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*

Jetty S.S. Ammiraju, Meizhong Luo, José L. Goicoechea, et al.

Genome Res. 2006 16: 140-147
Access the most recent version at doi:10.1101/gr.3766306

Resource

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*

Jetty S.S. Ammiraju,[1,3,12] Meizhong Luo,[1,3,12] José L. Goicoechea,[1,3,12] Wenming Wang,[1,3] Dave Kudrna,[1,3] Christopher Mueller,[1,3] Jayson Talag,[1,3,6] HyeRan Kim,[1,3] Nicholas B. Sisneros,[1,3] Barbara Blackmon,[5] Eric Fang,[5] Jeffery B. Tomkins,[5] Darshan Brar,[6] David MacKill,[6] Susan McCouch,[7] Nori Kurata,[8] Georgina Lambert,[2,3] David W. Galbraith,[2,3] K. Arumuganathan,[9] Kiran Rao,[1,3,4] Jason G. Walling,[10] Navdeep Gill,[10] Yeisoo Yu,[1,3] Phillip SanMiguel,[11] Carol Soderlund,[3,4] Scott Jackson,[10] and Rod A. Wing[1,3,13]

[1]Arizona Genomics Institute, [2]Department of Plant Sciences, [3]BIO5 Institute, and [4]Arizona Genomics Computational Laboratory, University of Arizona, Tucson, Arizona 85721 USA; [5]Clemson University Genomics Institute, Clemson University, Clemson, South Carolina 29634, USA; [6]Department of Plant Breeding and Genetics, International Rice Research Institute (IRRI), Los Baños, 4031 The Philippines; [7]Department of Plant Breeding, Cornell University, Ithaca, New York 14853, USA; [8]National Institute of Genetics (NIG), Shizuoka 411-8540, Japan; [9]Flow Cytometry and Imaging Core Laboratory, Benaroya Research Institute at Virginia Mason, Seattle, Washington 98101, USA; [10]Department of Agronomy, [11]Genomics Core Facility, Purdue University, West Lafayette, Indiana 47907, USA

Rice (*Oryza sativa* L.) is the most important food crop in the world and a model system for plant biology. With the completion of a finished genome sequence we must now functionally characterize the rice genome by a variety of methods, including comparative genomic analysis between cereal species and within the genus *Oryza*. *Oryza* contains two cultivated and 22 wild species that represent 10 distinct genome types. The wild species contain an essentially untapped reservoir of agriculturally important genes that must be harnessed if we are to maintain a safe and secure food supply for the 21st century. As a first step to functionally characterize the rice genome from a comparative standpoint, we report the construction and analysis of a comprehensive set of 12 BAC libraries that represent the 10 genome types of *Oryza*. To estimate the number of clones required to generate 10 genome equivalent BAC libraries we determined the genome sizes of nine of the 12 species using flow cytometry. Each library represents a minimum of 10 genome equivalents, has an average insert size range between 123 and 161 kb, an average organellar content of 0.4%–4.1% and nonrecombinant content between 0% and 5%. Genome coverage was estimated mathematically and empirically by hybridization and extensive contig and BAC end sequence analysis. A preliminary analysis of BAC end sequences of clones from these libraries indicated that LTR retrotransposons are the predominant class of repeat elements in *Oryza* and a roughly linear relationship of these elements with genome size was observed.

[Supplemental material is available online at www.genome.org. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: J. Mullet, T. Sasaki, M. Luo, A. Jetty, R.A. Wing, H.R. Kim, B. Gill, and S. McCouch.]

A finished, quality, whole genome sequence for key model animals and plants, such as *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Arabidopsis thaliana*, and *Oryza sativa*, provides an essential and powerful resource for comparative functional and evolutionary analysis of related genera and species. The recently finished rice genome (*O. sativa* ssp. *japonica*; International Rice Genome Sequencing Project [IRGSP] 2005) is considered the "Rosetta Stone" to unlock the secrets of all major cereal genomes that are used to feed the world (rice, sorghum, millet, corn, barley, oat, and wheat) as well as the wild relatives of rice within the genus *Oryza*.

Oryza is a complex but relatively small genus with two cultivated and 22 wild species (Ge et al. 1999). Morphological, cytological, and molecular divergence studies have classified the species of *Oryza* into 10 genome types, namely, AA, BB, CC,

[12]These authors contributed equally to the work.
[13]Corresponding author.
E-mail rwing@ag.arizona.edu; fax (520) 621-1259.
Article published online ahead of print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.3766306.

16:140–147 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org

BBCC, CCDD, EE, FF, GG, HHJJ, and HHKK (Aggarwal et al. 1997; Khush 1997; Ge et al. 1999) with the cultivated species, *O. sativa* (Asian rice) and *O. glaberrima* (African rice), designated as AA genome diploids ($2n = 24$). Within the genus, genome size varies several-fold (Iyengar and Sen 1978; Martinez et al. 1994; Uozu et al. 1997), polyploidy exists, and there are structural chromosomal changes between species (Huang and Kochert 1994; Jena et al. 1994; Hass et al. 2003). *Oryza* species have already provided genes for the hybrid rice revolution, yield enhancing traits (Xiao et al. 1996, 1998) and tolerance to biotic and abiotic stress (Brar and Khush 1997). However, genetic variation contained within the wild *Oryza* gene pool has been largely untapped.

To better understand wild rice species and take advantage of the rice genome sequence (IRGSP 2005), we have embarked on a comparative genomics program entitled the "*Oryza* Map Alignment Project" (OMAP). The long-term objective of this program is to create a genome-level closed experimental system for the genus *Oryza* by developing comparative BAC-based physical maps of all 10 genome types of the genus to study evolution, genome organization, domestication, gene regulatory networks, and crop improvement (Wing et al. 2005).

As a first step toward achieving this goal, we report the construction and detailed characterization of 12 high-quality BAC libraries from one cultivated (*O. glaberrima*) and 11 well-characterized wild species representing the 10 genome types of *Oryza*. We selected these species in consultation with breeders and basic researchers with emphasis on the presence of traits of potential agronomic importance (Supplemental Table 1) and, in some cases, the availability of mapping populations. Having convenient public access to the other nine genomes of *Oryza* in the form of BAC libraries will permit rapid advances in both basic and applied research for the most important food crop in the world.

