higher dependability. That is, rice genotypes at each SNP locus were classified into three clusters: (a) homozygous AA (Lemont), (b) homozygous BB (Teqing), and (c) heterozygous AB (Fig. S3). We then adjusted the genotyping clusters manually based on the control samples (genotyping of parents and F_1 of Lemont $\times$ Teqing) for each SNP locus. SNPs with call rates lower than 90 % were removed. Finally, 335 SNPs that with an estimated error rate lower than 1 %

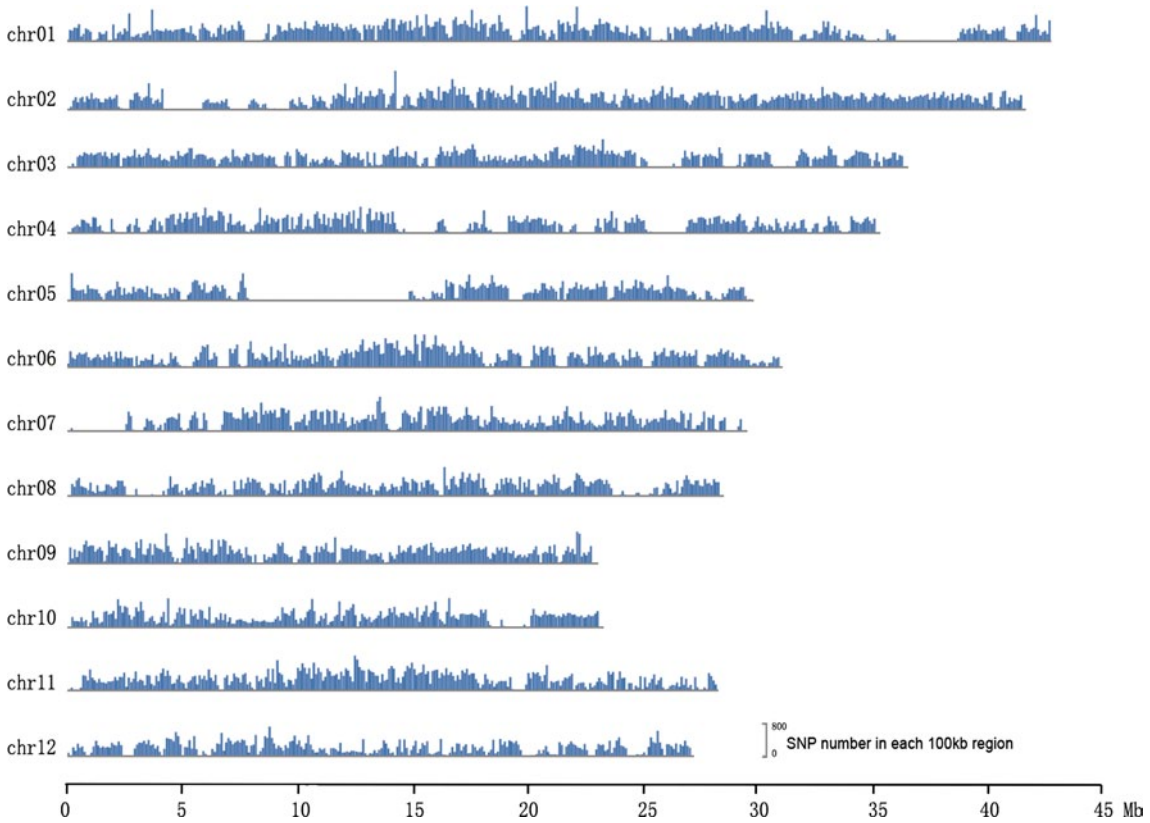

Fig. 2 Distribution of SNPs beween Teqing and Lemont across the 12 rice chromosomes. Window size was 100 kb

were retained for data analysis. Of the 335 SNPs, A/G (110) and C/T (108) accounted for 65 %, whereas A/C (37), A/T (31), C/G (20), and G/T (28) accounted for the remaining 35 %. According to the annotation of the MSU build 6.1 of the rice genome, 264 (62.9 %) of these SNPs are located in exons, of which 257 are non-synonymous mutations resulting in amino acid changes. The remaining SNPs include 10 (3.0 %) in introns, 6 (1.8 %) in UTRs, and 55 (16.4 %) in intergenic regions. Thus, following our stepwise strategy, most SNPs in the GoldenGate assay could be used for accurate and informative genotyping in a biparental population.

