Molecular Plant • Volume 5 • Number 1 • Pages 218–230 • January 2012 RESEARCH ARTICLE

Dynamic Intra-*Japonica* Subspecies Variation and Resource Application

Haiyan Lin[a], Peng Xia[a], Rod A. Wing[b], Qifa Zhang[a] and **Meizhong Luo[a,1]**

a National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China
b Arizona Genomics Institute, University of Arizona, Tucson, AZ 85721, USA

ABSTRACT We constructed a physical map of *O. sativa* ssp. *japonica* cv. ZH11 and compared it and its random sample sequences with the Nipponbare RefSeq derived from the same subspecies. This comparison showed that the two *japonica* genomes were highly syntenic but revealed substantial differences in terms of structural variations, rates of substitutions and indels, and transposable element content. For example, contractions/expansions as large as 450 kb and repeat sequences that were present in high copy numbers only in ZH11 were detected. In tri-alignment regions using the *indica* variety 93–11 sequence as an outgroup, we found that: (1) the substitution rates of the two *japonica–indica* inter-subspecies comparison combinations were close but almost a magnitude higher than the substitution rate between the *japonica* rice varieties ZH11 and Nipponbare; (2) of the substitutions found between ZH11 and Nipponbare, 47.2% occurred in ZH11 and 52.6% in Nipponbare; (3) of the indels found between ZH11 and Nipponbare, the indels that occurred in ZH11 were 15.8 times of those in Nipponbare. Of the indels that occurred in ZH11, 75.67% were insertions and 24.33% deletions. Of the indels that occurred in Nipponbare, 48.23% were insertions and 51.77% were deletions. The ZH11 comparative map covered four Nipponbare physical gaps, detected assembly errors in the Nipponbare sequence, and was integrated with the FSTs of a large ZH11 T-DNA insertion mutant library. ZH11 BAC clones can be browsed, searched, and obtained at our website, http://GResource.hzau.edu.cn.

Key words: Physical map; comparative genomics; *Japonica*; *Oryza*; rice; BAC.

INTRODUCTION

Comparisons are ubiquitous and serve as basic tools in genomics, genetics, and evolutionary studies. Large genome size differences, structural changes, and small SNPs are all determined through comparisons. Comparisons among different phylogenetic distances meet different expectations (Zhang and Wessler, 2004; Tang et al., 2010). Those between far phylogenetic distances are important for finding conservations, such as syntenies and conserved non-coding sequences (CNS) (Gale and Devos, 1998), while those between close phylogenetic distances, with extreme examples represented by those between recombinant inbred lines (RILs) or between wild-types and mutants, are valuable for finding variations that could provide clues to directly link DNA structures with functions.

Rice is a worldwide staple food crop. It belongs to the genus *Oryza*, which contains 23 species with 10 genome types (Ge et al., 1999; Ammiraju et al., 2006, 2008). The Asian cultivated rice species *O. sativa* contains two subspecies: *japonica* and *indica*. The genome sequences of the *japonica* variety Nipponbare and the *indica* variety 93–11 have been published and extensive inter-subspecies comparisons have since been made (Han and Xue, 2003; Ma and Bennetzen, 2004; IRGSP, 2005; Yu et al., 2005). A project called OMAP (Oryza Map Alignment Project, www.omap.org), which aims to align the physical maps of all 10 genome types of *Oryza* from representative species, was started several years previously (Wing et al., 2005). High-quality bacterial artificial chromosome (BAC) libraries of 12 *Oryza* species representing the 10 genome types were constructed for the project (Ammiraju et al., 2006). BAC-end sequences (BESs) and fingerprints were produced from these BAC libraries, and phase I physical maps were constructed for the 12 Oryza species (Kim et al., 2008). A global comparison between Nipponbare (AA genome type) and *O. punctata* (BB genome type) was performed, and large DNA expansions, contractions, inversions, and translocations were found (Kim et al.,

[1] To whom correspondence should be addressed. E-mail mzluo@mail.hzau.edu.cn, fax +86 27 87284213.

doi: 10.1093/mp/ssr085, Advance Access publication 7 October 2011
Received 27 June 2011; accepted 12 September 2011

2007). Insights into structures and the evolution of targeted orthologous regions of the *Oryza* species have been obtained through comparisons (Ammiraju et al., 2008; Lu et al., 2009; Ammiraju et al., 2010).

Intra-subspecies comparisons have been obtained with respect to SNPs (McNally et al., 2009; Huang et al., 2010), expression (Zhang et al., 2005), and targeted regions. McNally et al. (2009) resequenced and compared 20 diverse varieties and landraces in 100 Mb of the rice genome and showed the power of the technology in revealing the breeding history and relationships among the 20 varieties by the introgression patterns of the shared SNPs. Huang et al. (2010) resequenced at low coverage and compared 517 rice landraces, and identified ~3.6 million SNPs that were used to identify the loci for 14 agronomic traits in the population of *Oryza sativa indica* subspecies through genome-wide association studies (GWAS). However, no extensive comparisons with respect to structural variation, rates of substitutions and indels, and transposable element contents have been made within either *japonica* or *indica* subspecies because each subspecies has only one whole-genome sequence available. To perform intra-subspecies comparisons, at least one other independent whole-genome sequence or suitable physical map is required. The sequence reads produced by Huang et al. (2010) with next-generation sequencing technology were very short; without a good independent frame to guide sequence assembly, they cannot be used to effectively conduct comparative studies in the hopes of detecting large structural variations. Independently constructed physical maps through BAC fingerprints and with BESs can be used for this purpose (Wing et al., 2005), but are currently lacking for cultivated rice varieties. The OMAP project did not include any resources for these intra-subspecies genome comparisons (www.omap.org). In maize, Fu and Dooner (2002) found large variations in the *bz*1 regions between two inbred lines. Little is known about the extent of variations among rice intra-subspecies varieties or the extent of mutations occurring in individual varieties. The intra-subspecies variations have been almost totally neglected such that, in many reports, '*japonica*' and '*indica*' were used to directly replace or represent variety names. The levels of intra-subspecies variations represent the basic levels of inter-subspecies and interspecies variations. Without an understanding of the levels of intra-subspecies variations, the inter-subspecies and interspecies comparisons are imperfect.

O. sativa ssp. *japonica* cv. ZH11 (Zhonghua 11) is a typical *japonica* rice variety. It is an important host for *Agrobacterium*-mediated transformation because of its high native efficiency (Wu et al., 2003; Zhang et al., 2006; Chen et al., 2008). A large number of ZH11 T-DNA insertion mutants have been constructed, and T-DNA flanking sequence tags (FSTs) have been produced, both of which are available to the public (http://rmd.ncpgr.cn). ZH11 is also used to create ^{60}Co-gamma-ray mutants (Yan et al., 2007). Guo et al. (2009) used ZH11 in a mapping population to map two grain weight genes: GW3 and GW6. We are currently using ZH11 to create EMS mutants. Although the Nipponbare RefSeq is helpful for the identification and analysis of the genes of interest from mutants or for positional cloning in subsequent steps, allele sequence information and DNA templates from the host or parent ZH11 are eventually required to confirm the function through complementation.

We constructed an independent physical map for ZH11 and performed an intra-*japonica* comparison between ZH11 and Nipponbare. By analyzing the aligned regions of the comparative map and ZH11 BESs, which are random sample sequences of the genome, we found extensive and dynamic variations in structure, rates of substitutions and indels, and transposable element contents between the two *japonica* varieties. The ZH11 physical map covered four physical gaps and detected assembly errors in the Nipponbare RefSeq. The FSTs of a large ZH11 T-DNA insertion mutant library (http://rmd.ncpgr.cn) were integrated with our comparative map. Enlarged figures and detailed information of local regions can be browsed through our website (http://GResource.hzau.edu.cn). ZH11 BAC clones for FSTs and molecular markers can be found through our search tool by using the homologous Nipponbare sequences as bridges and can be obtained through our website.