Table 1. Nuclear DNA content of *Oryza* species estimated by flow cytometry

Species	Genome type	Nuclear DNA content: This study pg/2C	This study Mbp/1C	This study SD[e]	Previous reports Mbp/1C	References[a]
O. nivara	AA	0.93	448	0.01	760	1
O. rufipogon	AA	0.91	439	0.01	760	1
					459	2
O. glaberrima[b]	AA	ND[c]	ND[c]	ND[c]	809	1
					357	3
O. punctata	BB	0.88	425	0.18	539	1
					535	2
O. officinalis	CC	1.35	651	0.02	1201	1
					550	3
					697	2
O. minuta[b]	BBCC	ND[c]	ND[c]	ND[c]	1691	1
					1124	3
O. alta	CCDD	2.09	1008	0.019	—	—
O. australiensis	EE	2.00	965	0.8	1056	1
					946	2
					960	3
O. brachyantha	FF	0.75	362	0.07	346	1, 2
O. granulata	GG	1.83	882	0.28	907	1
O. ridleyi	HHJJ	2.66	1283	0.14	1568	1
					632–931	3
O. coarctata[d]	HHKK	ND[c]	ND[c]	ND[c]	—	—

[a]Reference sources: (1) Angiosperm C value database (unless indicated, the values are from Iyengar and Sen 1978). (2) Uozu et al. 1997. (3) Martinez et al. 1994. — not reported in C-value database.
[b]Genome sizes for *O. glaberrima* (0.72–0.76 pg/2C) and *O. minuta* (2.33 pg/2C) were adapted from Martinez et al. (1994).
[c]Not determined.
[d]Genome size of *O. coarctata* (2.66 pg/2C) was estimated based on *O. ridleyi* data.
[e]Standard deviation.

Results

Nuclear DNA content of *Oryza* species as measured by flow cytometry

The genome sizes of nine of the 12 *Oryza* accessions used to construct BAC libraries were determined by flow cytometry. The 1C values for *O. glaberrima* [AA; 357 Mb] and *O. minuta* [BBCC; 1124 Mb] were adopted from previous flow cytometric data (Martinez et al. 1994). The 1C value for *O. coarctata* [HHKK] was not measured because of quarantine restrictions. We therefore used the value estimated for *O. ridleyi* [HHJJ; 1283 Mb], which is also an allotetraploid species and shares the HH genome type with *O. coarctata*.

Table 1 compares the results of the nuclear DNA content analysis with previously reported studies. Single peaks obtained from our analysis indicated that the nuclei preparations did not contain dividing cells. The genome sizes of the various rice species vary by as much as 3.6-fold with *O. brachyantha* [FF] and *O. glaberrima* [AA] having the smallest (0.75 pg/2C and 0.74 pg/2C, respectively), while *O. minuta* [BBCC] and *O. ridleyi* [HHJJ], both tetraploids, have the largest (2.33 and 2.66 pg/2C). *O. alta* [CCDD] has a genome size of 1008 Mb, and this is the first report of a genome size for this species. Among the diploid species, *O. australiensis* [EE] (2.0 pg/2C) has the largest genome, followed by *O. granulata* [GG] (1.83 pg/2C). The other AA genome species, *O. nivara* and *O. rufipogon*, contain less nuclear DNA than the CC and EE genomes. Compared to the AA genome species *O. nivara* and *O. rufipogon*, their closest relative *O. punctata* [BB] has a 3%–5% smaller genome size (~425 Mb).

BAC library construction and characterization

BAC library construction followed standard protocols (Luo and Wing 2003). Briefly, megabase-size DNA for each accession was prepared from nuclei embedded in agarose plugs. HindIII partially digested, size-selected DNA fragments were then ligated into pIndigoBAC536 SwaI and transformed into *Escherichia coli*. Often, more than one ligation, having different insert sizes and transformation efficiencies, was used to achieve the required number of clones for 10-fold redundancy for each library. The number of clones per library ranged between 36,864 and 147,456, which were arrayed in 384-well microtiter plates (Table 2) and stored at −80°C.

To determine the average insert size and percent recombinant clones for each library, we analyzed 400–700 randomly picked clones, including clones from all the different ligations and at least one clone from every 384-well plate, depending on genome size. Insert sizes ranged from 10 kb to 300 kb, with a majority of fragments in the 120–150 kb size range (Supplemen-

Table 2. Characteristics of the *Oryza* BAC library resource

Species	Genome	Accession number	Total number of clones	% Non-insert-containing clones	% Total organellar DNA content	Avg insert size (kb)	Calculated genome coverage[a]
O. nivara	AA	W0106	55,296	0.0	4.1	161	19.0
O. rufipogon	AA	105491	64,512	0.2	3.7	134	18.9
O. glaberrima	AA	96717	55,296	0.5	3.2	130	19.3
O. punctata	BB	105690	36,864	0.8	1.0	142	12.0
O. officinalis	CC	100896	92,160	5.0	1.7	141	18.6
O. minuta	BBCC	101141	129,024	0.7	0.5	125	14.1
O. alta	CCDD	105143	92,160	0.0	0.4	133	12.1
O. australiensis	EE	W0008	92,160	0.4	2.3	153	14.2
O. brachyantha	FF	101232	36,864	0.0	1.7	131	13.1
O. granulata	GG	102118	73,728	0.7	2.7	134	10.8
O. ridleyi	HHJJ	100821	129,024	1.7	0.4	127	12.5
O. coarctata	HHKK	104502	147,456	0.6	1.5	123	13.8[b]

[a]Genome coverage after subtraction of organellar and non-insert-containing clones.
[b]Genome coverage estimated from adapted genome size value of *O. ridleyi.*

tal Fig. 1). Insert size distributions for the *O. nivara* and *O. australiensis* libraries (Supplemental Fig. 1) did not follow the expected Poisson distribution and may be explained by the use of multiple ligation mixes used to construct those libraries. The percentage of nonrecombinant clones was between 0% and 5%, indicating that more than 95% of the clones in these libraries contain inserts. The average insert sizes of these libraries ranged between 123 and 161 kb (Table 2).

