

The Complete Sequence of 340 kb of DNA around the Rice *Adh1-Adh2* Region Reveals Interrupted Colinearity with Maize Chromosome 4

Renato Tarchini,^{a,1} Phyllis Biddle,^a Robin Wineland,^b Scott Tingey,^a and Antoni Rafalski^{a,2}

^a DuPont Agricultural Biotechnology–Genomics, Delaware Technology Park, Suite 200, 1 Innovation Way, P.O. Box 6104, Newark, Delaware 19714-6104

^b Pioneer Hi-Bred International, Inc., 7300 N.W. 62nd Avenue, P.O. Box 1004, Johnston, Iowa 50131-1004

A 2.3-centimorgan (cM) segment of rice chromosome 11 consisting of 340 kb of DNA sequence around the alcohol dehydrogenase *Adh1* and *Adh2* loci was completely sequenced, revealing the presence of 33 putative genes, including several apparently involved in disease resistance. Fourteen of the genes were confirmed by identifying the corresponding transcripts. Five genes, spanning 1.9 cM of the region, cross-hybridized with maize genomic DNA and were genetically mapped in maize, revealing a stretch of colinearity with maize chromosome 4. The *Adh1* gene marked one significant interruption. This gene mapped to maize chromosome 1, indicating a possible translocation of *Adh1* after the evolutionary divergence leading to maize and sorghum. Several other genes, most notably genes similar to known disease resistance genes, showed no cross-hybridization with maize genomic DNA, suggesting sequence divergence or absence of these sequences in maize, which is in contrast to several other well-conserved genes, including *Adh1* and *Adh2*. These findings indicate that the use of rice as the model system for other cereals may sometimes be complicated by the presence of rapidly evolving gene families and microtranslocations. Seven retrotransposons and eight transposons were identified in this rice segment, including a *Tc1/Mariner*-like element, which is new to rice. In contrast to maize, retroelements are less frequent in rice. Only 14.4% of this genome segment consist of retroelements. Miniature inverted repeat transposable elements were found to be the most frequently occurring class of repetitive elements, accounting for 18.8% of the total repetitive DNA.

INTRODUCTION

Rice has become the model system for the molecular biology of grasses because its genome is amenable to analysis. The rice genome is relatively small (450 Mb; Arumuganathan and Earle, 1991), but its gene content is comparable to that of other grasses (Ahn and Tanksley, 1993). More than 45,000 rice expressed sequence tags (ESTs) are publicly available, and a dense EST-based restriction fragment length polymorphism map (Harushima et al., 1998) has been established and connected to a physical map (Sasaki, 1998). One of the objectives of our research reported here was to determine the sequence composition of the rice genome on a sizable contiguous sample of DNA; we also wanted to evaluate how its properties might affect a large-scale DNA-sequencing assembly process. The content, composition, and organization of the repetitive DNA fraction determine the assembly success of a shotgun DNA-

sequencing process of rice bacterial artificial chromosomes (BACs).

To what extent the information coming from the sequencing of the entire rice genome can be applied to other Gramineae species remains to be established. Comparative mapping has revealed the existence of a high degree of conservation in gene repertoire and order among the grass genomes (Ahn and Tanksley, 1993; Devos et al., 1994; Kurata et al., 1994; Gale and Devos, 1998). This has allowed the construction of a general grass consensus map (Moore et al., 1995; Gale and Devos, 1998) in which different grass genomes are described in terms of ancestral "rice linkage blocks." Bennetzen and co-workers have shown that microscale colinearity can be found in homologous regions of closely related species such as maize, sorghum, and rice (M. Chen et al., 1997; Bennetzen et al., 1998). However, exceptions to these findings have also been described (Bennetzen et al., 1998; Tikhonov et al., 1999). In particular, whereas the regions surrounding the alcohol dehydrogenase *Adh1* genes of maize and sorghum were found to be colinear, the corresponding regions of rice showed a lack of conservation (Tikhonov et al., 1999).

¹ Current address: Plant Research International, P.O. Box 16 6700 AA, Wageningen, The Netherlands.

² To whom correspondence should be addressed. E-mail j-antoni.rafalski@usa.dupont.com; fax 302-631-2607.