RESULTS

Generation of the ZH11 BAC Library, Fingerprints, and BESs

A BAC library of 36 864 clones, with an average insert size of 124 kb, was constructed for the *japonica* rice ZH11 and deposited in 96 × 384-well plates. The 18 432 BAC clones from the first 48 × 384-well plates were fingerprinted and end-sequenced. The average size per fingerprint band was 1 120 bp. In total, 35 919 BESs (Q > 16, length > 100 bp) were obtained, with a cumulative length of 24 072 531 bp. The statistics for the ZH11 BAC library, fingerprints, and BESs are listed in Supplemental Table 1.

Supposing ZH11 has a genome size of 389 Mb, which is similar to that of Nipponbare, and removing the clones and BESs of the organelles, the BAC library has 11× coverage of the genome. The successfully fingerprinted BAC clones have ~5.6× coverage of the genome. The BESs cover 5.8% of the genome. On average, every 11.5 kb of the genome has a BES.

Assembly of the ZH11 BAC Physical Map

Fingerprints were assembled into contigs with FPC (FingerPrinted Contigs, Soderlund et al., 1997) using standard protocols; 93.4% (16 365 clones) of the clones fell into 1093 contigs, and the remaining 1162 clones were singletons. The phase I FPC assembly covered approximately 82% of the ZH11 genome (318 Mb; 285 264 CB units), with an average contig size of approximately 292 kb (261 CB units). The phase I FPC assembly was manually edited as described by Kim et al. (2007). The resulting phase II FPC assembly consisted of 648 contigs (16 042 clones; ~291 Mb; 260 030 CB units) and 1 485 singletons. The longest contig (contig 38) in this assembly was about 3 Mb.

Comparative Physical Mapping of Intra-*Japonica* Subspecies

An intra-*japonica* subspecies comparative physical map was constructed between the ZH11 phase II physical map and the Nipponbare RefSeq (Figure 1). Blat software (Kent, 2002) matched 21 320 masked-BESs (accounting for 59.36% of the total BESs) from the ZH11 phase II physical map to the Nipponbare RefSeq; according to the match data, SyMAP (Soderlund et al., 2006) anchored 575 ZH11 contigs, with a total length of 249 048 CB (~279 Mb) containing 15 768 clones, to the Nipponbare RefSeq. Of the 15 768 clones in the anchored contigs, 12 093 clones (76.69%) were directly anchored by BESs: 5811 by double ends (paired-ends) and 6282 by single ends (Table 1; note that the repetitive BES sequences were masked before mapping). The alignment formed 63 synteny blocks and displayed an overall collinearity (Supplemental Figure 1). Seventy-three small contigs accounting for ~12 Mb could not be aligned. These unanchored contigs may represent unique regions of the ZH11 genome relative to the Nipponbare RefSeq.

The BESs on the anchored contigs are valuable tools for detecting structural changes between the genomes compared. Because the orientation and relative location of every BES are fixed on the contigs, single-end BESs are as valuable as paired-end ones. Expansions and contractions between the two *japonica* genomes can be reflected from the relative distances of the BESs on the anchored clones. Inside the contigs of the comparative map, expansions and contractions can be visualized by non-parallel alignments (Figure 1; enlarged figures and detailed information for the local regions can be browsed at our website, http://GResource.hzau.edu.cn). An approximately 450-kb contraction of ZH11 relative to Nipponbare (ZH11 contig 1750 compared to the 16.59–17.29-Mb region of Nipponbare RefSeq Chromosome 6; Supplemental Figure 2) and many smaller expansions/contractions were found on the comparative map. Figure 2A shows an approximately 140-kb contraction of ZH11 relative to Nipponbare as an example. Similarly, translocations and inversions between the two japonica genomes and assembly errors of either the Nipponbare RefSeq or the contigs could be detected by the locations and orientations of the ZH11 BESs relative to the Nipponbare RefSeq. If a ZH11 contig was anchored to different chromosomes of the Nipponbare RefSeq by BESs, it was inferred that a translocation or an assembly error must exist. If a group of BESs was anchored to the Nipponbare RefSeq with an abnormal orientation relative to other BESs of the same contig, it was inferred that an inversion or assembly error must exist. We did not find any ZH11 contigs spanning different chromosomes of the Nipponbare RefSeq. Of the 575 anchored ZH11 contigs, 558 showed normally oriented BESs, and 17 contained regions with abnormally oriented BESs. Of the 12 093 ZH11 BAC clones directly anchored to the Nipponbare RefSeq by BESs, 12 047 showed normally oriented BESs, whereas 46 showed abnormally oriented BESs at one end (Table 1). Of the 46 abnormally oriented BESs, seven were individually aligned; others were aligned in groups of two (five locations), three (two locations), seven (one location), and eight BESs (two locations), respectively (Table 2). On the comparative map, these BESs showed an outward direction at one end of their respective clones and crosswise alignments with those in the group (Figure 3A).

Yu et al. (2006) reported a local misassembly of an 80.4-kb region on the IRGSP Nipponbare chromosome 11. They compared the IRGSP Nipponbare BESs with the IRGSP Nipponbare assembly in this region and found an abnormal match for

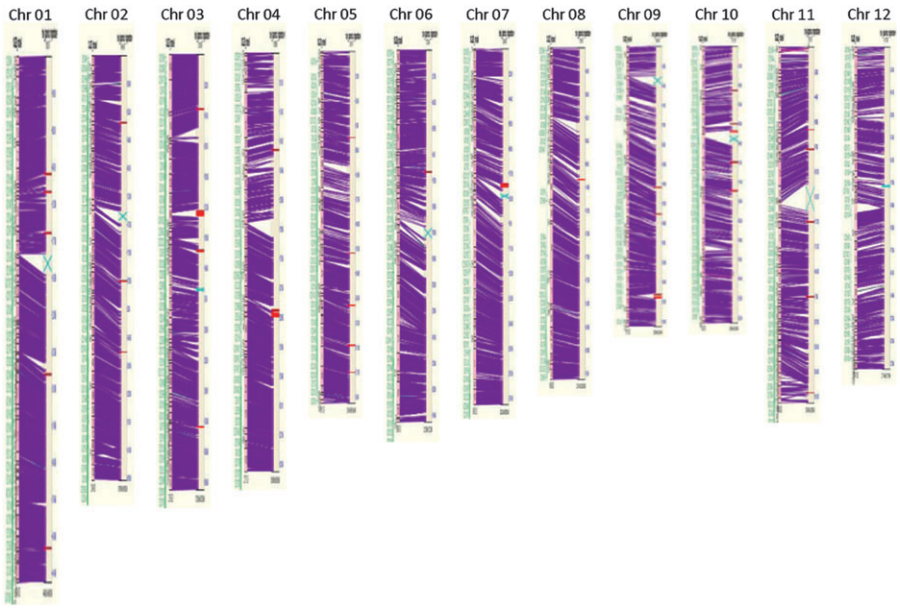

Figure 1. The SyMap Alignment of the ZH11 Phase II Physical Map to the Nipponbare RefSeq (IRGSP, Build 4.0).
The boxes on the left represent the ZH11 contigs. The bars on the right represent the Nipponbare RefSeq. The green Xs on nine chromosomes of the Nipponbare RefSeq represent the centromere locations. The centromere locations of chromosomes 4, 5, and 8 are not available. The purple lines represent the alignments between the ZH11 BESs and the Nipponbare sequence. The red boxes represent the physical gaps of the Nipponbare RefSeq. The green labels on the left represent the IDs of the ZH11 FSTs (http://rmd.ncpgr.cn). Zoomable figures and detailed information for local regions with a direct link to the RMD database (http://rmd.ncpgr.cn) for each FST ID can be browsed through our website (http://GResource.hzau.edu.cn).

the four pairs of BESs, indicating that this region was mistakenly inverted during assembly and should be flipped. Upon examining our intra-*japonica* subspecies comparative map, we found that an abnormal alignment of a group of two ZH11 BESs (Table 2, contig 146) was also located in this region. This result indicates that the abnormal alignment in this region is not due to an inversion between the two *japonica* genomes; rather, it is caused by the misassembly of the Nipponbare RefSeq. We analyzed the IRGSP Nipponbare assembly regions corresponding to the last three abnormally aligned locations in Table 2 using the method described in Yu et al. (2006) and found that several IRGSP Nipponbare BESs in each location were also abnormally matched to the IRGSP Nipponbare assembly. Figure 3 shows the BES alignments of the ZH11 contig 613 (Figure 3A) and the IRGSP Nipponbare BAC clones (Figure 3B) to the same IRGSP Nipponbare assembly region. In Figure 3B, the abnormal BESs were not cross- and collectively aligned as in Figure 3A, because only the BES information of the Nipponbare BACs was used in this analysis and no fingerprinting information was used. In this condition, the orientations of the correlated BAC clones were not fixed, and the software aligned the BESs in parallel by default. These analyses demonstrated that all the abnormal alignments analyzed were caused by the misassembly of the Nipponbare RefSeq and not caused by the inversion between the two *japonica* genomes. Because the assembly errors of the Nipponbare RefSeq were confirmed, possible assembly errors in the ZH11 contigs could be ruled out.