To estimate the percentage of organellar DNA content, the libraries were screened with three chloroplast and four mitochondrial probes. Results showed that the libraries contained approximately 0.09%–3.9% chloroplast and 0%–0.7% mitochondrial DNA sequences (Supplemental Table 2), which is typically observed using similar DNA preparations (Luo et al. 2001).

By using the genome size, average insert size, and the number of clones for each library, after subtraction of organellar and nonrecombinant contaminants, we estimate that the theoretical genome coverage of each *Oryza* library is between 10.8- and 19.3-fold (Table 2).

Estimation of genome coverage by hybridization and contig analysis

To independently assess the genome coverage of each BAC library, a probe set representing a single locus from each of the 12 rice chromosomes (Supplemental Table 2) was hybridized to each library, and positive BAC clones were analyzed for their ability to assemble into FPC contigs. For 8 of the 12 libraries (*O. nivara* [AA], *O. rufipogon* [AA], *O. glaberrima* [AA], *O. punctata* [BB], *O. minuta* [BBCC], *O. australiensis* [EE], *O. brachyantha* [FF], and *O. coarctata* [HHKK]), preliminary FPC/BES physical maps were available and composed of a calculated minimum of 8.6× genome coverage per library. Using these FPC maps, contigs for all but 13 out of the 120 possible contigs were identified (Supplemental Table 2). Upon manual inspection of each FPC contig, we immediately noticed that not all clones in each contig were identified by hybridization. We therefore performed an extended analysis to determine if any BAC end sequences derived from the clones in the FPC contigs, which were not identified by hybridization, could be mapped to the predicted location on the sequenced rice genome. In 93 out of 107 contigs analyzed, at least one BAC clone could be confirmed to be in the correct orthologous position but was not detected by hybridization (Supplemental Table 2). The number of BAC clones identified by the extended analysis were then combined with the hybridization data and used to estimate the genome coverage of each of the eight BAC libraries and the results are shown in Table 3A. The hybridization/BES/FPC analysis revealed that all eight libraries covered their corresponding genomes by at least 10-fold (Table 3A).

For the four remaining BAC libraries, clones that hybridized to the 12-locus probe set were picked, end sequenced, fingerprinted, assembled into contigs individually, and analyzed as above. Results were obtained similar to those using the whole genome FPC assemblies for the *O. officinalis* [CC], *O. alta* [CCDD], and *O. ridleyi* [HHJJ] libraries, with coverages ranging between 10- and 14-fold (Table 3B, Supplemental Table 2). However, analysis of the *O. granulata* [GG] library resulted in only 6.3-fold genome coverage, 42% lower than mathematically predicted.

Repeat content estimates from pilot BAC end sequences

To obtain a preliminary view of the major repetitive element content of the 12 *Oryza* species under investigation, we generated nearly 6.7 Mb of sequence from 623 to 3658 BAC ends from each library. These sequences represent a total of 60 to 862 kb and approximately 0.01% to 0.1% of each of the *Oryza* genomes (Table 4). The TIGR and University of Georgia (UGA) (Jiang and Wessler 2001) *O. sativa* (Nipponbare) repeat databases (http://www.tigr.org/tdb/e2k1/plant.repeats/) were combined and utilized for repeat detection using RepeatMasker (http://www.repeatmasker.org/). The UGA database was then used to estimate the fraction of interspersed repeats belonging to five broad repeat categories: LTR-retrotransposons, LINEs, SINEs, DNA elements, and unclassified (Table 4). Sixteen percent to 49% of sequence generated from each species was detected as repetitive by RepeatMasker using the combined databases, where LTR-retrotransposons were the predominate class for every species. If *O. coarctata* [HHKK] is excluded, because its genome size is unknown, then a roughly linear relationship between genome size and repeat content is observed, with *O. brachyantha* [FF] having the lowest LTR retrotransposon content and *O. australiensis* [EE] the highest.

Discussion

New and confirmed genome size data for nine *Oryza* species

Accurate genome size data is a critical basis for the development of whole genome analysis platforms. The *Oryza* BAC library resource project began using genome size data summarized in the RBG Kew Gardens Angiosperm DNA C-value data base and the Martinez et al. (1994) and Uozu et al. (1997) publications. We observed inconsistencies between studies that used different accessions and methods. The most noticeable were for the following species: *O. rufipogon* [AA], *O. glaberrima* [AA], *O. officinalis* [CC], *O. brachyantha* [FF], and *O. ridleyi* [HHJJ], where both Iyengar and Sen (1978) and flow cytometry data were available (Table 1).

Table 3A. Genome coverage estimations for eight *Oryza* species based on hybridization and extended analysis utilizing whole genome FPC physical maps and BAC end sequences

	Genome									
	AA			BB	BBCC		EE	FF	HHKK	
	O. nivara	*O. rufipogon*	*O. glaberrima*	*O. punctata*	*O. minuta*		*O. australiensis*	*O. brachyantha*	*O. coarctata*	
Number of hits	255	153	149	122	153	78	77	128	147	79
Number of loci	12	12	12	12	12	9	7	12	12	7
HX coverage[a]	21.3	12.8	12.4	10.2	12.8	8.7	11	10.7	12.3	11.3
					11.0[b]				11.9[b]	
FPC X coverage[c]	17.5	9.7	12.2	11	8.6		9.6	9.1	8.7	

[a]Genome coverage based on the total number of hybridization and BES hits identified by extended analysis divided by the total number of loci per genome in the diploids or subgenome in the tetraploids (dispersed clones and undetected homeologous contigs [see Supplemental Table 2] were not taken into account for estimating genome coverage).
[b]Average HX coverage of both subgenomes for each tetraploid species (see Supplemental Table 2 for details).
[c]Calculated coverage of the FPC physical maps (excluding singletons).

Our genome size measurements were found to be within a 7% range of flow cytometry data previously reported for *O. rufipogon*, *O. officinalis*, *O. australiensis*, and *O. brachyantha* compared either to Uozu et al. (1997) or Martinez et al (1994). However, with *O. ridleyi* [HHJJ], our genome size data was 64% higher than previously reported even though the same accession was used (Martinez et al. 1994).