We also compared the SNP markers with 160 SSR markers previously generated for the same population (Zheng et al. 2007). The sequences of the SSR markers were obtained from RiceData (http://www.ricedata.cn/) and Gramene (http://www.gramene.org/). To obtain the physical locations of the SSR markers, the sequences of the SSR markers were aligned to the reference Nipponbare genome using BLAST. Overall, 151 SSR markers were successfully located, and the physical locations of the other nine SSR markers were calculated based on the physical and genetic locations of the flanking markers. The distributions of SNP and SSR markers across the chromosomes are shown in Fig. S2b, which shows that the SNP markers designed in this study had higher resolution and were more evenly distributed than the SSR markers.

Genotyping analysis of the ILs

The results of genotyping the 143 ILs are shown in Fig. S4a (raw data are presented in Table S3). The proportions of the entire genomes of these lines consisting of introgression fragments from Lemont ranged from 0.5 to 26.2 %. At each SNP locus, from 2 to 88 ILs contained introgression fragments, with an average of 19.5 (Fig. S4b). We compared genotypes obtained using SNPs with those obtained using SSRs (Zheng et al. 2007), and the results showed a good match. On average, 91.2 % of the genotypes detected by SSRs were consistent with those detected by SNPs. Meanwhile, SNP genotyping provided some precise results that were not detected using SSRs (Fig. S5). Locus Os06-26104673-LT, was remarkably enriched with 88 ILs that had introgressed Lemont fragments, indicating a very high segregation distortion at this position.

To better understand the mechanism of segregation distortion in the ILs, a Chi square test was performed on the SNP loci to determine whether the population showed normal Mendelian segregation (Fig. 3). Most of the *P* values

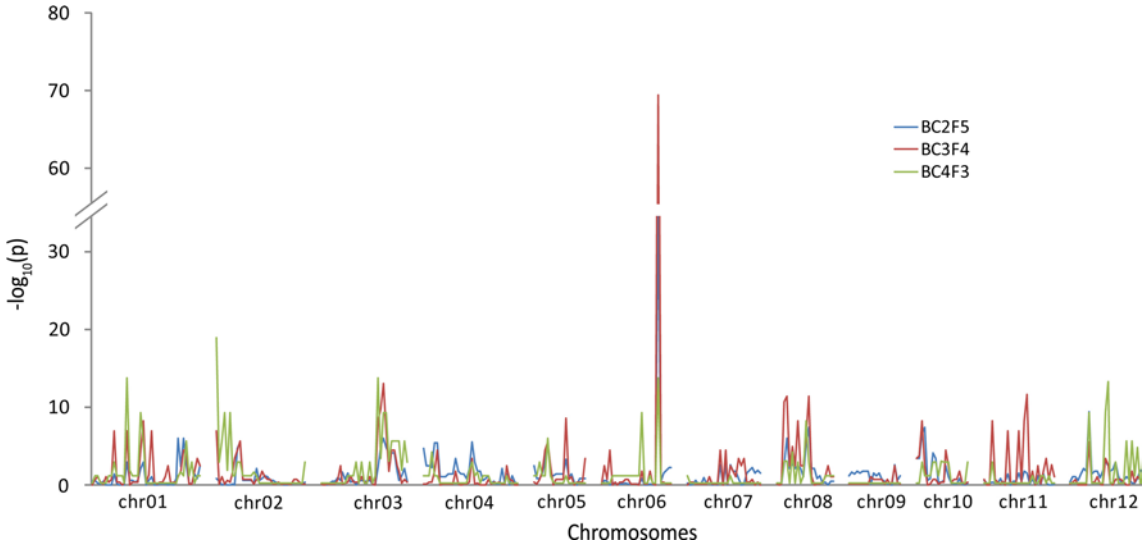

Fig. 3 Segregation distortion loci detected in three different generations of introgression lines derived from Teqing and Lemont. *Y* axis indicates –log10 of *P* values for the Chi square test