The *indica* rice 93–11 genome was sequenced (Yu et al., 2005) and used as an outgroup in this study. To identify the differences between intra- and inter-subspecies, an inter-subspecies

Table 1. Summary of the Comparisons of the ZH11 BESs Embedded in the Physical Map with Nipponbare and 93–11 Sequences.

ZH11	Nipponbare	93–11
BES with hits	21 320	19 975
Clones anchored by BES	12 877	12 394
Clones in anchored contigs	15 768	15 687
without BES hits	3 675	4 583
with BES hits	12 093	11 104
with paired BES hits	5 811	4 818
with single-end BES hits	6 282	6 286
with abnormally oriented BES hits	46	265
without abnormally oriented BES hits	12 047	10 839
Clones in unanchored contigs	274	355
Anchored contigs	575	564[a]
with abnormally oriented BES hits	17	99
without abnormally oriented BES hits	558	467
Unanchored contigs	73	84

a Two contigs were anchored to two locations each.

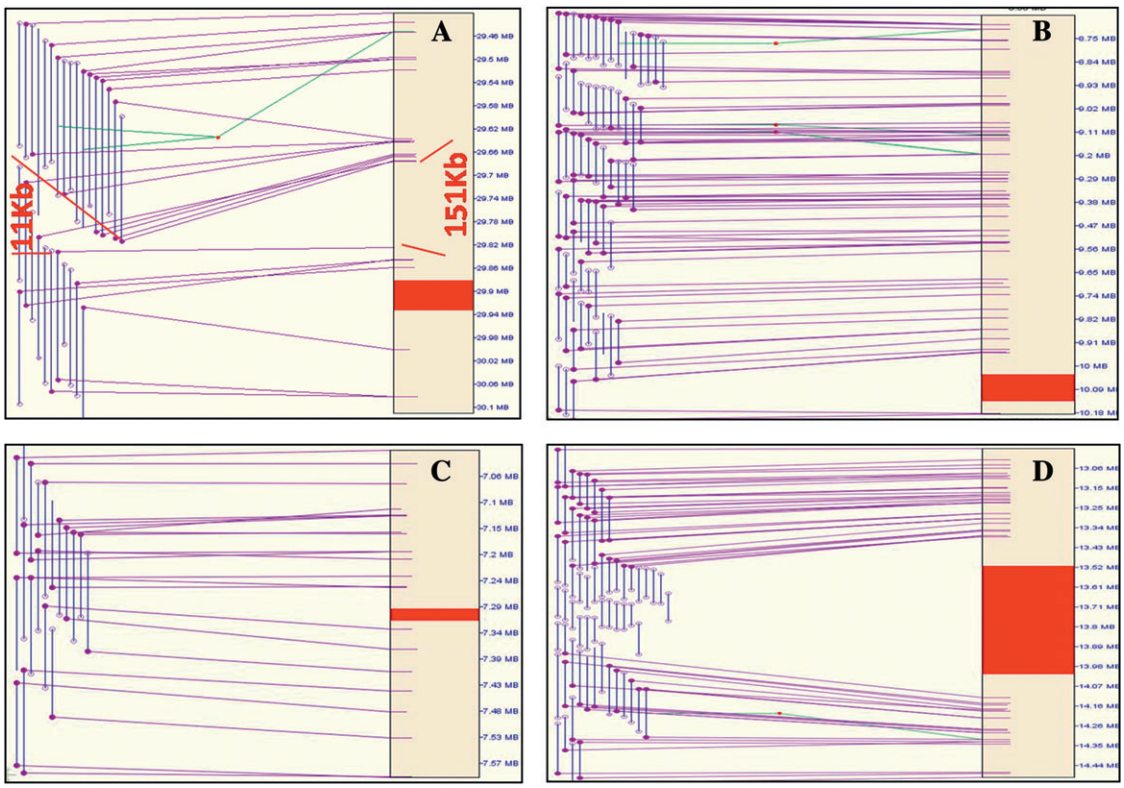

Figure 2. Alignments of ZH11 Contig 63 **(A)**, contig 194 **(B)**, contig 691 **(C)**, and contig 1490 **(D)** to the Nipponbare RefSeq.
The red boxes show the physical gaps of the Nipponbare RefSeq. A contraction of approximately 140 kb in ZH11 Contig 63 is indicated in (A). The vertical blue lines on the left side of each panel represent BAC clones in a contig.

Table 2. The Corresponding Chromosome Locations and Contigs of the Nipponbare RefSeq Hit by the Abnormally Oriented BESs in ZH11 Contigs.

ZH11 contig	Chromosome	Start bp	Stop bp	No. of abnormally oriented BESs[a]	Corresponding Nipponbare contigs
66	chr06	10 399 889	11 236 281	1	52
599	chr01	38 934 576	39 614 444	1	6
1096	chr04	15 895 672	16 870 900	1	37
938	chr06	16 368 319	17 080 056	1	52, 53
1727	chr11	28 796 354	29 388 937	1	86
623	chr12	20 149 929	21 237 384	1	96
733	chr12	21 455 062	21 772 079	1	96
700	chr01	3 062 808	3 737 068	2	1
94	chr04	1 098 024	2 391 778	2	31, 32, 33
840	chr04	23 943 970	25 698 055	2	40
9	chr05	25 833 795	27 230 261	2	49
146	**chr11**	**18 291 376**	**18 884 114**	**2**	**79**
206	chr04	19 797 193	20 949 495	3	37
905	chr06	18 574 089	18 715 694	3	53
578	**chr04**	**27 030 057**	**27 744 684**	**7**	**40**
613	**chr12**	**25 130 098**	**25 614 675**	**8**	**96**
312	**chr05**	**28 642 090**	**29 178 130**	**8**	**50**

a Hit to each location. The abnormally oriented BESs in the bold rows were investigated in detail (see text).