No flow cytometry data were available for *O. nivara* [AA], and its genome size was estimated by Iyengar and Sen (1978) to be 760 Mb, almost twice that of cultivated rice. We measured the *O. nivara* genome size to be 448 Mb, which is much closer to the other AA genome diploids *O. sativa* and *O. rufipogon*. One possible explanation to account for the large differences in genome size estimations between Iyengar and Sen (1978) and the other flow cytometric data reported here and elsewhere is that the 1C values reported by Iyengar and Sen (1978) for 5 of 10 species (i.e., *O. nivara*, *O. rufipogon*, *O. glaberrima*, *O. officinalis*, and *O. ridleyi*) were actually 2C values (Table 1). If this were the case, then all of the genome size data reported by Iyengar and Sen (1978), except for *O. ridleyi*, would fall within 21% of the data measured by flow cytometry.

The discrepancy between genome size values measured by flow cytometry for *O. ridleyi* may be explained by the use of contaminated or heterozygous germplasm in the Martinez et al. (1994) study. The accessions used for the *Oryza* BAC library project were genetically homozygous and have been extensively used in breeding programs as donors for important agronomical traits.

BAC library coverage estimations

For a BAC library to be useful for positional cloning, physical mapping, and genome sequencing, it must have a minimum of 5–10 × coverage across the entire genome. Genome coverage for the *Oryza* BAC library resource was determined mathematically and by hybridization/BES/FPC analysis, and in all but one case (*O. granulata*), both measurements resulted in a minimum of 10-fold redundancy. For the majority of libraries, the extended analysis resulted in lower genome coverages primarily because not all of the clones in a given contig could be identified by hybridization or BES analysis. We suspect that some of the clones that were not identified by hybridization, including the clones identified by BES alone, were undetected due to technical issues associated with colony blot hybridization. These include the use of locus-specific probes from a single species [AA] to hybridize to distantly related species, uniform hybridization and washing conditions across all libraries, and decreasing filter quality due to multiple hybridizations. For clones that were identified by BESs alone, it is possible that they are false positives and were derived from paralogous sequence duplications in the genome. This is unlikely, however, as we only analyzed BESs from clones in contigs identified by hybridization. The issues raised above may be particularly relevant for analysis of the *O. granulata* [GG] library, which is the most basal of the *Oryza* species, and was the only library that showed less than 10-fold genome coverage by hybridization/contig analysis even though it was predicted to contain 10.8 genome equivalents.

We were unable to detect robust contigs for 19 out of 216 predicted contigs, assuming the syntenic relationships between these species and the reference *japonica* genome were maintained throughout evolution (Supplemental Table 2). The majority (13) of the "missing" contigs were from the four *Oryza* polyploid libraries. For

Table 3B. Genome coverage estimations for four *Oryza* species based on hybridization and contig analysis (see Methods for details)

	Genome					
	CC	CCDD		GG	HHJJ	
	O. officinalis	*O. alta*		*O. granulata*	*O. ridleyi*	
Number of hits	168	116	90	76	139	86
Number of loci	12	11	9	12	12	10
HX coverage[a]	14	10.5	10	6.3	11.6	8.6
		10.3[b]			10.2[b]	
BAC library coverage[c]	18.6	12.1		10.8	12.5	

[a]HX coverage and [b]average HX coverage are as described for Table 3A, except that these values are obtained from specific FPC and BES assemblies as small projects (see Methods for details).
[c]Calculated genome coverage estimations from Table 2.

Table 4. Analysis of repetitive sequences from pilot BAC end sequences of *Oryza* BAC libraries

Oryza species	Genome		# reads	GC %	Total detected			ISR[b]	% of interspersed					Low complexity		Simple	
	Type	Size (Mb)			Length (kb)[a]	kb	%	kb	LTR	LINE	SINE	DNA	UNC	kb	%	kb	%
glaberrima	AA	357	2304	41.7	1362	396	29	215	77.0	3.9	0.7	17.8	0.6	5.1	0.4	4.8	0.4
rufipogon	AA	439	2249	42.1	1571	574	37	285	78.6	2.2	0.9	17.6	0.7	5.7	0.4	5.4	0.3
nivara	AA	448	2213	42.8	1458	519	36	236	78.9	3.1	1.1	16.5	0.4	5.1	0.4	5.5	0.4
punctata	BB	425	1503	41.8	1062	385	36	243	83.2	1.1	0.2	15.5	0.0	3.7	0.4	2.7	0.3
officinalis	CC	651	2910	42.9	2070	859	41	425	77.3	1.1	0.3	21.1	0.2	7.3	0.4	5.8	0.3
minuta	BBCC	1124	3212	42.5	2243	859	38	469	74.6	0.9	0.3	24.1	0.1	7.9	0.4	3.7	0.2
alta	CCDD	1008	2196	43.0	1274	485	38	201	79.3	2.4	0.1	18.0	0.1	3.8	0.3	3.9	0.3
australiensis	EE	965	2254	44.4	1572	768	49	354	80.9	0.8	0.0	18.2	0.0	5.4	0.4	4.0	0.3
brachyantha	FF	362	623	40.9	357	60	17	20	85.4	3.3	0.0	11.2	0.0	2.3	0.6	2.4	0.7
granulata	GG	882	2105	45.6	1452	507	35	203	87.6	1.7	0.1	10.5	0.0	4.5	0.3	2.6	0.2
ridleyi	HHJJ	1283	3658	43.4	2579	862	33	503	69.9	0.2	0.1	29.7	0.1	8.9	0.4	5.7	0.2
coarctata	HHKK	ND[c]	3764	40.3	2577	412	16	233	82.6	2.2	0.2	14.7	0.2	18.3	0.7	4.6	0.2

[a]Length excluding Ns.
[b]Interspersed repeats and includes LTR elements, LINE elements, SINE elements, DNA transposons, UNC (unclassified) and simple sequence repeats.
[c]Not determined.

the remaining six cases, BAC clones were identified by hybridization but could not assemble into contigs and were thus classified as "dispersed" (Supplemental Table 2). For *O. minuta* [BBCC], 9 of 12, *O. alta* [CCDD], 9 of 11 (1 locus was dispersed), *O. coarctata* [HHKK], 7 of 12, and *O. ridleyi* [HHJJ], 10 of 12 probes identified clones that assembled into two contigs (Table 3A,B; Supplemental Table 2). Although further work is required to elucidate if these duplicate contigs are derived from orthologous positions on each genome type, it is not unexpected that all loci were not represented twice per polyploid genome. Several studies have demonstrated that rapid gene loss and genomic rearrangements are a consequence of polyploidization (Ozkan et al. 2001; Shaked et al. 2001). For the purposes of determining genome coverage, duplicate contigs were treated as independent loci.