for segregation distortion were less than 10^{-5}, except for a peak at Os06-26104673-LT for all BC_2F_5, BC_3F_4 and BC_4F_3 sub-populations. Segregation distortions varied according to populations and chromosomes, while *P* values in BC_4F_3 population were lower than in the other two sub-populations. In these three sub-populations, *P* value peaks were observed in regions located on chromosomes 3, 6, and 8, coinciding with the *ga-14*, *S8* and *qSS-8* loci previously reported to be associated with pollen or spikelet sterility (Lu et al. 2000; Chen et al. 2006; Wan et al. 1993; Wang et al. 2005).

Linkage maps and QTL affecting agronomic traits

After removing 14 segregation distortion markers, 321 SNP markers were kept for linkage and QTL analysis. Then a genetic linkage map was constructed when the total size is 1,791 cM with an average distance of 5.6 cM between adjacent markers. The average size of the 12 chromosomes is 149.3 cM, ranging from 81.9 cM for chromosome 6 to 247.5 cM for chromosome 1 (Fig. S6). This map is larger than the genetic map constructed using 160 SSR markers by Xu et al. (2005), which had a total genomic size of 1,677 cM with an average distance of 10.5 cM between adjacent markers.

Quantitative trait loci affecting grain shape and milling quality traits detected by the 321 SNPs and 160 SSRs were compared. Thirty QTL were detected by the 321 SNPs: one for grain length (GL), ten for grain width (GW), five for grain thickness (GT), three for ratios of GL/GT (RGLT) one for ratios of GL/GW (RGLW), four for brown rice percentage (BR), four for milled rice percentage (MR), and two for head rice percentage (HR) (Table 1). At all loci except for *QGl4*, *QGt1*, *QGt5*, *QGt11*, *QRglt3a*, *QRglt4*, *QRglt5,* and *QRglw3* the Lemont alleles were associated with lower trait values. For grain shape traits, only eight QTL for were also identified by the 160 SSRs, including one for GW, three for GT, and four for RGLT; the Lemont alleles at *QGw7b*, *QGt5*, *QGt6*, and *QGt7b* reduced the trait values, whereas those at the other four QTL increased RGLT (Table 1). QTL affecting GT and GW were detected at five genomic regions, and the Lemont alleles in the regions (Os01-20463894-LT–Os01-21379853-LT on chromosome 1 and Os11-15445681-LT–Os11-17512870-LT on chromosome 11) had opposite effects on GT and GW, whereas the Lemont alleles in the other three regions (Os07-22344509-LT–Os07-23234709-LT and RM18-25652668–RM478-25949521 on chromosome 7, and Os10-12700842-LT–Os10-13707801-LT on chromosome 10) had the same effects on GT and GW. For three milling quality traits, no QTL were identified by the 160 SSRs when 10 QTL were detected by 321 SNPs. This may be due to the density of different markers (Fig. S7) and relatively low heritabilities (72.1 % for BR, 82.3 % for MR, and 70.2 % for HR, when 94.4 % for GL, 92.1 % for GW, and 93.7 % for GT). The Lemont alleles in these regions had opposite effects on these milling quality traits.

GW5 is an important cloned gene affecting grain shape. QTL loci adjacent to the gene were detected for GT and RGLT traits by QTL mapping (Fig. 4). For GT, a significant QTL *QGt5* was detected by both SNP and SSR markers, which could be accurately delimited to about 0.8 Mb using SNP markers but only to a 4.4-Mb interval using the

Table 1 QTLs for grain shape and milling quality traits detected by SNP and SSR markers in the 143 introgression lines derived from the cross between Teqing and Lemont