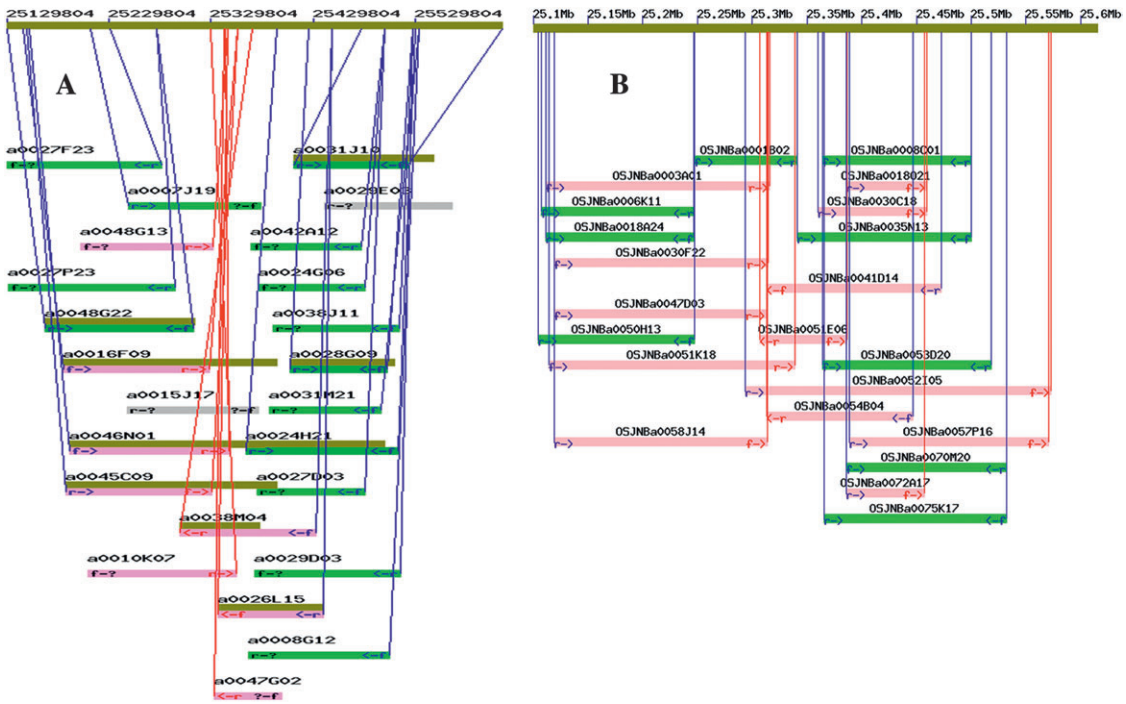

Figure 3. The Alignment View of ZH11 Contig 613 **(A)** and Nipponbare Paired-BESs from AGI **(B)** to Chr12: 25 130 098–25 614 675 bp of the Nipponbare RefSeq.

The golden bars on the top represent the Nipponbare RefSeq. Pink bars represent the BAC clones with abnormally aligned BESs. Green bars represent the BAC clones with only one end sequence available. Dark-green bars represent the BAC clones with both end sequences. Red arrows on pink bars show the abnormally aligned BESs, and the red lines indicate the abnormal alignments. In (A), the golden bars on the pink and dark-green bars show the corresponding lengths on the Nipponbare RefSeq defined by the sequences aligned to the paired-BESs of the respective BAC clones beneath the golden bars. In (B), only BES information of the Nipponbare BACs was used and no fingerprint information was used. In this condition, the orientations of the correlated BAC clones were not fixed, and the software aligned the BESs in parallel by default.

comparative physical map between the *japonica* rice ZH11 phase II physical map and the *indica* rice 93–11 genome sequence was constructed and analyzed in parallel. As shown in Table 1, fewer ZH11 contigs and clones were anchored, and more ZH11 contigs and clones were unanchored, to the 93–11 sequence than to the Nipponbare RefSeq. The results indicate a closer relationship between ZH11 and Nipponbare (intra-*japonica* subspecies) than between ZH11 and 93–11 (inter-subspecies). A total of 265 anchored BESs showed an abnormal orientation on the inter-subspecies comparative map. However, more work is needed to confirm whether they reflect real structural differences (translocation and inversion) between the inter-subspecies genomes or whether they were resulted from misassembly of either the 93–11 sequence or the ZH11 contigs.

On the intra-*japonica* comparative physical map, four ZH11 contigs were found to span four Nipponbare physical gaps (Figure 2). Contig 63 (Figure 2A), contig 194 (Figure 2B), contig 691 (Figure 2C), and contig 1490 (Figure 2D) span the fifth gap on chromosome 11, the fourth gap on chromosome 10, the first gap on chromosome 11, and the second gap on chromosome 3 of the Nipponbare assembly (IRGSP, 2005), respectively. These contigs all have a deep coverage of BAC clones in the corresponding regions with at least four clones (Figure 2B) spanning each gap. The gap-corresponding region on contig 1490 (Figure 2D) contains an approximately double coverage of BACs compared with the flanking regions, implying that a duplication may exist in this region. The region contains 55 unanchored BESs, of which 52 are repetitive sequences that can be masked by RepeatMasker. The remaining three unmasked BESs (OSJZBa0024K22.r, OSJZBa0003O13.f, and OSJZBa0047H14.r) may match the Nipponbare RefSeq in the gap.

ZH11 Highly Frequent Sequences

Out of the 35 919 BESs (24 072 531 bp), RepeatMasker identified 18 626 BESs (51.86%) containing a total of 7 685 317 bp (31.93%) of the known rice repeat sequences collected in the GIRI and TIGR repeat databases. After the removal of the known rice repeat sequences, rice organelle sequences, and SSRs by RepeatMasker, the remaining sequences were searched against themselves by blastn with the condition 'Match length $\geqslant$ 50,E-Value $\leqslant 10^{-50}$' and 975 BESs were found to contain sequence regions with more than six HSPs (High Score Pairings) among the ZH11 BESs. These 975 original BESs (not masked) were categorized into 200 groups (200 sequence contigs) by PCAP.REP (Huang et al., 2003). The consensus sequences of the 200 sequence contigs were masked again by RepeatMasker, and 266 sequences of $>$50 bp were identified, of which 249 had at least six HSPs among the ZH11 BESs. These sequences were considered to be new rice repetitive sequences that were not collected in the GIRI and TIGR repeat databases.

Of the 249 new rice repetitive sequences, 231 had more than six HSPs, nine had fewer than six HSPs, and nine had no HSPs on the whole Nipponbare RefSeq. The latter 18 were considered to be ZH11 highly frequent sequences (HFSs). Because the BESs accounted for only 5.8% of the genome, the entire ZH11 genome was estimated to contain ~310 ZH11 HFSs, with at least 103 copies for each.

When the 18 ZH11 HFSs were used as core sequences to search the NCBI GSS database, they mainly hit the BESs of the wild rice species from the Oryza Map Alignment Project (www.omap.org). Table 3 lists the search results for different *Oryza* species. For wild rice species, we used the hits per every 100 Mb BESs to represent the repeat frequency because different wild rice species have different genome sizes and contain different numbers of BESs. For the *japonica* rice Nipponbare and *indica* rice 93–11, the total hits on the whole genomes are shown. For our ZH11 HFSs, six HSPs among 24 072 531 bp of BESs were set as the minimum standard, which corresponds to 24.92 hits per 100 Mb of corresponding sequences (6/24 072 531 $\times$ 100 000 000 bp; in this analysis, the organelle sequences in the BESs were neglected because wild rice BESs all contain a certain number of organelle sequences). As shown in Table 3, different core sequences had different hit numbers to different *Oryza* species. All of the core sequences had few or no hits to Nipponbare and 93–11. It is interesting that *O. rufipogon* and *O. australiensis* had high hit numbers for complementary sets of core sequences, and only *O. australiensis* showed hits for nine core sequences.

The core sequence 11.1.1 had only one and five hits to the entire Nipponbare and 93–11 genomes, respectively, and was used as a representative of the ZH11 HFSs for an overgo hybridization to confirm the analysis results (Figure 4). In agreement with the computer analysis, ZH11 contained high copy numbers of core sequence 11.1.1, whereas Nipponbare and 93–11 contained a number undetectable via the overgo hybridization. Of the other investigated rice varieties, namely ZH15, Dongjin, MDJ 8', MH63, IR24, IRBB23, and ZS97, only Dongjin and MH63 contained high and moderate copy numbers of this core sequence, respectively. ZH11 contained considerably more copies than Dongjin and MH63. ZH11 shared one band with Dongjin and two bands with MH63.

Substitutions and Indels between ZH11 and Nipponbare

SNPs, MNPs, and indels were analyzed by comparing the masked-BESs on the ZH11 contigs (11 932 497 bp) to the orthologous regions of the Nipponbare RefSeq. In total, 25 673 SNPs (25 673 bp) and 3114 MNPs (7 258 bp) were identified, which accounted for 0.215 and 0.061% of the analyzed sequence, respectively. In total, 8 911 (9 774 bp) insertions and 2 469 (3 160 bp) deletions were identified, which accounted for 0.082 and 0.026% of the analyzed sequence, respectively. The SNP frequency was much higher than the frequencies of the other two polymorphism types, indicating that simple variations may occur or be retained more easily. The distribution frequency of these polymorphic sites on chromosomes in 250-kb windows is shown in Supplemental Figure 3. They were distributed largely unevenly, with high densities on the ends of most chromosomes. An exception was found on chromosome 6: the densities of the polymorphic sites on the center region were much higher than those on the two sides. Ma and Bennetzen (2004) reported that, between the *japonica* rice

Table 3. The HSPs in ZH11 BESs and *Oryza* Sequences in the GSS Database (Measured for Every 100 Mb).