Regarding dispersed loci, five of the six were identified from the *O. australiensis* [EE] library. This observation may be indicative of large genome rearrangements in the EE genome and corresponds well with the EE genome being the largest of all the diploids (Table 1) and the most highly repetitive of all the *Oryza* species (Uozu et al. 1997; Table 4). Preliminary analysis of BAC end sequences of the clones identified in these dispersed loci show that the majority share significant sequence similarity with a number of different classes of transposable elements (data not shown), suggesting these loci may be located in repetitive regions of the EE genome.

Differentiation of colinear and homeologous BACs in the tetraploids: Opportunities to reconstitute the genomes of extinct diploid counterparts

Fingerprinting methods have recently been used to dissect the subgenomes of tetraploids (Cenci et al. 2003). However such differentiation depends on the extent of sequence divergence of the two diploid counterparts in the tetraploid species (Cenci et al. 2003). Recently created polyploids like wheat exhibit very little intraspecific genetic variation due to genetic bottlenecks imposed during polyploidization. However, all the polyploids in the genus *Oryza* are either highly polymorphic or exhibit at least the same level of genetic variation as the diploids. For these reasons the polyploids are considered as older or ancient (Jena and Kochert 1991; Wang et al. 1992; Ge et al. 1999).

Although diploid counterparts for the BBCC tetraploid exist, living ancestor diploid species for the DD, HH, JJ, and KK genomes have not been identified and are presumed extinct. The differentiation of both subgenomes in the tetraploid libraries of *O. alta* [CCDD], *O. ridleyi* [HHJJ], and *O. coarctata* [HHKK] by fingerprinting/BES methods offers a unique opportunity to reconstitute these genomes and develop genome-wide physical maps for these genomes.

A preliminary survey of repeat content from *Oryza* species and their correlation with respective genome sizes

Possible mechanisms for the genome size variation among the *Oryza* specices include insertion and deletion of a variety of DNA sequences (SanMiguel and Bennetzen 1998; Devos et al. 2002; Feng et al. 2002; Han and Xue 2003; Edwards et al. 2004; Feltus et al. 2004; Ma and Bennetzen 2004). Although insertions have been largely attributed to amplifications of retrotransoposons (Devos et al. 2002; Ma and Bennetzen 2004; Ma et al. 2004), as well as genome-specific unique sequences (Zhao et al. 1989; Uozu et al. 1997), deletions include all classes of DNA sequences through homologous recombination and illegitimate recombination (Ma and Bennetzen 2004).

Genome-wide BAC end sequences in combination with physical maps are important resources for gaining insights regarding genome sequence composition and organization (Mao et al. 2000; Messing et al. 2004). To explore the possible relationship between repeat elements and genome sizes among the *Oryza* species, we estimated the repeat content from BAC end sequences from the *Oryza* BAC libraries. Repeat databases derived from the *O. sativa* genome sequence successfully detected repeats in all 12 rice species considered here.

LTR-retrotransposons frequently dominate plant genomes. In this study, the largest, *O. australiensis* [EE], and smallest genome sizes, *O. brachyantha* [FF], excluding *O. coarctata* [HHKK], correlated with the abundance of LTR retrotransposons. These results are in agreement with Uozu et al. (1997), who demonstrated good correlation between genome size of *O. australiensis* and *O. brachyantha* with overall chromosome size and morphology. Both metaphase and prometaphase chromosomes of *O. australiensis* were much larger than those of any other diploid *Oryza* species with a high degree of heterochromatin condensation, whereas *O. brachyantha* chromosomes showed the opposite pattern.

We are further exploring the causes for this dynamic variation in the sizes of nuclear genomes by sequencing an orthologous region on chromosome 11 across all the genomes of the *Oryza*. In combination with a well-defined phylogeny, studies with this new BAC library resource will add directionality to the analysis of genome size evolution in the genus *Oryza* and may answer questions regarding mechanisms involved in such events.

Utilization of the *Oryza* BAC library resource

The *Oryza* BAC library resource is the first description of a comprehensive collection of libraries that represent all the genome types of an entire genus. To add additional value to these libraries, we have already generated BAC end sequence and fingerprint databases for eight of the 12 libraries and expect to have similar data for the remaining four libraries in public databases by the end of 2005 (OMAP Consortia, unpubl.). This library resource is publicly available in the form of whole libraries, filters, and individual clones, through our BAC/EST Resource Center (http://www.genome.arizona.edu/orders) and has already been extensively used worldwide for the analysis of genome evolution and organization, positional cloning, and gap closure in the *japonica* reference sequence.

For example, an emerging picture in rice evolution is that the genomes of Asian rice (*O. sativa* ssp. *indica* and *japonica*) have undergone rapid genome expansion in comparison to *O. glaberrima*, which diverged from a common ancestor around 0.64 MYA (Ma and Bennetzen 2004). However, no information is available regarding evolutionary trends relative to immediate ancestors of Asian cultivated rice, *O. nivara* and *O. rufipogon*, as well as the other nine genome types of the genus *Oryza*. To obtain a broader understanding of *Oryza* genome evolution and the consequences of domestication, we and others are using the *Oryza* BAC library resource to investigate key loci and whole chromosomes across all genomes by comparative physical mapping and genome sequencing. To illustrate, we utilized the *O. nivara* BAC library and end sequence and fingerprint databases to reconstruct *O. nivara* chromosome 3 with only 16 small gaps. Detailed comparative analysis showed that *O. sativa* ssp. *japonica* rice chromosome 3 is about 20% larger than its progenitor *O. nivara* chromosome 3, thereby supporting and extending the concept of rapid genome expansion in cultivated rice (Rice Chromosome 3 Sequencing Consortium 2005).