Marker	QTL	Trait	Chr.	Pos.	Left marker	Right marker	LOD	PVE (%)[a]	Add[b]
321 SNPs									
	QGl4	GL	4	103	Os04-28545803-LT	Os04-29499584-LT	12.3	13.8	0.8
	QGw1	GW	1	103	Os01-20463894-LT	Os01-21379853-LT	5.9	4.3	−0.1
	QGw3	GW	3	57	Os03-09537796-LT	Os03-10696698-LT	10.2	11.3	−0.2
	QGw7a	GW	7	122	Os07-23234709-LT	Os07-24210590-LT	10.8	10.1	−0.2
	QGw8	GW	8	57	Os08-17449468-LT	Os08-18396373-LT	7.0	6.0	−0.1
	QGw9a	GW	9	19	Os09-07237057-LT	Os09-08118408-LT	5.4	8.1	−0.2
	QGw9b	GW	9	85	Os09-19102345-LT	Os09-20111040-LT	8.7	10.3	−0.2
	QGw10	GW	10	66	Os10-12700842-LT	Os10-13707801-LT	7.5	8.5	−0.2
	QGw11a	GW	11	54	Os11-08413857-LT	Os11-09429954-LT	6	4.7	−0.1
	QGw11b	GW	11	69	Os11-12596218-LT	Os11-13616357-LT	5.8	4.4	−0.1
	QGw11c	GW	11	84	Os11-15445681-LT	Os11-17512870-LT	7.4	5.6	−0.1
	QGt1	GT	1	103	Os01-20463894-LT	Os01-21379853-LT	4.6	3.5	0.1
	QGt5	GT	5	31	Os05-04578273-LT	Os05-05389716-LT	5.5	4.4	0.1
	QGt7a	GT	7	120	Os07-22344509-LT	Os07-23234709-LT	8	5.8	−0.1
	QGt10	GT	10	68	Os10-12700842-LT	Os10-13707801-LT	8.8	7	−0.1
	QGt11	GT	11	85	Os11-15445681-LT	Os11-17512870-LT	5.6	4.2	0.1
	QRglt3a	RGLT	3	116	Os03-16699322-LT	Os03-17511092-LT	7.4	6.9	0.3
	QRglt4	RGLT	4	103	Os04-28545803-LT	Os04-29499584-LT	7.2	11.0	0.4
	QRglt5	RGLT	5	33	Os05-04578273-LT	Os05-05389716-LT	10.1	8.3	0.2
	QRglw3	RGLW	3	58	Os03-09537796-LT	Os03-10696698-LT	9.4	2.8	0.4
	QBr3	BR	3	71	Os03-09537796-LT	Os03-10696698-LT	5.3	9.4	−0.1
	QBr4a	BR	4	57	Os04-18321788-LT	Os04-19483066-LT	5.4	4.5	−0.1
	QBr4b	BR	4	266	Os04-31485158-LT	Os04-32513565-LT	6.1	4.6	−0.1
	QBr8	BR	8	23	Os08-04488034-LT	Os08-07745351-LT	5.3	6.2	−0.1
	QMr2	MR	2	117	Os02-17799722-LT	Os02-18559893-LT	6.6	5.6	−0.1
	QMr4a	MR	4	20	Os04-04560663-LT	Os04-05587643-LT	6.0	4.5	−0.1
	QMr4b	MR	4	57	Os04-18321788-LT	Os04-19483066-LT	6.8	5.5	−0.1
	QMr4c	MR	4	265	Os04-30541070-LT	Os04-31485158-LT	6.1	4.9	−0.1
	QHr2	HR	2	115	Os02-17799722-LT	Os02-18559893-LT	7.9	5.7	−0.1
	QHr4	HR	4	122	Os04-28545803-LT	Os04-29499584-LT	6.9	11.8	−0.1
160 SSRs									
	QGw7b	GW	7	89	RM18-2565266	RM478-259495	7.7	9.3	−0.3
	QGt5	GT	5	60	RM574-3451030	RM289-7807830	4.2	7.7	−0.1
	QGt6	GT	6	99	RM30-27252300	RM439-29625189	4.3	9.2	−0.1
	QGt7b	GT	7	90	RM18-25652668	RM478-25949521	6.8	7.7	−0.1
	QRglt2a	RGLT	2	41	RM53-4414267	RM324-11389874	6.8	12.1	0.4
	QRglt2b	RGLT	2	175	RM208-35135799	RM482-35272560	5.7	10.8	0.4
	QRglt3b	RGLT	3	99	RM16-23126231	RM168-28091639	6.7	11.9	0.4
	QRglt12	RGLT	12	95	RM463-22092389	RM270-24969001	6.1	12.1	0.4

[a] Phenotypic variation explained

[b] Additive effect

SSR markers (Fig. S8). For RGLT, the QTL (*QRglt5*) adjacent to the *GW5* locus was detected only by SNP and not by SSR markers.