Our goal was to establish whether use of the rice genome would be feasible as a surrogate genome for map-based cloning of maize genes. To use rice as the surrogate genome, one would first genetically map a trait in maize and then use flanking maize markers to isolate the corresponding genetic interval in rice. Candidate genes would be identified from the rice BAC contig, and corresponding genes in maize would be isolated and analyzed. A similar approach was used by Kilian et al. (1995) for barley. Complete sequencing of a several hundred kilobase interval of the rice genome and investigation of the syntenous region in maize allowed us to predict some of the difficulties that could be encountered in using this approach. The basis for the reported lack of sequence conservation between the regions surrounding the *Adh1* genes of maize and rice was investigated further, and the composition of transposable elements in this genome segment was established.

RESULTS

Physical Mapping of a BAC Contig in the *Adh1* Region

A set of minimally overlapping clones consisting of BAC85C11, BAC178G5, and BAC62F3 covering the *Adh1* region was assembled by hybridization with the *Adh1* probe; the set was extended by polymerase chain reaction (PCR) identification of additional BACs, using primers corresponding to the ends of the *Adh1*-containing BACs (Figure 1B). Additional BAC sub-clones were used to complete the contig (see Methods).

Sequence of the *Adh1* Region

The sequence of this region consists of 339,485 bp of DNA and is equivalent to ~ 2.3 centimorgans (cM) of rice chromosome 11S, as determined from the published relative map positions of restriction fragment length polymorphism markers C410 (GenBank accession number D15288), C477 (GenBank accession number D15339), C496 (GenBank accession numbers D15347 and D22600), and R682 (GenBank accession number D23967) on the 1998 version of the Rice Genome Research Program genetic map of chromosome 11, as shown in Figures 1A and 1B (see also <http://www.dna.affrc.go.jp:84/publicdata/geneticmap98/chr11pre.html>). The distance between C410 and R682 is 330,670 bp. Those previously mapped markers correspond to gene 33 (DUPR11.33), function unknown, and gene 1 (DUPR11.1), which is *Adh2*, in our sequence (Table 1).

The physical to genetic distance ratio is, on average, 147 kb of DNA per cM. This estimate is only approximately half of the expected average of 295 kb/cM, calculated by assuming a total genetic map length of 1522 cM (<http://www.dna.affrc.go.jp:82/>) and a 450-Mb genome size for rice (Arumuganathan and Earle, 1991).

Gene Repertoire

We identified 33 putative genes (Table 1 and Figure 2), with an average of one gene per 10.3 kb of DNA. Thirteen of these genes were confirmed by the identification of corresponding cDNA clones, and one was confirmed by the isolation of a rapid amplification of cDNA ends (RACE)-PCR product. Nineteen of the genes are homology-based or GENSCAN predictions for which no transcripts were identified.

Adh1 and *Adh2* are the only known genes of rice present in the region. Twelve of the genes are similar in structure to other plant genes with known function. Three major classes

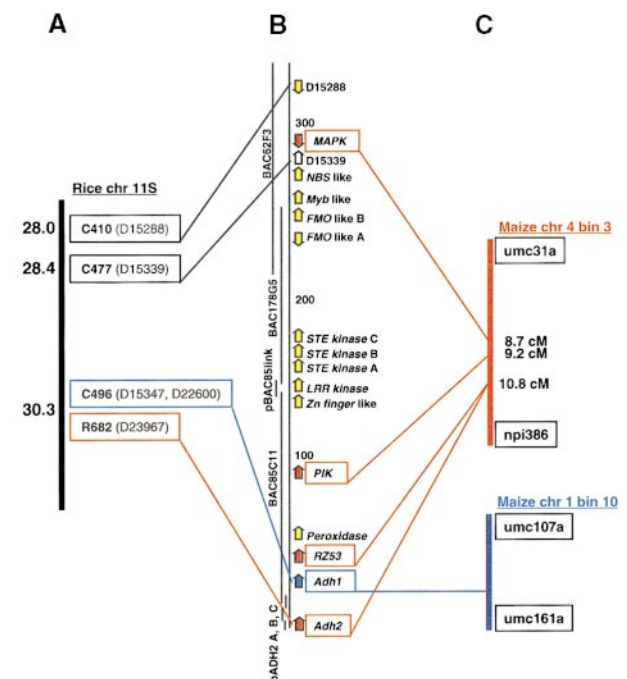


Figure 1. Genetic and Physical Map of Rice Chromosome 11S around the *Adh1* and *Adh2* Loci and Genetic Map of the Corresponding Regions of Maize Chromosomes 1 and 4.