To further explore genome expansion relative to the other AA genomes and *O. punctata* [BB], we utilized the extended analysis data generated in this study for the *Adh1* gene, which is a standard locus that has been used to study genome evolution across the plant kingdom. We measured the distances between paired BAC ends mapped on to the reference *O. sativa* genome and compared these distances with BAC clone insert sizes. The results indicated that the orthologous region in the reference *O. sativa* genome is larger by 50 kb (28%), 19.1 kb (11.3%), 35.1 kb (14.8%), and 28.2 kb (9.4%) relative to *O. punctata*, *O. glaberrima*, *O. rufipogon*, and *O. nivara*, respectively (Supplemental Table 3). Analysis of large and contiguous sequences generated from orthologous *Adh1* regions from these species indicate that this dynamic variation is not only highlighted by insertion of transposable elements, but involves multiple genetic mechanisms (J. Ammiraju, Y. Yu, R.T. Mueller, J. Currie, H.R. Kim, J.L. Goicoechea, and R.A. Wing, unpubl.).

In summary, this comparative structural analysis provides a previously unavailable glimpse through the window of rice evolution and confirms that the rice genome has undergone rapid changes after divergence from progenitors.

Methods

Plant material

Young leaf tissue was collected from clonally propagated single plants at IRRI from *O. brachyantha* (Acc. 101232), *O. alta* (Acc. 105143), *O. officinalis* (Acc. 100896), *O. ridleyi* (Acc. 100821), *O. punctata* (Acc. 105690), *O. coarctata* (Acc. 104502), *O. minuta* (Acc. 101141), and *O. granulata* (Acc. 102118). For *O. glaberrima* variety CG14 (Acc. 96717), *O. rufipogon* perennial type (Acc. 105491), *O. nivara* (Acc. W0106), and *O. australiensis* (W0008), tissue samples were obtained from inbred seedling material propagated at IRRI, Cornell, and NIG, respectively.

Genome size determination by flow cytometry

Samples for flow cytometric analysis were prepared from seedling tissue as described by Arumuganathan and Earle (1991a,b) and Galbraith et al. (1983). Three to 5 measurements, on a minimum of 2000 nuclei per analysis, were made on two separate days with fresh preparations made each day. Cell clumps and debris were excluded from analysis by using red fluorescence and forward angle light scatter gates. Chicken red blood cells (3.0 pg/nucleus), *Nicotiana tobacum* var. Xanthi (11 pg/ 2C nucleus), *A. thaliana* ecotype Columbia (0.47 pg/2C nucleus), and *Oryza sativa* ssp. *japonica* cv Nipponbare (0.91 pg/2C) were used as internal standards. Values for nuclear DNA content were estimated by a comparison of nuclear peaks from the *Oryza* species on the linear scale, with the peak for chicken red blood cells (CRBC) included as an internal standard in each run. The conversion factor for picograms to base pairs is 1 pg = 0.965×10^9 bp (Bennett et al. 2000).

BAC library construction

All protocols used for megabase-size DNA preparation, library construction, picking, and arraying were as previously described (Luo and Wing 2003; Kudrna and Wing 2004) except the following: (1) To reduce organelle contamination in the nuclei preparations, nuclei isolation buffer containing 0.5% TritonX-100 was used during the nuclei washing steps (Georgi et al. 2002); (2) all libraries were constructed in the HindIII site of the vector pIndigoBAC536 SwaI. This vector is identical to pIndigoBAC536 (H. Shizuya et al. unpubl.) except for the addition of two SwaI sites near and internal to two NotI sites that flank the *LacZ* gene (M. Luo, A. Jetty, and R.A. Wing, unpubl.); (3) all ligations were transformed into DH10B T1 phage resistant *E. coli* cells (Invitrogen).

Insert size analysis

BAC plasmid DNA was isolated from randomly picked clones from each *Oryza* library, in a 96-well format, using a simplified high throughput method (H.R. Kim and R.A. Wing, unpubl.) that is based on conventional alkaline lysis methods (Sambrook and Russell 2001). BAC DNA (~500 ng) was digested with NotI and resolved on CHEF (Bio-Rad) gels as previously described (Luo and Wing 2003).

BAC library screening

High density colony filters for each library were prepared using a Genetix Q-bot (Genetix). Each 22.5 × 22.5 cm filter (Hybond-N+: Amersham) contained 18,432 independent clones arrayed in a 4 × 4 double spotted pattern. All hybridizations followed Chen et al. (2000), and the addresses of BAC clones that hybridized with specific probes were recorded and input as "markers" into FPC (Soderlund et al. 2000).

Organellar DNA content estimation

To estimate the percentage of chloroplast and mitochondrial DNA content in each library, one high-density filter from each library was screened with a pool of three barley chloroplast probes, *ndhA*, *rbcL*, and *psbA* (obtained from J. Mullet, Texas A&M University), and with a pool of four rice mitochondrial probes, *atpA*, *cob*, *atp9*, and *coxE* (obtained from T. Sasaki, MAFF, Japan) separately.

Probes for BAC library nuclear genome coverage estimation

Gene-specific probes for *Hd1* (Yano et al. 2000) and *Adh1* (Tarchini et al. 2000) were PCR amplified from Nipponbare genomic DNA, using the primers Hd1F 5′-TTCTCCTCTCCAAAGA TTCC-3′ and Hd1R 5′-GCTTTTGTTTGGAGAATGTT-3′ and Adh1F 5′-GGAAGCCCATTTACCATTT-3′ and Adh1R 5′-GCCC AGGATACACAGAAGA-3′, respectively, and gel purified. Rice cDNA R2277 (Li and Gill, 2002) was obtained from B. Gill, Kansas State University. These probes map to chromosomes 6, 11, and 1 (Table 4). cDNA RFLP markers that map to the remaining nine rice chromosomes were obtained from S. McCouch, Cornell University (Supplemental Table 2). Inserts were gel purified using a QIAEX II (Qiagen) kit and labeled with α^{32}P dCTP using a decaprimeII random prime labeling Kit (Ambion).