Core sequence	ZH11 BES	*Rufipogon*[a]	*Nivara*[a]	*Glaberrima*[a]	*Punctata*[a]	*Miuta*[a]	*Officinalis*[a]	*Alta*[a]	*Australiensis*[a]	*Granulata*[a]	*Ridleyi*[a]	*Nipponbare*[b]	*93–11*[b]
7.1.1	**33.23**	**34.00**	21.13	12.82	2.04		4.17	1.33		1.08		1	1
9.1.1	**49.85**	8.00			8.16	22.11	12.50	1.33		5.38	1.55	1	3
11.1.1	**33.23**	10.00		5.13	10.20	14.74	6.94	4.00		3.23	3.88	1	5
21.1.2	**29.08**	**26.00**	12.68	2.56	4.08		1.39	1.33		2.15	0.78	3	2
96.1.2	**29.08**	**26.00**	12.68	2.56	4.08		1.39	1.33		2.15	0.78	3	2
45.1.1	**24.92**	6.00	2.82				1.39	2.67				1	
51.1.1	**29.08**								**45.88**				
77.1.1	**29.08**	**34.00**	11.27	17.95	2.04	4.21	4.17		3.53			4	3
82.1.1	**29.08**								**27.06**				
86.1.2	**29.08**								**27.06**				
91.1.1	**20.77**								7.06				
92.1.1	**20.77**								7.06				
98.1.1	**29.08**								22.35				
104.1.1	**33.23**								22.35				
110.1.1	**33.23**								23.53				
121.1.1	**29.08**	12.00	4.23	2.56	2.04	2.11		5.33	1.18	1.08	2.33	6	4
121.1.2	**24.92**	16.00	5.63	5.13	4.08	2.11		6.67	2.35		2.33	3	5
125.1.1	**49.85**								**47.06**				

a The numbers larger than the threshold of **24.92 HSPs/100 Mb** are shown in bold.
b The numbers in the entire genome.

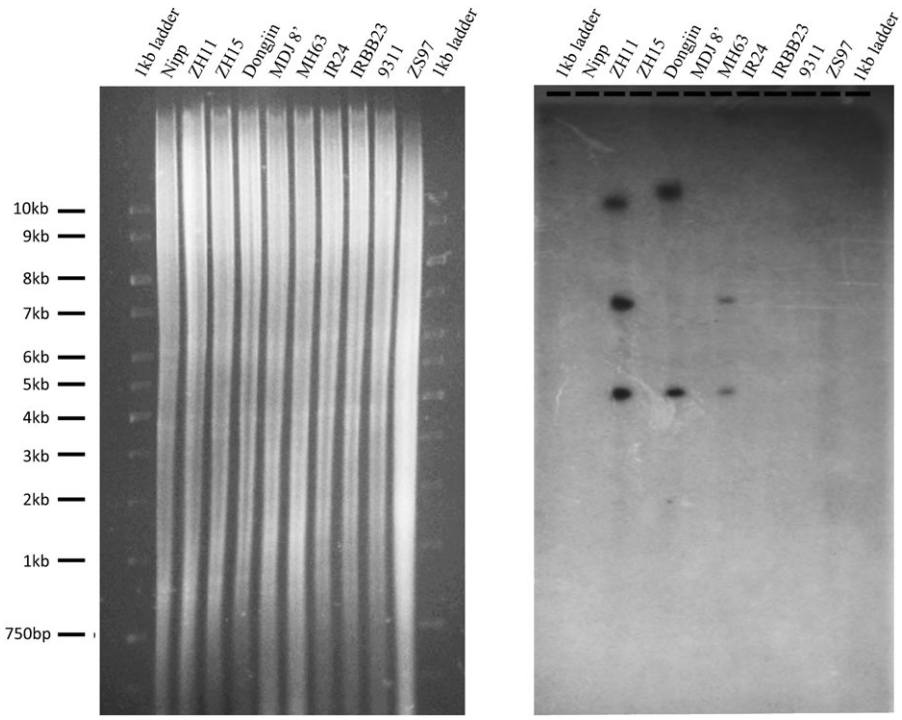

Figure 4. Overgo Hybridization with an 'Overgo' Probe for the ZH11 Highly Frequent Core Sequence 11.1.1. Genomic DNAs from various rice varieties (denoted at the top of the blot) were digested with *Hin*dIII and separated on an agarose gel. Size standards were loaded on both sides of the gel. The ethidium-bromide-stained image is shown as a loading control on the left of the blot.

variety Nipponbare and the *indica* rice variety GLA4, the rate of substitutions was linearly correlated with that of small indels. In agreement with this finding, our results also showed that, between the *japonica* rice varieties ZH11 and Nipponbare, the regions with higher densities of point mutations had higher densities of small indels.

BESs are single-path sequencing reads and may contain sequencing errors. To check the accuracy rates of the

polymorphic sites, we resequenced the BES regions of 244 SNPs and seven indels. All indels were verified. For the SNPs, 242 were verified and two were found to be sequencing errors. The overall accuracy rate of the polymorphic sites was 99.2%. The error rate was tightly correlated to the Phred quality of each nucleotide. The two BAC-end sequencing errors both had a Phred quality of below 20 (Supplemental Table 2).

Substitutions and Indels that Occurred in ZH11 and Nipponbare

To find the relative levels of substitutions and indels that occurred in ZH11 and Nipponbare, we used 93–11 as the outgroup reference. The *indica* rice variety 93–11 is the sister subspecies of the *japonica* subspecies. Thus, it should be an appropriate outgroup reference for intra-*japonica* genome analysis. Using Blat and ClustalW, tri-alignment blocks of >50 bp length and >98% homology with Nipponbare and 93–11 were identified, for a cumulative 5 852 773 bp of ZH11 masked-BESs. A high stringency (>98% homology) was used to ensure that only the orthologous regions were analyzed. Substitutions and indels in these tri-alignment blocks were compared. For the substitutions and insertions, only those with a Phred Q>20 were considered to ensure high accuracy. In total, 1650 substitutions between ZH11 and Nipponbare, 13 614 substitutions between ZH11 and 93–11, and 14 298 substitutions between Nipponbare and 93–11 were identified. These corresponded to substitution rates of 0.028, 0.233, and 0.244%, respectively. The substitution rates of the two *japonica–indica* inter-subspecies comparison combinations were close and almost a magnitude higher than the substitution rates between ZH11 and Nipponbare within the *japonica* subspecies. The substitution rate of 0.028% between ZH11 and Nipponbare was much less than the sum percentage of the SNP and MNP above because the comparison was limited to the conservative tri-alignment blocks and used a high stringency and a high Phred Q value.

Of the 1 650 substitutions between ZH11 and Nipponbare, 779 (47.2%) were assigned to ZH11 because the nucleotides at these sites were the same in Nipponbare and 93–11. Similarly, 868 (52.6%) were assigned to Nipponbare because the nucleotides at these sites were the same in ZH11 and 93–11. The substitution rate in ZH11 was lower than that in Nipponbare. Three (0.2%) sites could not be assigned to the varieties because they were also different in the 93–11 outgroup.