(A) Genetic map of rice 11S. Map positions of markers are shown in centimorgans (<http://www.dna.affrc.go.jp:84/publicdata/geneticmap98/chr11pre.html>).

(B) Physical map of the 339,485-bp segment of rice 11S. Genes indicated in red have been genetically mapped on maize chromosome 4, as shown in (C). The *RZ53* map position is from Ahn and Tanksley (1993). Genes shown in yellow do not cross-hybridize with maize genomic DNA. Blue indicates the *Adh1* gene; its position on the genetic map of maize chromosome 1 has been inferred from published information (MaizeDB; <http://www.agron.missouri.edu/>). The gene indicated in white has not been hybridized with maize genomic DNA. MAPK, mitogen-activated protein kinase; PIK, phosphatidylinositol 4-kinase.

(C) Segments of maize chromosomes 1 and 4 containing maize homologs of rice chromosome 11 genes identified in (B).

of genes, the products of which may be related to disease resistance, were found. These include a protein kinase-like protein that contains a leucine-rich repeat (LRR) and is similar to Xa-21 (DUPR11.16), a nucleotide binding site-containing protein similar to a lettuce disease resistance protein (DUPR11.30), and three gene products similar to

S-domain receptor-like kinases (DUPR11.18, DUPR11.19, and DUPR11.20; Table 1). No function could be attributed to 18 of the putative genes. We found evidence of transcription for six of them: matching ESTs were found in five instances, and a RACE product was isolated for the remaining one (Table 1). Two of the transcriptionally active genes