In conclusion, the SNP markers generated by our strategy were highly accurate, evenly distributed, and of high quality for gene mapping and QTL identification.

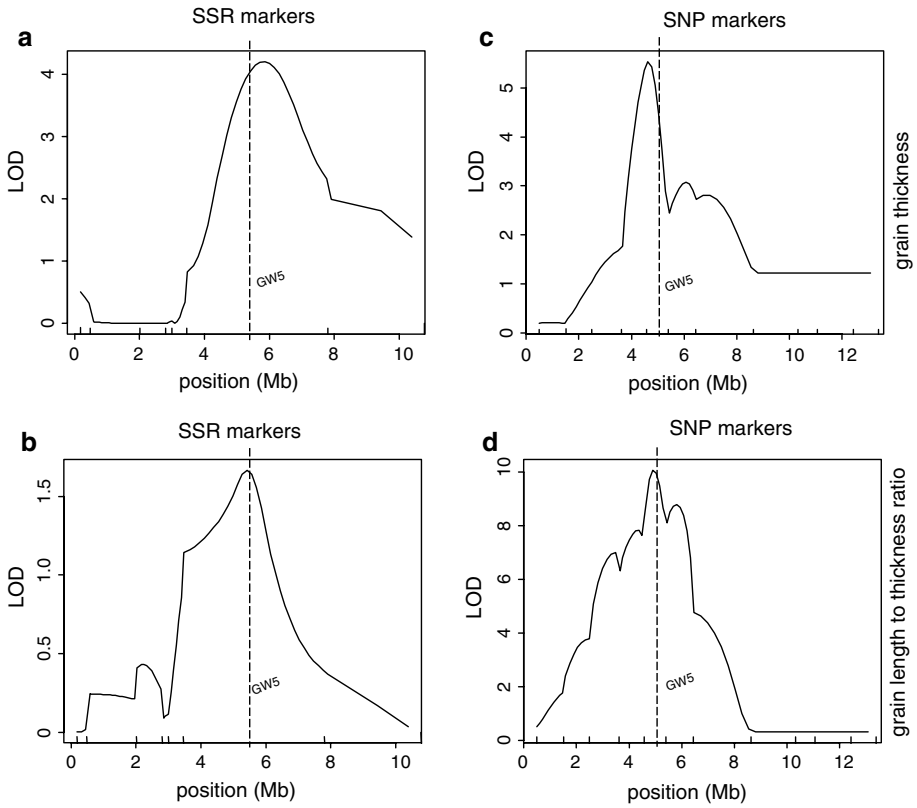

Fig. 4 Comparison of QTL mapping using SNP and SSR markers. **a** LOD curves of QTL mapping for grain thickness on chromosome 5 using SSR. *Short bars* on *X* axis indicate the position of SSR markers. **b** LOD curves of QTL mapping for grain length to thickness ratio on chromosome 5 using 160 SSR markers. **c** LOD curves of QTL mapping for grain thickness on chromosome 5 using all 321 SNP markers. *Short bars* on *X* axis indicate the position of SNP markers. **d** LOD curves of QTL mapping for grain length to thickness ratio on chromosome 5 using all 321 SNP markers

Discussion

SNP selection from parental resequencing

Traditionally, to genotype a biparental population, marker pools including SSRs and other markers were used to identify polymorphisms between the parents and only the polymorphic markers were kept (McCough and Doerge 1995). Every marker needed specific pairs of primers, and most laboratories could only handle stocks of primers for detecting a few hundred markers. Accordingly, there frequently were not enough informative markers after the preliminary screening, especially if the distribution of these markers was uneven (Fig.S2b; Zheng et al. 2007).