In total, 2 974 indels between ZH11 and Nipponbare, 4 608 indels between ZH11 and 93–11, and 1 922 indels between Nipponbare and 93–11 were identified. All of the indels identified here were small (1–2 bp), as determined by our settings, and contributed little to genome size variations in the compared regions. Of the 2 974 indels between ZH11 and Nipponbare, 1 689 (56.79%) insertions and 543 (18.26%) deletions were assigned to ZH11, and 68 (2.29%) insertions and 73 (2.45%) deletions were assigned to Nipponbare. The insertion number was about three times the deletion number in ZH11, whereas it was close to the deletion number in Nipponbare. The indels in ZH11 (2232) were 15.8 times of those in Nipponbare (141). Thus, the frequency of indels between ZH11 and Nipponbare (intra-subspecies comparison combination) was higher than that between Nipponbare and 93–11 (inter-subspecies comparison combination). Of the indels, 601 could not be assigned to varieties because they were also different in the 93–11 outgroup (Table 4).

Integration of ZH11 T-DNA Mutant Flanking Sequences into the Comparative Physical Map

The 16 158 ZH11 T-DNA FSTs were downloaded from the RMD database (http://rmd.ncpgr.cn) and mapped to the ZH11–Nipponbare comparative map. Of these sequences, 3154 (19.52%) were directly matched to 8519 BESs of the 7422 anchored ZH11 BACs on the 575 contigs. The rest of the FSTs were integrated into the comparative physical map through matches to the Nipponbare RefSeq (Figure 1). For all mapped FSTs, regardless of whether they were directly or indirectly mapped, ZH11 BAC clones and the corresponding Nipponbare RefSeq information can be browsed and obtained through our website (http://GResource.hzau.edu.cn). A link directly to the RMD database (http://rmd.ncpgr.cn) was established for every mapped FST. We also established a search tool (http://GResource.hzau.edu.cn). The corresponding ZH11 BAC clones for the additional ZH11 FSTs produced in the above laboratory or in other laboratories as well as for the molecular markers and all sequences homologous to the Nipponbare RefSeq can be found via the search tool.

DISCUSSION

Intra-*Japonica* Subspecies Comparative Physical Mapping

Although extensive genome comparisons among species of the genus *Oryza* (www.omap.org) (Kim et al., 2007) and

Table 4. Variations in the Tri-Alignment Regions of ZH11, Nipponbare, and 93–11.

Variation	Counts	Percentages
Substitutions		
Between Nipponbare and 93–11	14 298	
Between ZH11 and 93–11	13 614	
Between ZH11 and Nipponbare	1 650	
Occurring in ZH11	779	47.2
Occurring in Nipponbare	868	52.6
Not determined	3	0.2
Indels		
Between Nipponbare and 93–11	1 922	
Between ZH11 and 93–11	4 608	
Between ZH11 and Nipponbare	2 974	
Insertions occurring in ZH11	1 689	56.79
Insertions occurring in Nipponbare	68	2.29
Deletions occurring in ZH11	543	18.26
Deletions occurring in Nipponbare	73	2.45
Not determined	601	20.21

between the rice subspecies *japonica* and *indica* (Han and Xue, 2003; Ma and Bennetzen, 2004) have been performed, no extensive genome comparisons between varieties within a subspecies have been performed to date because of the lack of a second independent genome sequence or a comparable physical map for a variety in either subspecies. We constructed a *de novo* physical map containing a large number of embedded BESs for the *japonica* rice variety ZH11 and performed an extensive intra-*japonica* comparison between our map and the Nipponbare RefSeq. Although the Nipponbare RefSeq was used to find errors and discrepancies of the ZH11 contig assemblies during manual editing of the ZH11 physical map, the ZH11 physical map is independent because the correction of errors and verification of discrepancies were accomplished through the analysis of ZH11 fingerprinting data.

Intra-*Japonica* Subspecies Variation

The comparative physical map showed that the two *japonica* rice genomes are generally highly syntenic and conserved. No translocations were found. We did not find any very large contractions/expansions (>500 kb). Although 17 locations were detected that contained a total of 46 abnormally aligned BESs, the validated locations were all found to be assembly errors of the Nipponbare RefSeq instead of inversions between the two *japonica* rice genomes. These results contrast with those from comparisons between species genomes of the genus *Oryza* spanning ~15 million years of evolution (Kim et al., 2007; Ammiraju et al., 2008; Hurwitz et al., 2010). Hurwitz et al. (2010) identified 674 expansions, 611 contractions, and 140 putative inversions, accounting for 25–38, 4–19, and 14–19 Mb, respectively, in *O. nivara*, *O. rufipogon*, and *O. glaberrima* genomes relative to the *O. sativa* cv. Nipponbare genome. At least 16 cases of large DNA contraction/expansion (>500 kb), inversion, and translocation were found between *O. punctata* (BB genome type) and *O. sativa* cv. Nipponbare genomes (Kim et al., 2007).

However, substantial variations between the two *japonica* varieties were also found. Contractions as large as 450 kb in ZH11 relative to Nipponbare, along with many smaller contractions/expansions, were found. SNPs, MNPs, and indels were detected for 0.215, 0.061, and 0.108% of the analyzed sequence, respectively. The frequency of SNPs was about 3.5 and 2 times the frequencies of MNPs and indels, respectively. An intriguing discovery was that the rates of substitutions and indels occurring in different individual varieties were very different. Even within a variety, the rates of insertions and deletions can be very different. These results indicate that different mechanisms may exist for substitution, insertion, and deletion. Our case provides an example that the indel frequencies between varieties do not necessarily reflect phylogenetic distances.

Another intriguing discovery is that the ZH11 genome contained many HFSs that were not found or were only found in a few copies in the Nipponbare RefSeq. In 5.8% of the ZH11 genome sampling sequences (BESs), 18 highly frequent core sequences with at least six HSPs each were found. The 18 core sequences were all found, in either high or low copy numbers, in at least one different *Oryza* species (Table 3), indicating that they are ancient sequences. Nine were found in many *Oryza* species, whereas the other nine were found only in *O. australiensis* and were not found in the sequenced Nipponbare and 93–11 genomes. Interestingly, the wild rice species *O. rufipogon* and *O. australiensis* each contained an almost complementary subset of these core sequences, many with high copy numbers. These two wild rice species may employ a mechanism to maintain and amplify the subset HFSs. In fact, it was reported that *O. australiensis* experienced recent species-specific bursts of three retrotransposon families, namely RIRE1, Kangourou, and Wallabi, which resulted in a twofold increase in genome size. These three families had only 2, 10, and 16 complete copies, respectively, in the Nipponbare RefSeq (Piegu et al., 2006). The 18 ZH11 HFSs did not include any homologs of the three retrotransposon families because the three family sequences were collected in the repeat sequence database and masked in our analyses. We confirmed the variety-specific amplification of the core sequence 11.1.1 by overgo hybridization (Figure 4). We could not get any full-length sequences for the ZH11 HFSs from the BESs; more sequence information is needed to elucidate their structure, organization, and evolution. Our results demonstrate that the ZH11 HFSs must be proliferated after speciation and not horizontally transferred. First, they are ancient sequences and exist in at least one wild rice species. Second, many exhibited much higher copy numbers in ZH11 than in other species. Third, for sequence 11.1.1, ZH11 had higher copy numbers and more bands than Dongjin and MH63 (Figure 4).

Repetitive sequences are major evolutionary forces and genome components (Ma et al., 2004). To investigate whether the ZH11 HFSs caused the ZH11 genome size to surpass the Nipponbare genome size, we estimated the sizes of the aligned ZH11 contigs and compared them to the corresponding regions of the Nipponbare RefSeq. The distribution frequency of the size differences between the ZH11 contigs and the corresponding regions of the Nipponbare RefSeq is shown in Supplemental Figure 4. The comparison revealed that parts of the contigs had larger sizes (expanded) and that parts of the contigs had smaller sizes (contracted), whereas the rest had the same sizes. The total aligned contigs were, on average, 8% smaller than the corresponding regions of the Nipponbare RefSeq. The estimation depends on the accuracy of the average BAC insert size, and an underestimation of as much as 8% may still be possible for BAC clone sampling. However, our BES analysis revealed that the ZH11 genome contained 31.93% of the known rice repeat sequences. According to the recent report that the Nipponbare genome contains 38.87% of the known rice repeat sequences (Feng et al., 2009), the ZH11 genome is thus about 7% smaller than the Nipponbare genome. Although about 310 ZH11 HFSs were predicted, some of them may have come from the same repetitive sequence units, and their total contribution to the genome size could be limited. We found at least 231 new rice repetitive sequences (not collected in the repeat

sequence database); their relative contribution to the sizes of the two genomes is not known. Our analysis procedure could not detect the new repetitive sequences that exist in high copy numbers only in Nipponbare. Our result showed that the largest contractions/expansions identified were contractions of ZH11 relative to Nipponbare (Supplemental Figure 2 and Figure 2A). The Nipponbare genome size was also reported to be larger than those of the *indica* cultivars GLA4 (Ma and Bennetzen, 2004) and 93–11 (Han and Xue, 2003). It is possible that most of the Nipponbare size increase occurred after its divergence from ZH11. Nevertheless, the ZH11 genome size is not likely to be larger than that of Nipponbare.