BAC end sequencing and repeat analysis of the *Oryza* species

BAC ends were sequenced using BigDye v3.1 (Applied Biosystems) with T7 (5′-TAATACGACTCACTATAGGG-3′) and BES_HR primers (5′-CACTCATTAGGCACCCCA-3′). Cycle sequencing was performed using the following conditions: 150 cycles of 10 sec at 95°C, 5 sec at 55°C, and 2.5 min at 60°C, followed by DNA purification using CleanSeq (Agencourt). Samples were eluted into 20 µL of water and separated on ABI 3730xl DNA sequencers. Sequence data were collected and extracted using ABI sequence analysis software. Phred software (Ewing and Green 1998; Ewing et al. 1998) was used for base calling, and vector and low quality sequences were removed using the program Lucy (Chou and Holmes 2001). All sequences were submitted to the GSS section of GenBank.

Repeat analysis was undertaken using "RepeatMasker" version 3.0.5 (http://www.repeatmasker.org/). The program was run in "sensitive mode" and using cross_match version 0.990329 as the search engine and a custom repeat library composed of both the TIGR *Oryza* Repeat Database (http://www.tigr.org/tdb/e2k1/plant.repeats/) and a database for transposable elements from Jiang and Wessler 2001.

FPC/BES contig assembly and analysis to estimate genome coverage of the *Oryza* BAC libraries

Genome coverage estimates utilized (1) hybridization data from the 12 chromosome specific probes, (2) BAC end sequence data from the positively hybridizing clones, and (3) fingerprint/contig data either from existing whole genome FPC assemblies (extended analysis) derived from the *Oryza* Map Alignment Project (http://www.omap.org) or specific FPC assemblies from only the clones that hybridized with a given probe (small project).

Extended analysis

This strategy was used for the species with high coverage FPC/BES phase I physical maps (*O. australiensis* [EE] [63,368 clones], *O. brachyantha* [FF] [25,216 clones], *O. glaberrima* [AA] [33,065 clones], *O. nivara* [AA] [51,056 clones], *O. punctata* [BB] [34,224 clones], *O. rufipogon* [AA] [33,023 clones], *O. minuta* [BBCC] [83,592 clones], and *O. coarctata* [HHKK] [50,146 clones]). First, an incremental FPC build was constructed by implementing the CpM (Clone plus marker) function on phase I physical maps as described above at a $1e^{-50}$ cutoff. End merges of contigs were then performed at a cutoff of $1e^{-21}$–$1e^{-18}$. Blast analysis was carried out in parallel for all the BAC end sequences from the positive hybridization hits against *O. sativa* pseudo-molecules representing the 12 chromosome of rice (GenBank accession numbers AP008207–AP008218). Alignments larger than 100 bp and that map to an interval of 200 kb flanking the position of the marker in reference genome, *O. sativa* ssp. *japonica*, were further included in the analysis. A contig was considered positive when a majority of the clones in it were hit by both hybridization and BES analysis. Blast analysis of BES from the clones that were mapped within a 50-CB (metric of FPC) unit interval flanking the position of the marker in the "positive contig" was also carried out against the *O. sativa* pseudomolecules, to identify positive clones that were not identified by hybridization.

Small projects

For those libraries without FPC/BES physical maps (*O. officinalis* [CC], *O. granulata* [GG], *O. ridleyi* [HHJJ], and *O. alta* [CCDD]) positive clones from hybridizations were fingerprinted and end sequenced. Fingerprints were generated using a modified SNaPshot fingerprinting method (Luo et al. 2003; H.R. Kim and R.A. Wing, unpubl.). Trace files were processed with GeneMapper v. 3.0 (ABI) to generate size files that were assembled with FPC (Soderlund et al. 2000) projects for every marker tested per species. These projects were initially assembled very stringently. The cutoff values were then gradually reduced until clones began to form into contigs. At that particular cutoff, singletons were incorporated in a new contig. End-to-end merges and reanalysis of the resulting contigs were then performed in cycles, until all the clones were added. The initial and final cutoff values of these analyses were chosen based on the number of clones involved in the analysis and the nature of the species (Soderlund et al. 2000).

GenBank accession numbers of BAC end sequences

CL610447–CL612660 (*O. nivara*); CL792274–CL794523 (*O. rufipogon*); CW623334–CW624836 (*O. punctata*); CZ157233–CZ160142 (*O. officinalis*); CZ027313–CZ030524 (*O. minuta*); CZ115907–CZ118102 (*O. alta*); CL903491–CL905744 (*O. australiensis*); CL553094–CL553716 (*O. brachyantha*); CZ155128–CZ157232 (*O. granulata*), CZ160143–CZ163800 (*O. ridleyi*), CZ163801–CZ167564 (*O. coarctata*); CW652102–CW654406, CW662310–CW662313 (*O. glaberrima*).

Acknowledgments

We thank Olin Feuerbacher, Samina Makda, Miriam Eaton, Elena Ruiz, Noreen Lyle, Angelina Angelova, Diana Stum, Elizabeth Ashley, Marina Wissotski, and Danielle Yost for technical assistance. The work was funded by grants from the National Science Foundation (R.A.W. and S.J.: IOB-0208329; R.A.W., S.J., and P.S.M: DBI-0321678).

References

Aggarwal, R.K., Brar, D.S., and Khush, G.S. 1997. Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. *Mol. Gen. Genet.* **254:** 1–12.

Arumuganathan, K. and Earle, E.D. 1991a. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Reporter* **9:** 208–218.

———. 1991b. Estimation of nuclear DNA content of plants by flow cytometry. *Plant Mol. Biol. Reporter* **9:** 229–233.