Table 1. Identified and Predicted Genes

ID ^a	Designation	Strand	Position ^b	Identification Method ^c	Best Protein Homolog ^d	BLASTX E Value ^e	DuPont ESTs	Predicted Gene Product
DUPR11.1	Adh2	+	<5529..8604	Contains M36469, X16297, D23967	P04707	0	rr1.pk0002.b4	Rice alcohol dehydrogenase 2 (EC 1.1.1.1)
DUPR11.2	HYP protein	+	<21150..>24962	Prediction	AAD25557.1	1.00 E-30		Hypothetical protein
DUPR11.3	Adh1	+	<37765..40900	Contains X16296, D15347, D22600	S04571	0	rlr24.pk0070.g12	Rice alcohol dehydrogenase 2 (EC 1.1.1.1)
DUPR11.4	RZ53	+	<42569..43105	Contains AA231931, AA231852	—	No hits	rsr9n.pk004.e21	Unknown
DUPR11.5	Peroxidase like	+	<51800..53606	Contains D48298, D48378	AAD37375.1	2.00 E-91	rs11n.pk007.g5	Putative peroxidase
DUPR11.6	NifS like	+	<77129..>80844	Prediction	AE001527	1.00 E-29		Similar to <i>H. pylori</i> NifS-like protein (AE000556)
DUPR11.7	PIK	+	<90248..100554	Contains D24320, D22005, C72631	CAB37928	0	rls2.pk0022.f10	Phosphatidylinositol kinase
DUPR11.8	REP31	+	<105386..>107812	Prediction	AAD21743.1	9.00 E-9		Hypothetical protein, REP family member A
DUPR11.9	REP34	+	<113109..>116512	Prediction, apparent pseudogene	CAB10188.1	7.00 E-06		REP family member B
DUPR11.10	REP37	+	<118957..>121953	Prediction	CAB10175.1	6.00 E-11		Hypothetical protein, REP family member C
DUPR11.11	REP39	+	<124746..>126658	Prediction	CAB10175.1	2.00 E-7		Hypothetical protein, REP family member D
DUPR11.12	REP43	+	<135666..138397	RACE-PCR product	CAB10175.1	9.00 E-9		Unknown, REP family member E
DUPR11.13	REP45	+	<139106..>141138	Prediction	CAB10175.1	7.00 E-12		Hypothetical protein, REP family member F
DUPR11.14	Zn-finger like	+	<143232..>144488	Prediction	U90439	3.00 E-25		Similar to Arabidopsis Cys3His zinc finger protein isolog (U90439)
DUPR11.15	Unknown	—	146521..>148569	Contains AA751852	—	No hits	rls48.pk0010.b2	Unknown
DUPR11.16	LRR kinase like	+	<150766..161394		AAD55655.1	1.00 E-156	rlr72.pk0008.f11	Similar to LRR/receptor-like protein kinase
DUPR11.17	REP54	+	<161667..163895		CAB10285.1	2.00 E-9	rlr2.pk0017.e2	Unknown, REP family member G
DUPR11.18	STE kinase like A	+	<164742..>167204	Prediction	CAA09029	0		Similar to S-domain receptor-like protein kinase, gene family member A
DUPR11.19	STE kinase like B	+	<171273..>173739	Prediction, apparent pseudogene	CAA09029	1.00 E-180		Similar to S-domain receptor-like protein kinase, gene family member B
DUPR11.20	STE kinase like C	+	<182063..>184516	Prediction, apparent pseudogene	CAA09029	0		Similar to S-domain receptor-like protein kinase, gene family member C
DUPR11.21	REP63	+	<185828..>188063	Prediction, apparent pseudogene	CAB10188.1	5.00 E-4		REP family member H
DUPR11.22	REP67	+	<194126..>196227	Prediction	CAB10188.1	1.00 E-10		Hypothetical protein, REP family member I
DUPR11.23	REP69	+	<199310..>207442	Prediction, apparent pseudogene	AC005679	1.00 E-5		REP family member J
DUPR11.24	REP77	+	<219593..>221221	Prediction	AC005679	6.00 E-8		Hypothetical protein, REP family member K
DUPR11.25	REP82	—	<233157..>234908	Prediction	AC005679	9.00 E-7		Hypothetical protein, REP family member L
DUPR11.26	FMO like A	—	<237296..>239012	Prediction	AC002411	6.00 E-95		Putative dimethylaniline monooxygenase, gene family member
DUPR11.27	REP84	+	<241468..>254983	Prediction, apparent pseudogene	AC005679	9.00 E-7		REP family member M
DUPR11.28	FMO like B	+	<260372..>264609	Prediction	AC002411	4.00 E-91		Putative dimethylaniline monooxygenase, gene family member B
DUPR11.29	MYB like	—	<270379..>273737	Contains AU032361, AU032360	AF062888	2.00 E-39		Myb-like protein
DUPR11.30	NBS like	+	<284125..285885	Transcribed apparent pseudogene	AF017751	7.00 E-31	rr1.pk0026.c10	Similar to (AF017751) <i>L. sativa</i> resistance protein candidate
DUPR11.31	Unknown D15339	+	<287097..288403	Contains D15339	—	No hits		Unknown
DUPR11.32	MAPK like	—	288757..>293123	Contains D46594	U58918	1.00 E-140	rlr24.pk0099.c2	Putative MAPK
DUPR11.33	Unknown D15288	—	323101..330671	Contains D15288, D42495	CAA18174.1	4.00 E-82	rlr24.pk0091.f6	Unknown

^a As listed in GenBank AF172282.

^b The carrots (<...>) indicate that the sequence range is an approximation.

^c GenBank accession numbers are provided when appropriate. If no exact match to a genomic or EST sequence was identified, then the putative gene is listed as Prediction.

^d GenBank accession number of closest protein homolog. Dash indicates no protein homologs found in GenBank.

^e For definition of Blast Expect E values, see <http://www.ncbi.nlm.nih.gov/BLAST/>.