Compared with SSRs, SNP markers are relatively abundant across the rice genome and can be detected by high-throughput techniques. Recently, next-generation sequencing has provided a powerful method to detect SNPs. Smith et al. (2008) reported that sequencing at a depth of 10–15× using the Solexa Analyzer permitted accurate and cost-effective detection of polymorphisms. Yamamoto et al. (2010) discovered 67,051 SNPs between two closely related rice cultivars (Koshihikari and Nipponbare), with an average density of 5.7 kb/SNP after sequencing to a depth of 15× coverage with the Illumina system. In our study, to identify the informative SNPs, we resequenced the parents Teqing and Lemont to a depth of about 14× and detected 728,362 "high quality" SNPs between them. This was enough to select markers not only for the preliminary QTL analysis, but also for the subsequent fine mapping (Fig. 2, Fig. S2a). As shown in Fig. S2b, because of the huge pools of SNPs, it was easy to select evenly distributed SNPs for QTL analysis, which showed great advantages over SSR markers. PCR validation confirmed that SNPs with more than 9× coverage during resequencing were 100 % correct (Table S2).

Advantages of genotyping biparental populations with GoldenGate assays based on parental resequencing

The BeadXpress platform was used to genotype the IL population derived from Teqing and Lemont using the selected SNPs and the GoldenGate assay. This platform has been demonstrated to be efficient for medium-throughput SNP genotyping (Chen et al. 2011). In this study, the genotyping experiment was completed within 3 days by one person after obtaining the genomic DNAs. In contrast, genotyping these lines with 160 SSR markers took about 3 months of work by two highly experienced scientists who assayed on average twenty 96-well gels per day without taking account of possible repeats or trouble-shooting (Zheng et al. 2007). Thus, the strategy used in this study dramatically accelerated the genotyping process and provided more information with much lower cost per marker.

Recently, next-generation sequencing (NGS) has been used to genotype populations, which has increased the resolution of QTL mapping (Huang et al. 2009; Wang et al. 2011; Xie et al. 2010; Yu et al. 2011). However, to minimize the relatively higher costs, the strategies for genotyping populations by resequencing commonly choose a relatively low coverage (often less than 2×). As indicated by our screening of SNPs from the sequencing data of Teqing and Lemont, the repeatability of SNPs decays significantly with decreasing depth of resequencing (Table S1). Furthermore, it has two other disadvantages and is difficult to apply widely. First, library construction and resequencing for several hundred plants is time-consuming and expensive. Second, the huge amount of data generated by resequencing requires complicated bioinformatics analysis. Comparing with NGS, the GoldenGate assay is much easier, since it only needs genomic DNA, avoiding the complexity of library construction. Moreover, the generated data are direct SNP alleles which do not need complicated analysis. Most importantly, the SNPs that perform well in GoldenGate assays can be directly used as molecular markers for genetic research and breeding.

Using GoldenGate assays, many SNP arrays have been developed in plants. Hyten et al. (2008) developed a 384 SNP GoldenGate assay in soybean based on published soybean SNP data. When this set of SNPs was used to genotype three recombinant inbred line populations, about 40 % of the SNPs showed polymorphisms in each population. Thomson et al. (2012) developed several versatile SNP arrays in rice, which are very useful for wide applications. However, the efficiency of these versatile SNP arrays was relatively low when applied to specific populations. Our method took a different approach in that it was designed for a specific population. Since it was based on resequencing the whole genomes of the parents, this guaranteed that all the selected SNPs were informative for genotyping the population derived from the biparental cross. It thus traded efficiency for versatility, but this will be a useful tradeoff in analyzing any RIL population, especially as the costs of sequencing continue to decrease.

Comparison of genotyping and QTL analysis based on SNPs and SSRs

We compared genotypes obtained using SNP markers with those previously obtained using SSR markers (Zheng et al. 2007). On average, 91.2 % of genotypes detected by SSRs were consistent with those detected by SNPs. However, genotyping by SNP provided more precise results, such as the successful detection of small introgressions that were not recognized by SSRs (Fig. S5). The genotypes obtained using SNP markers had another advantage in linkage map construction. The known physical positions of SNP markers generated by resequencing could detect false double crossovers between adjacent markers, which would otherwise be incorrectly incorporated in genetic maps based on markers such as SSRs (Yu et al. 2011).