Bennett, M.D., Bhandol, P., and Leitch, I.J. 2000. Nuclear DNA amounts

in angiosperms and their modern uses—807 new estimates. *Ann. Bot.* **86:** 859–909.
Brar, D.S. and Khush, G.S. 1997. Alien introgression in rice. *Plant Mol. Biol.* **35:** 35–47.
Cenci, A., Chantret, N., Kong, X., Gu, Y., Anderson, O.D., Fahima, T., Distelfeld, A., and Dubcovsky, J. 2003. Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. durum). *Theor. Appl. Genet.* **107:** 931–939.
Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., et al. 2000. An integrated physical and genetic map of the rice genome. *Plant Cell* **14:** 537–545.
Chou, H.H. and Holmes, M.H. 2001. DNA sequence quality trimming and vector removal. *Bioinformatics* **17:** 1093–1094.
Devos, K.M., Brown, J.K., and Bennetzen, J.L. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12:** 1075–1079.
Edwards, J.D., Lee, V.M., and McCouch, S.R. 2004. Sources and predictors of resolvable indel polymorphism assessed using rice as a model. *Mol. Genet. Genomics* **271:** 298–307.
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., Wendle, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.
Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N., and Paterson, A.H. 2004. An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.* **14:** 1812–1819.
Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J., Liu, Y., Hu, X., et al. 2002. Sequence and analysis of rice chromosome 4. *Nature* **420:** 316–320.
Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., and Firoozabady, E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220:** 1049–1051.
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.* **96:** 14400–14405.
Georgi, L.L., Wang, Y., Yvergniaux, D., Ormsbee, T., Inigo, M., Reighard, G., and Abbott, A.G. 2002. Construction of a BAC library and its application to the identification of simple sequence repeats in peach [*Prunus persica* (L.) Batsch] *Theor. Appl. Genet.* **105:** 1151–1158.
Han, B. and Xue, Y. 2003. Genome wide intraspecific DNA-sequence variations in rice. *Curr. Opin. Plant Mol. Biol.* **6:** 134–138.
Hass, B.L., Pires, J.C., Porter, R., Phillips, R.L., and Jackson, S.A. 2003. Comparative genetics at the gene and chromosome levels between rice (*Oryza sativa*) and wildrice (*Zizania palustris*). *Theor. Appl. Genet.* **107:** 773–782.
Huang, H. and Kochert, G. 1994. Comparative RFLP mapping of an allotetraploid wild rice species (*Oryza latifolia*) and cultivated rice (*O. sativa*). *Plant Mol. Biol.* **25:** 633–648.
International Rice Genome Sequencing Project. 2005. The map based sequencing of the rice genome. *Nature* **436:** 793–800.
Iyengar, G.A.S. and Sen, S.K. 1978. Nuclear DNA content of several wild and cultivated *Oryza* species. *Env. Exp. Bot.* **18:** 219–224.
Jena, K.K. and Kochert, G. 1991. Restriction fragment length polymorphism analysis of CCDD genome species of the genus *Oryza* L. *Plant. Mol. Biol.* **16:** 831–839.
Jena, K.K., Kush, G.S., and Kochert, G. 1994. Comparative RFLP mapping of a wild rice, *Oryza officinalis*, and cultivated rice, *O. sativa*. *Genome* **37:** 382–389.
Jiang, N. and Wessler, S.R. 2001. Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell* **13:** 2553–2564.
Khush, G.S. 1997. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35:** 25–34.
Kudrna, D.A. and Wing, R.A. 2004. Genetic conservation of genomic resources. In *Encyclopedia of plant and crop sciences* (ed. R.M. Goodman), pp. 1–5. Dekker Publishers, New York.
Li, W. and Gill, B.S. 2002. The colinearity of the Sh2/A1 orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the triticeae. *Genetics* **160:** 1153–1162.
Luo, M. and Wing, R.A. 2003. An improved method for plant BAC library construction. In *Plant functional genomics* (ed. E. Grotewold), pp. 3–20. Human Press Inc., Totowa, NJ.
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.* **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.
Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., et al. 2000. Rice transposable elements: A survey of 73,000 sequence-tagged-connectors. *Genome Res.* **10:** 982–990.
Martinez, C.P., Arumuganathan, K., Kikuchi, H., and Earle, E.D. 1994. Nuclear DNA content of ten rice species as determined by flow cytometry. *Jpn. J. Genet.* **69:** 513–523.
Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F., Fuks, G., Soderlund, C.A., Mayer, K.F., et al. 2004. Sequence composition and genome organization of maize. *Proc. Natl. Acad. Sci.* **101:** 14349–14354.
Ozkan, H., Levy, A.A., and Feldman, M. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13:** 1735–1747.
Rice Chromosome 3 Sequencing Consortium. 2005. Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. *Genome Res.* **15:** 1284–1291.
Sambrook, J. and Russell, D.W. 2001. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
SanMiguel, P. and Bennetzen, J.L. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **82:** 37–44.
Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A.A. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13:** 1749–1759.
Soderlund, C., Humphray, S., Dunham, A., and French, L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.* **10:** 1772–1787.
Tarchini, R., Biddle, P., Wineland, R., Tingey, S., and Rafalski, A. 2000. The complete sequence of 340 kb of DNA around the rice Adh1-adh2 region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* **12:** 381–391.
Uozu, S., Ikehashi, H., Ohmido, N., Ohtsubo H., Ohtsubo E., and Fukui, K. 1997. Repetitive sequences: Cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol. Biol.* **35:** 791–799.
Wang, Z.Y., Second, G., and Tanksley, S.D. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor. Appl. Genet.* **83:** 565–581.
Wing, R.A., Ammiraju, J.S.S., Luo, M., Kim, H.R., Yu, Y., Kudrna, D., Goicoechea, J., Wang, W., Nelson, W., Soderlund, C., et al. 2005. The *Oryza* Map Alignment Project: The golden path to unlocking the genetic potential of wild rice species. *Plant Mol. Bio.* **59:** 53–62.
Xiao, J.H., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D., Li, J.M., and Yuan, L.P. 1996. Genes from wild rice improve yield. *Nature* **384:** 223–224.
Xiao, J., Li, J., Grandillo, S., Ahn, S.N., Yuan, L., Tanksley, S.D., and McCouch, S.R. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* **150:** 899–909.
Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., et al. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* **12:** 2473–2484.
Zhao, X., Wu, T., Xie, Y., and Wu, R. 1989. Genome specific repetitive sequences in the genus *Oryza*. *Theor. Appl. Genet.* **78:** 201–209.

Web site references

http://www.genome.arizona.edu; Arizona Genomics Institute.
http://www.omap.org; *Oryza* Map Alignment Project.
http://www.rbgkew.org.uk/cval/homepage.html; Royal Botanical Gardens, plant DNA c-values database.
http://www.repeatmasker.org/; RepeatMasker software.
http://www.tigr.org/tdb/e2k1/plant.repeats/; The Institute of Genome Research plant repeat database.

Received January 31, 2005; accepted in revised form September 22, 2005.