Utilization of Resources

We report here a rice intra-subspecies comparative physical map that provides a broad view of the extent of variations among rice intra-subspecies varieties and mutations occurring in individual varieties. This study also provides a platform for whole-genome sequencing and assembly of the ZH11 genome as well as comparisons among rice varieties of both *japonica* and *indica* subspecies.

The Nipponbare RefSeq still contains many physical gaps, although exhaustive efforts have been applied to close those gaps (Ammiraju et al., 2005; IRGSP, 2005). Possible reasons for the inability to fill the gaps could be that these genomic regions may be recalcitrant to cloning or detection. The ZH11 comparative map covered four Nipponbare physical gaps. The gap-spanning ZH11 BAC contigs provide tools and guidance to explore the gaps. The ZH11 comparative map also detected assembly errors in the Nipponbare RefSeq that could not be easily identified by other approaches.

Mutants are important resources for gene function studies because they provide direct clues to link genes with phenotypes (Krishnan et al., 2009). ZH11 is relatively easily transformed with *Agrobacterium* and has been extensively used as a host to construct T-DNA insertion mutant libraries for rice (Zhang et al., 2006, 2007). A large ZH11 T-DNA insertion library with a large number of FSTs was released (Zhang et al., 2007), and additional insertion lines and FSTs are being generated (http://rmd.ncpgr.cn). In addition, ZH11 is also being used or has been used to generate an EMS mutant library (in progress), a ^{60}Co-gamma radiation mutant library (Yan et al., 2007), and a genetic mapping population (Guo et al., 2009) to clone functional genes. The extensive variations between varieties found in this study demonstrate that sequence information and DNA templates from the host or parent are eventually required to confirm function through complementation. Although most FSTs and molecular markers cannot be directly matched to the ZH11 sequences because of the lack of the ZH11 whole-genome sequence, they can be indirectly matched to the corresponding ZH11 BAC clones for DNA templates through the RefSeq. A comparative map provides the ability to use the reference sequence to find host or parent DNA templates. The ZH11-Nipponbare comparative map presented here adds a new and powerful dimension for accelerated gene cloning, functional analyses, and comparative genomics.

METHODS

BAC Library Construction

The BAC vector was prepared and the BAC library was constructed as previously described (Luo et al., 2001; Luo and Wing, 2003). Megabase genomic DNA was isolated from young ZH11 seedlings, and *Hin*dIII restriction digestion was used to construct the BAC library. Copies of the BAC library were stored at –80°C at both the Arizona Genomics Institute (www.genome.arizona.edu) and the Genome Resource Laboratory of Huazhong Agricultural University (http://GResource.hzau.edu.cn). BAC clones and filters are available at our websites. Insert sizing was performed for 229 randomly selected BAC clones from the library on 1% agarose CHEF gels (Bio-Rad) with a 5–15-s linear ramp time at 6 V cm^{-1} and 14°C in 0.5X TBE buffer for 16 h.

BAC-End Sequencing

The BAC clones from the first 48 × 384-well plates (plate Nos 1–48; a total of 18 432 clones) were sequenced at both ends as previously described (Luo et al., 2006). A T7 primer (5′ TAA TAC GACTCACTA TAGGG 3′) and a BES_HR primer (5′ CACTCA TTAGGCACCCCA 3′) were used as forward and reverse primers, respectively. Sequencing was performed on ABI 3730 × l DNA capillary sequencers with default conditions and the accompanying software (Applied Biosystems, ABI, Foster City, CA). Sequences were basecalled using the program Phred (Ewing and Green, 1998; Ewing et al., 1998), and vector and low-quality (Phred value <16) sequences were removed by LUCY (Chou and Holmes, 2001). All sequences were deposited in GenBank.

Fingerprinting and FPC Assembly

The same set of BAC clones was also fingerprinted. Fingerprinting and FPC assembly were performed as in Kim et al. (2007). The SNaPshot technique (Luo et al., 2003) was used to produce BAC fingerprints. FPC software was used to assemble the fingerprints into contigs (Soderlund et al., 2000). BAC clones with fewer than 25 or more than 180 fingerprint bands were excluded from FPC assembly. This primary FPC assembly was defined as a phase I physical map as in Kim et al. (2007).

Contig Alignment and Manual Editing

Repeat, organelle, and SSR sequences were masked from the BESs by RepeatMasker software. The masked BESs on contigs were aligned to the Nipponbare RefSeq (IRGSP, Build 4) through sequence-similarity searches using BLAT (Kent, 2002). The alignments between the ZH11 FPC contigs and the Nipponbare RefSeq were displayed by SyMap (Soderlund et al., 2006). The alignments of the contigs were manually edited against the Nipponbare RefSeq as in Kim et al. (2007).

Discrepant regions were inspected and re-analyzed. The contigs were merged if both the neighboring ends overlapped or were bridged by at least two paired clones and the merged region showed an improved alignment to the Nipponbare RefSeq. Only regions inside contigs were compared with the Nipponbare RefSeq.

BES Analyses

Repetitive sequences were identified by RepeatMasker (www.repeatmasker.org) against a local database combining the rice repetitive sequences downloaded from GIRI (Genetic Information Research Institute: www.girinst.org/) and TIGR (The Institute for Genomic Research: www.tigr.org/). The ZH11 HFSs were identified using the method and standards in Aggarwal et al. (2009). SNPs, MNPs, and indels were identified by Blastn (Altschul et al., 1990). Validation of variations between the ZH11 BESs and the Nipponbare RefSeq was accomplished by resequencing the PCR products from the BAC ends. Organelle sequences were identified by Blastn against the rice organelle database in GenBank.

Overgo Hybridization

Overgo hybridization was performed as previously described (Luo et al., 2006). Genomic DNA was purified from rice seedlings using CTAB and was digested with *Hin*dIII. DNA (10 μg) samples were separated on 1% agarose gel and were transferred to a Hybond N^+ membrane. The two Overgo primers for the ZH11 highly frequent core sequence 11.1.1 were Primer F: 5′ CTACTCCTTACCTTATCCGTTGCCGCCT 3′ and Primer R: 5′ AAAATTGAGGACCCACATCTAGGCGGCA 3′. The eight nucleotides at the 3′ ends are complementary to each other. The probe was prepared by incorporating both ^{32}P-dCTP and ^{32}P-dATP through Klenow synthesis.

Substitution and Indel Analysis of the Tri-Alignment Regions of ZH11, Nipponbare, and 93–11

Substitution (SNP and MNP) and indel analyses of the tri-alignment regions were performed as described in Ma and Bennetzen (2004). Orthologous sequences among the ZH11 BESs and Nipponbare and 93–11 genome sequences were identified by Blat (Kent, 2002) and then aligned by ClustalW (Thompson et al., 1994). Substitutions and indels occurring in individual ZH11 and Nipponbare varieties of the *japonica* subspecies were determined using 93–11 of the *indica* subspecies as an outgroup.

Integration of the ZH11 T-DNA FSTs with the Comparative Physical Map

ZH11 T-DNA FSTs were downloaded from RMD (Rice Mutant Database, http://rmd.ncpgr.cn) and compared with ZH11 BESs and the Nipponbare RefSeq by Blastn (Altschul et al., 1990). The positions of the FSTs and annotations from RAP2 (http://rapdb.dna.affrc.go.jp/) for the Nipponbare RefSeq were displayed on the comparative map by SyMap (Soderlund et al., 2006).