Pan et al. (2012) reported high-density markers could improve QTL mapping power and resolution. By progressing from QTL mapping with 220 SSRs to QTL mapping with 1536 SNPs, they reduced the size of the cloned QTL *Pa19* locus (*FatB* gene controlling palmitic acid) from 3.5 to 1.72 Mb, and the cloned QTL *Oil6* locus (*DGAT1-2* gene controlling oil concentration) from 10.8 to 1.62 Mb. In our study, a cloned QTL *QGt5* was detected by both SNP and SSR markers and accurately delimited to ~0.8 Mb using SNP markers, but only to a 4.4-Mb interval using the SSR markers (Fig. S8). Our results clearly show that high-density SNP markers improved QTL mapping resolution.

Furthermore, to compare the QTL analyses based on equal numbers of SNP and SSR markers, we selected 160 SNPs that were evenly distributed for analysis, as the previous QTL mapping study used 160 SSR markers (Zheng et al. 2007). The accuracies for QTL mapping were similar between these two marker sets (160 SNPs and 160 SSRs, Fig. S9), but LOD values detected by the SNPs were much larger than those detected by the SSRs (Fig. S10). As well, the SNPs were selected from a huge SNP pool here to ensure their better even distribution than the SSRs', which would minimize the possible false negatives for the QTL analysis caused by marker gaps.

In summary, following our strategy, all SNPs were selected from a huge SNP pool; thus the SNPs were evenly distributed along chromosomes and provide much higher resolution for QTL mapping. Each SNPs was from biparental sequencing and thus would be effective for genotyping the population. So we suggest it is the best strategy currently available to genotype a biparental population and map its important agronomic traits.

Acknowledgments This work was supported by grants from the National Program on Key Basic Research Project of China (973 Program: 2011CB100101), Ministry of Agriculture of China (948 Program: 2011-G2B), National High Technology Research and Development Program of China (863 Program: 2012AA10A304), National Natural Science Foundation of China (U1031001) and the Program of International Science and Technology Cooperation (2012DFB32280).

Conflict of interest The authors declare that they have no conflict of interest.

References

Ahn SN, Suh JP, Oh CS, Lee SJ, Suh HS (2002) Development of introgression lines of weedy rice in the background of Tongil-type rice. Rice Genet Newsl 19:14

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517

Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res 21:610–617

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43:956–963

Chen H, He H, Zou Y, Chen W, Yu R, Liu X, Yang Y, Gao YM, Xu JL, Fan LM, Li Y, Li ZK, Deng XW (2011) Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). Theor Appl Genet 123:869–879

Chen J, Jiang L, Wang C, Ikehashi H, Zhai HQ, Wan JM (2006) Mapping of loci for pollen sterility of indica-japonica hybrids in rice (*Oryza sativa* L.). Acta Agron Sin 32:515–521

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genom 10:582

Doi K, Iwata N, Yoshimura A (1997) The construction of chromosome introgression lines of African rice (*Oryza glaberrima Steud.*) in the background of japonica (*O. sativa* L.). Rice Genet Newsl 14:39–41

Doi K, Sobrizal K, Ikeda K, Sanchez PL, Kurakazu T, Nagai Y, Yoshimura A (2003) Developing and evaluating rice chromosome segment substitution lines. Proc Int Rice Congress, pp 275–287

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003) Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69–78

Ghesquiere A, Sequier J, Second G, Lorieux M (1997) First steps toward a rational use of African rice, Oryza glaberrima, in rice breeding: a contig line concept. Euphytica 96:31–39

Hamilton JP, Buell CR (2012) Advances in plant genome sequencing. Plant J 70:177–190

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

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 (2012a) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012b) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44:32–39

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li H, Hearne S, Banziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity (Edinb) 105:257–267