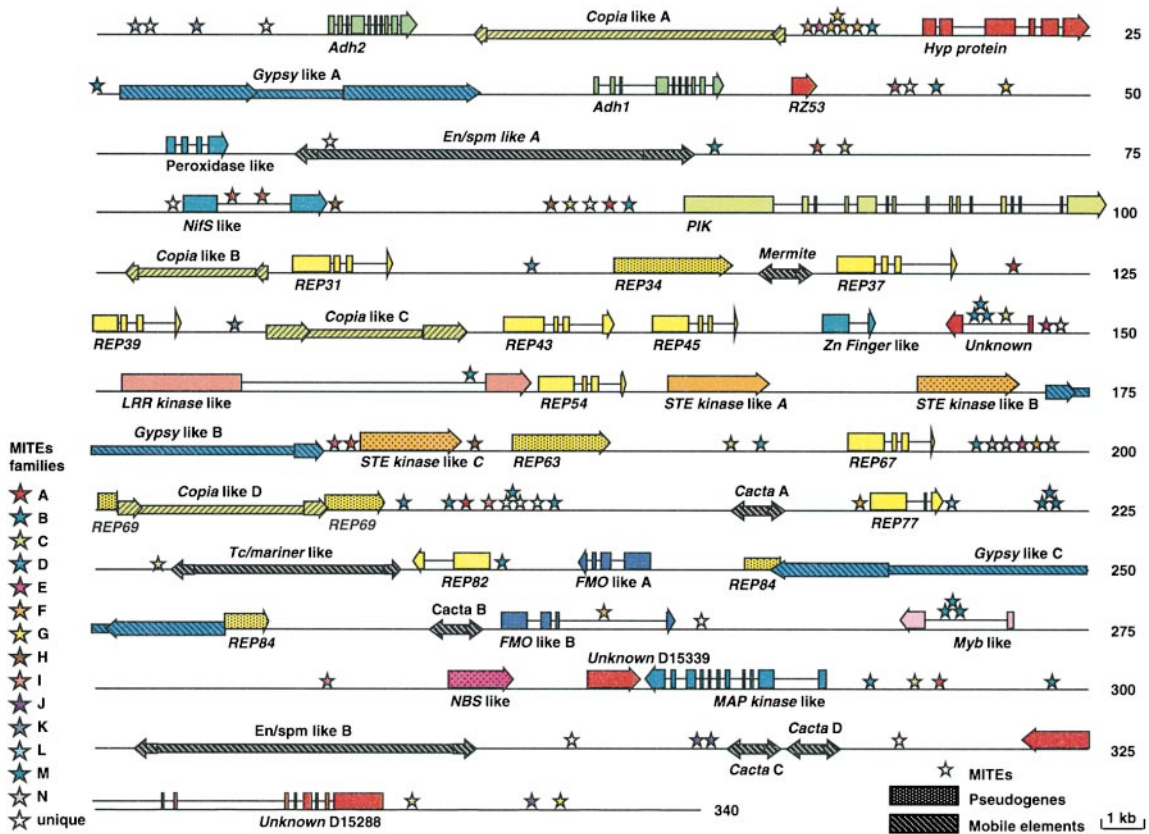


Figure 2. Gene and Repetitive Element Map of the 339,485-bp Segment of Rice Chromosome 11S.

Positions of genes reported in Table 1 and transposable/repetitive elements reported in Table 2 are shown. Arrows indicate direction of transcription. MITEs (Wessler et al., 1995) are indicated with colored stars. Different classes of MITEs are shown in different colors (R. Tarchini, unpublished results). NBS, nucleotide binding site.

(*REP54* and *REP43*) are members of a family that is represented by 13 different copies. Nine of those copies have the potential to encode functional proteins; the others are apparently pseudogenes.

Repetitive DNA

The region contains numerous class I (retrotransposon) and class II (transposon) mobile genetic elements (Grandbastien, 1992), miniature inverted repeat transposable elements (MITEs; Wessler et al., 1995) (Table 2 and Figure 2), and simple sequence repeats (Weber and May, 1989). Four *Ty1-Copia*-like and three *Ty3-Gypsy*-like retrotransposons were found. Most of them are not highly similar to any well-characterized rice retrotransposons. The identified structural similarities are to partial sequences of repetitive DNA in rice

or are limited to the highly conserved reverse transcriptase domain. Some of the elements show nucleotide sequence similarities to retrotransposons of other species. *Copia C* resembles the wheat *Tar1* element (Matsuoka and Tsunewaki, 1997), and *Copia D* is similar to *BARE-1*, a *Copia*-like retroelement of barley (Manninen and Schulman, 1993). *Gypsy C* is highly similar to *retosat2* of rice (GenBank accession number AF111709). It also resembles the maize retrotransposon *Tekay* (GenBank accession number AF050455), the *dea1* element from the evolutionarily more distant pineapple (*Ananas comosus*) (Thomson et al., 1998), and the *Lilium henryi del* retrotransposon (Smyth et al., 1989). With the possible exception of *Gypsy C*, all of the elements are most probably inactive, given the presence of frameshifts, missense mutations, or deletions and rearrangements.