Accession Numbers

The BES sequences reported in this paper have been deposited in the GenBank database (accession nos GS599267∼GS635185).

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

FUNDING

This work was supported by a grant from the National Natural Science Foundation of China for International Collaboration (30620120431) and the 111 Project B07041.

ACKNOWLEDGMENTS

We thank Dave Kudrna, Nicholas B. Sisneros, Qiang Song, Fusheng Wei, José L. Goicoechea, HyeRan Kim, Yeisoo Yu, Jetty S.S. Ammiraju, and other members of the Arizona Genomics Institute for production of the BAC library, fingerprints, and BESs and for student training. We also thank Yonglong Pan and Jun Xu for BAC library insert sizing. No conflict of interest declared.

REFERENCES

Aggarwal, R., Benatti, T.R., Gill, N., Zhao, C., Chen, M.S., Fellers, J.P., Schemerhorn, B.J., and Stuart, J.J. (2009). A BAC-based physical map of the Hessian fly genome anchored to polytene chromosomes. BMC Genomics. **10,** 293.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

Ammiraju, J.S., et al. (2005). Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. Nipponbare) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome. Theor. Appl. Genet. **111,** 1596–1607.

Ammiraju, J.S., et al. (2006). The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. Genome Res. **16,** 140–147.

Ammiraju, J.S., et al. (2008). Dynamic evolution of oryza genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. Plant Cell. **20,** 3191–3209.

Ammiraju, J.S., et al. (2010). Spatio-temporal patterns of genome evolution in allotetraploid species of the genus Oryza. Plant J. **63,** 430–442.

Chen, S.Y., Wang, A.M., Li, W., Wang, Z.Y., and Cai, X.L. (2008). Establishing a gene trap system mediated by T-DNA(GUS). in rice. J. Integr. Plant Biol. **50,** 742–751.

Chou, H.H., and Holmes, M.H. (2001). DNA sequence quality trimming and vector removal. Bioinformatics. **17,** 1093–1104.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8,** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8,** 175–185.

Feng, Q., Huang, T., Zhao, Q., Zhu, J., Lin, Z., and Han, B. (2009). Analysis of collinear regions of Oryza AA and CC genomes. J. Genet. Genomics. **36,** 667–677.

Fu, H., and Dooner, H.K. (2002). Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl Acad. Sci. U S A. **99,** 9573–9578.

Gale, M.D., and Devos, K.M. (1998). Comparative genetics in the grasses. Proc. Natl Acad. Sci. U S A. **95,** 1971–1974.

Ge, S., Sang, T., Lu, B.R., and Hong, D.Y. (1999). Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl Acad. Sci. U S A. **96,** 14400–14405.

Guo, L., Ma, L., Jiang, H., Zeng, D., Hu, J., Wu, L., Gao, Z., Zhang, G., and Qian, Q. (2009). Genetic analysis and fine mapping of two genes for grain shape and weight in rice. J. Integr. Plant Biol. **51,** 45–51.

Han, B., and Xue, Y. (2003). Genome-wide intraspecific DNA-sequence variations in rice. Curr. Opin. Plant Biol. **6,** 134–138.

Huang, X., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42,** 961–967.

Huang, X., Wang, J., Aluru, S., Yang, S.P., and Hillier, L. (2003). PCAP: a whole-genome assembly program. Genome Res. **13,** 2164–2170.

Hurwitz, B.L., Kudrna, D., Yu, Y., Sebastian, A., Zuccolo, A., Jackson, S.A., Ware, D., Wing, R.A., and Stein, L. (2010). Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus Oryza. Plant J. **63,** 990–1003.

International Rice Genome Sequencing Project (IRGSP) (2005). The map-based sequence of the rice genome. Nature. **436,** 793–800.

Kent, W.J. (2002). BLAT-The BLAST-LIKE Alignment Tool. Genome Res. **12,** 656–664.

Kim, H., et al. (2008). Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of genus Oryza. Genome Biol. **9,** 45–59.

Kim, H., San Miguel, P., Nelson, W., Collura, K., Wissotski, M., Walling, J.G., Kim, J.P., Jackson, S.A., Soderlund, C., and Wing, R.A. (2007). Comparative physical mapping between *Oryza sativa* (AA genome type) and *O. punctata* (BB genome type). Genetics. **176,** 379–390.

Krishnan, A., et al. (2009). Mutant resources in rice for functional genomics of the grasses. Plant Physiol. **149,** 165–170.

Lu, F., et al. (2009). Comparative sequence analysis of MONOCULM1-orthologous regions in 14 Oryza genomes. Proc. Natl Acad. Sci. U S A. **106,** 2071–2076.

Luo, M., and Wing, R.A. (2003). An improved method for plant BAC library construction. Methods Mol. Biol. **236,** 3–20.

Luo, M., et al. (2006). Construction of a nurse shark (*Ginglymostoma cirratum*) bacterial artificial chromosome (BAC) library and a preliminary genome survey. BMC Genomics. **7,** 106.

Luo, M., Wang, Y.H., Frisch, D., Joobeur, T., Wing, R.A., and Dean, R.A. (2001). Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium wilt (Fom-2). Genome. **44,** 154–162.

Luo, M.C., Thomas, C., You, F.M., Hsiao, J., Ouyang, S., Buell, C.R., Malandro, M., McGuire, P.E., Anderson, O.D., and Dvorak, J. (2003). High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics. **82,** 378–389.

Ma, J., and Bennetzen, J.L. (2004). Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl Acad. Sci. U S A. **101,** 12404–12410.

Ma, J., Devos, K.M., and Bennetzen, J.L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. **14,** 860–869.

McNally, K.L., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc. Natl Acad. Sci. U S A. **106,** 12273–12278.

Piegu, B., et al. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. **16,** 1262–1269.

Soderlund, C., Humphray, S., Dunham, A., and French, L. (2000). Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. **10,** 1772–1787.

Soderlund, C., Longden, I., and Mott, R. (1997). FPC: a system for building contigs form restriction fingerprinted clones. Comput Appl. Biosci. **13,** 523–535.

Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: a systeming for discovering and viewing sytenic regions of FPC maps. Genome Res. **16,** 1159–1168.

Tang, H., Bowers, J.E., Wang, X., and Paterson, A.H. (2010). Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc. Natl Acad. Sci. U S A. **107,** 472–477.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22,** 4673–4680.

Wing, R.A., et al. (2005). The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol. Biol. **59,** 53–62.

Wu, C., Li, X., Yuan, W., Chen, G., Kilian, A., Li, J., Xu, C., Zhou, D.X., Wang, S., and Zhang, Q. (2003). Development of enhancer trap lines for functional analysis of the rice genome. Plant J. **35,** 418–427.

Yan, C., Yan, S., Zeng, X., Zhang, Z., and Gu, M. (2007). Fine mapping and isolation of Bc7(t), allelic to OsCesA4. J. Genet. Genomics. **34,** 1019–1027.

Yu, J., et al. (2005). The genomes of Oryza sativa: a history of duplications. PLoS Biol. **3,** e38.

Yu, J., Ni, P., and Wong, G.K. (2006). Comparing the whole-genome-shotgun and map-based sequences of the rice genome. Trends Plant Sci. **11,** 387–391.

Zhang, J., et al. (2007). Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by

analyzing 13 804 T-DNA flanking sequences from an enhancer-trap mutant library. Plant J. **49,** 947–959.

Zhang, J., Feng, Q., Jin, C., Qiu, D., Zhang, L., Xie, K., Yuan, D., Han, B., Zhang, Q., and Wang, S. (2005). Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63. Plant J. **42,** 772–780.

Zhang, J., Li, C., Wu, C., Xiong, L., Chen, G., Zhang, Q., and Wang, S. (2006). RMD: a rice mutant database for funtional analysis of the rice genome. Nucleic Acids Res. **34,** 745–748.

Zhang, X., and Wessler, S.R. (2004). Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. Proc. Natl Acad. Sci. U S A. **101,** 5589–5594.