Lu CG, Zou JS, Ikahashi H (2000) Gamete abortion locus detected by segregation distortion of isozyme locus Pgi1 in *indica-japonica* hybrids of rice (*Oryza sativa* L.). Breed Sci 50:235–239

Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong T, Crouch JH, Xu Y (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120:93–115

Li ZK, Fu BY, Gao YM, Xu JL, Ali J, Lafitte HR, Jiang YZ, Rey JD, Vijayakumar CH, Maghirang R, Zheng TQ, Zhu LH (2005) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Mol Biol 59:33–52

Maughan PJ, Yourstone SM, Jellen EN, Udall JA (2009) SNP discovery via genomic reduction, barcoding and 454-pyrosequencing in Aramanth. Plant Genome 2:260–270

Maughan PJ, Yourstone SM, Byers RL, Smith SM, Udall JA (2010) Single-nucleotide polymorphism genotyping in mapping populations via genomic reduction and next-generation sequencing: proof of concept. Plant Genome 3:166–178

McCough SR, Doerge RW (1995) QTL mapping in rice. Trends Genet 11(12):482–487

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of the rice chromosomes. Theor Appl Genet 76:815–829

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res 35:D883–D887

Pan Q, Ali F, Yang X, Li J, Yan J (2012) Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS One 7:e52777

Parida SK, Mukerji M, Singh AK, Singh NK, Mohapatra T (2012) SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. BMC Genomics 13:426

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel

two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253

Saintenac C, Jiang D, Wang S, Akhunov E (2013) Sequence-based mapping of the polyploid wheat genome. G3 (Bethesda) 3:1105–1114

Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, Shen L, Donahue WF, Tusneem N, Stromberg MP, Stewart DA, Zhang L, Ranade SS, Warner JB, Lee CC, Coleman BE, Zhang Z, McLaughlin SF, Malek JA, Sorenson JM, Blanchard AP, Chapman J, Hillman D, Chen F, Rokhsar DS, McKernan KJ, Jeffries TW, Marth GT, Richardson PM (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res 18:1638–1642

Schouten HJ, van de Weg WE, Carling J, Khan SA, McKay SJ, van Kaauwen MP, Wittenberg AH, Koehorst-van Putten HJ, Noordijk Y, Gao Z, Rees DJ, Van Dyk MM, Jaccoud D, Considine MJ, Kilian A (2012) Diversity arrays technology (DArT) markers in apple for genetic linkage maps. Mol Breed 29:645–660

Sobrizal Ikeda K, Sanchez PL, Doi K, Angles ER, Khush GS, Yoshimura A (1999) Development of Oryza glumaepatula introgression line in rice (*O. sativa* L.). Rice Genet Newsl 16:107–108

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung CW, Reynolds A, Scheffler B, Eizenga G, McClung A et al (2012) Highthroughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Mol Breed 29:875–886

Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Sun CQ (2006) Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor Appl Genet 112:570–580

van Orsouw NJ, Hogers RC, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H, van Eijk MJ (2007) Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS One 2:e1172

Wan J, Yanagihara S, Kato H, Ikehashi H (1993) Multiple alleles at a new locus causing hybrid sterility between a Korean indica variety and a javanica variety in rice (*Oryza sativa* L.). Jpn J Breed 43:507–516

Wang C, Zhu C, Zhai H, Wan J (2005) Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). Genet Res 86:97–106

Wang L, Wang A, Huang X, Zhao Q, Dong G, Qian Q, Sang T, Han B (2011) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theor Appl Genet 122:327–340

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210

Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA 107:10578–10583

Xu JL, Lafitte HR, Gao YM, Fu BY, Torres R, Li ZK (2005) QTLs for drought escape and tolerance identified in a set of random introgression lines of rice. Theor Appl Genet 111:1642–1650

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, Shibaya T, Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267

Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS One 6:e17595

Zhao K, Wright M, Kimball J, Eizenga G, McClung A, Kovach M, Tyagi W, Ali ML, Tung CW, Reynolds A, Bustamante CD, McCouch SR (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS One 5:e10780

Zheng TQ, Xu JL, Li ZK, Zhai HQ, Wan JM (2007) Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (*Oryza sativa* L.). Plant Breeding 126:158–163

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989