No long interspersed nuclear elements were found by searching for the reverse transcriptase domain in the vicinity

of oligo(dA/dT) sequences. All reverse transcriptase domains found were within long terminal repeat (LTR)-containing elements described above.

Six of the eight identified transposons can be ascribed to the *CACTA* family of DNA transposable elements (Gierl et al., 1989), based on the presence of the characteristic *CACTA* motif in the terminal inverted repeats. The two major rice elements of this group are unique in sequence, but both show similarities to the coding portions of the maize *Enhancer/Suppressor mutator (En/Spm)* system (Pereira et al.,

1986) and the snapdragon *Tam1* element (Nacken et al., 1991) (Table 2). More than 95% of the *En/Spm A* nucleotides are identical to those of *RIM2*, a rice transposon protein-like cDNA induced by *Magnaporthe grisea* (GenBank accession number AF121139). *CACTA A*, *CACTA B*, *CACTA C*, and *CACTA D* are highly homologous to each other and can be considered members of the same family of defective transposable elements. The remaining two transposable elements are a small, presumably defective, novel transposable element and a *Tc1/Mariner*-like element (Hartl et al., 1997),

Table 2. Transposons and Retrotransposons Identified in This Study

Class	Position	5' LTR	3' LTR	Main Features ^a	Size (bp)	Total	%
Retrotransposons							
Gypsy like							
Gypsy-like A	25901..34850	25901..29267	31484..34850	5-nucleotide direct repeats (TTATC), >99.9% identical LTRs, no evident CDS LTRs homologous to dbj AP000364.1 AP000364 (BlastN E = 0) LTRs homologous to GYPSY C LTRs	8,950		
Gypsy-like B	173805..180871	173805..174506	180170..180871	4-nucleotide direct repeats (TATT), 97% identical LTRs, degenerated CDS CDS homologous to gb AAD27548.1 AF111709.2 (BlastX E = 2e-18)	7,067		
Gypsy-like C	242372..253905	250834..253905	242372..245442	5-nucleotide direct repeats (TTAGC), >99.9% identical LTRs, complete CDS LTRs homologous to dbj AB026295.2 AB026295 (BlastN E = 0) CDS homologous to gb AF111709.1 AF111709 (BlastN E = 0)	11,534	27,551	
Copia like							
Copia-like A	9612..17334	17060..17334	9612..9868	5-nucleotide direct repeats (TAAAC), >97.59% identical LTRs, complete CDS CDS homologous to dbj BAA01703 (BlastX E = e-110)	7,723		
Copia-like B	<101197..>104216	ND ^b	ND	LTRs not defined, complete CDS CDS homologous to dbj BAA81872.1 (BlastX E = e-134)	3,020		
Copia-like C	129515..134466	129515..130607	133134..134466	2-nucleotide direct repeats (TA), 96% identical LTRs, incomplete CDS Homologous to rice dbj AB017983.1 (BlastN E = e-123) and dbj BAA22288 (BlastX E = e-119)	4,952		
Copia-like D	200748..206421	200748..201250	205919..206421	5-nucleotide direct repeats (TTTTG), 99.6% identical LTRs, complete CDS Homologous to rice gb AF161269.1 (BlastN E = 0) and gj 4234854 (BlastX E = e-100)	5,674	21,369	
Total						48,920	14.41
Transposons							
En/Spm like							
En/Spm like	55331..65393			3-nucleotide direct repeats (CAC), 15-nucleotide imperfect terminal inverted repeats Homologous to rice RIM2, gb AF121139.1 (BlastN E = 0)	10,063		
Mermite	117189..118589			7-nucleotide direct repeats (CAAAATT), 270-nucleotide almost perfect terminal inverted repeats, no CDS No significant homologies to other elements	1,401		
Tc1/Mariner like	227524..232786			2-nucleotide direct repeats (TA), 28-nucleotide perfect terminal inverted repeats, complete CDS Homologous to rice dbj AB023482.2 (BlastN E = e-67) and soybean Mariner transposase gj 3386534 (BlastP E = e-108)	5,263		
En/Spm-like B	301270..309485			3-nucleotide direct repeats (GTA), 33-nucleotide imperfect terminal inverted repeats Homologous to pir S29329 (BlastX E = 0)	8,216		
CACTA-like A	216205..217591			3-nucleotide direct repeats (TTT), 21-nucleotide imperfect terminal inverted repeats, no CDS, no homologies	1,387		
CACTA-like B	258636..259738			3-nucleotide direct repeats (TAC), 21-nucleotide imperfect terminal inverted repeats, no CDS, no homologies	1,102		
CACTA-like C	316465..317581			3-nucleotide direct repeats (ACT), 25-nucleotide imperfect terminal inverted repeats, no CDS, no homologies	1,117		
CACTA-like D	317852..318889			3-nucleotide direct repeats (GTA), 20-nucleotide imperfect terminal inverted repeats, no CDS, no homologies	1,038		
Total						29,587	8.71
MITES (total)				78 individual MITES, 12 unique, 66 organized in 14 families		18,130	5.34
Total						96,637	28.46

^a CDS, coding sequence.

^b ND, not defined.

providing evidence for the presence of this family of transposable elements in rice.

MITEs are the numerically most abundant class of repetitive DNA elements found in this region (Figure 2). We were able to define a total of 78 MITEs (an average of one every 4.36 kb) and to classify them according to their overall sequence similarities (R. Tarchini, unpublished results). Of these, 66 were assigned to 14 different families. Each family contains at least two and at most 13 members.

Among the simple sequence repeats containing at least six repeats, AT/TA repeats were the most abundant (11), followed by AG/TC (10) and AC/GT (four), confirming previous observations of relative simple sequence repeat abundance reported in the literature (Lagerkrantz et al., 1993). Ten of the 11 AT/TA repeats are concentrated in a 94-kb interval (nucleotides 210,277 to 304,540).

Genetic Mapping with Rice Probes in Maize

Thirteen loci from the sequenced region were tested for their sequence conservation and genetic map position in maize (Figure 1C). Five rice genes hybridized well with maize genomic DNA: *Adh1*, *Adh2*, *RZ53*, phosphatidylinositol 4-kinase (*PIK*), and mitogen-activated protein kinase (*MAPK*) genes. Three probes, *Adh2*, *PIK*, and *MAPK*, identified useful polymorphism on the available maize mapping populations, ALEB9 and DRAG2, each of which contained 86 individuals (see Methods). *Adh2* was mapped to chromosome 4 in the DRAG2 population, 10.8 cM downstream of *umc31a*. *PIK* and *MAPK* were mapped as 8.7 and 9.2 cM, respectively, downstream of *umc31a* in the ALEB9 population. In all cases, the loci were placed on the map with LOD scores >20 (Lander et al., 1987). However, because of the modest size of the mapping population, the absolute order of the markers on maize chromosome 4 must be considered provisional, and differences from the order of *MAPK*, *PIK*, *RZ53*, and *Adh2* loci observed in rice cannot be excluded. In rice, *Adh1* (GenBank accession numbers D15347 and D22600) and *Adh2* (GenBank accession number D23967) map to chromosome 11S position 30.3, unknown sequence DUPR11.31 (GenBank accession number D15339) maps to chromosome 11S position 28.4, and unknown sequence DUPR11.33 (GenBank accession number D15288) maps to chromosome 11S position 28.0 (Figures 1A and 1B; <http://www.dna.affrc.go.jp:84/publicdata/geneticmap98/chr11pre.html>).

DISCUSSION

Genomic Organization

Analysis of the sequence of 339,485 bp of the rice *Adh1* region revealed the presence of 33 putative genes corre-

sponding to one gene per 10.3 kb. This is ~1.4 times higher than the average of one gene every 14 kb, calculated by assuming a total of 30,000 genes, and approximately one-half of the reported gene density for Arabidopsis (4.6 kb per gene; Bevan et al., 1998).

The comparison of the physical interval with the genetic distance shows that the frequency of recombination in this region is twice the average for rice. The ratio of physical to genetic distance in this region is 147 kb/cM, as compared with the rice genome average of 295 kb/cM. This result may be explained by a local variation in the recombination frequency or by an error in estimating the genetic distance.

Gene sequences comprise a total of 101,923 bp of annotated features (including exons, introns, and 5' and 3' untranslated regions for which ESTs were available) corresponding to 30% of the 399,485 bp that were sequenced. An interesting feature of the region is the frequent occurrence of gene duplication. Examples include two alcohol dehydrogenases, two monooxygenases, three serine/threonine kinase-like genes, and a family of 13 genes of unknown function clustered in a 150-kb interval. All of the duplicated genes are closely spaced and occur mostly in the same orientation (Table 1 and Figure 2). Unequal crossing-over followed by sequence divergence is a likely explanation, as has been proposed for the evolution of multigene families (Nei et al., 1997; Zhang et al., 1988). An exception is a small inversion involving the region containing *REP82* and one of the two copies of the flavine monooxygenase-like (*FMOA*) gene.

The identification of genes was aided by the availability of EST collections. Eleven of the 33 candidate genes were confirmed by the identification of homologous ESTs in our collection of >120,000 rice ESTs. Eight of the 11 rice genes homologous to DuPont rice ESTs also have corresponding GenBank sequences (Table 1). Two predicted genes (a *Myb*-like gene and an unknown gene) correspond to the publicly available ESTs, but there is no corresponding EST in our collection.

The main repetitive DNA fraction present in the region had sequence similarities to different classes of mobile genetic elements. Except for the non-LTR retrotransposons, all major classes of known transposable elements are represented. Excluding gene families from the calculation, the repetitive fraction comprises 28.46% of the region sequence, of which half is represented by retrotransposons. MITEs alone constitute 18.8% of the repetitive DNA (on average, one MITE per 4.36 kb; 5.34% of the total DNA) and are the third most abundant components after transposons (8.7% of the total DNA). In contrast, the retrotransposon fraction in maize accounts for 60% of the genome, and MITEs appear to be rare (M. Morgante, personal communication). This observation provides additional evidence for the proposal (SanMiguel et al., 1996) that retrotransposon proliferation contributed greatly to the genome expansion observed in maize.

All of the transposons and retrotransposons except the

MITEs appear to be dispersed, with no obvious clustering. This is in contrast to what was observed with maize (SanMiguel et al., 1996). For maize, nested retroelements are the main component of the intergenic regions. This difference may be adequately explained by differences in the density of retroelements. Also, as SanMiguel et al. (1996) hypothesized, several successive invasions of different retroelement families may have contributed to the complexity of the genomic organization in maize. In rice, most of the homologs to mobile genetic elements reside between genes. Insertions in putative coding regions can be seen in two cases (*REP69* and *REP84*; Figure 2). In both instances, the inserted element is a retrotransposon and the disrupted gene is a member of the REP gene family. This indicates that duplications leading to the expansion of that gene family are antecedent to the insertion of the retroelements. The insertion of *Gypsy B* between *Kinase A* and *B* might provide further evidence in favor of this view. MITE-to-MITE distances follow a gamma distribution, with skewness of 1/8, indicating a statistically significant clustering of MITEs in the region examined here. The data available were insufficient to determine possible association of MITEs with coding sequences, as has been proposed for maize (Wessler et al., 1995).

Data from the comparative analysis of the orthologous *Adh1* regions of maize and sorghum suggest that the dramatic increase in size of the maize *Adh1* region dates back only 3 million years and that almost all of the retrotransposon insertions occurred in the last 6 million years (SanMiguel et al., 1998). This is long after the germplines of maize and sorghum diverged.

Retrotransposons with LTRs, such as the rice *Gypsy A* and *Gypsy C*, increased the difficulty of assembly of the genomic sequence, a problem that is likely to be encountered in the future. Cataloging these elements may allow automated highlighting of problematic regions.

Gene Repertoire

The presence of a cluster of genes structurally similar to disease resistance genes is an interesting feature of the region. One example is an LRR protein kinase-like gene showing similarity to the rice *Xa-21* (Wang et al., 1996) gene product. The gene encodes a putative 1099-amino acid protein with 25 LRRs, which is followed by a potential transmembrane domain and a serine/threonine kinase-like domain. Although the similarity to *Xa-21* at the protein level is <50% and no significant nucleotide similarity is detectable, the presence of a single intron in the same position in the two genes suggests that they might be paralogs. Another example is a single-copy gene encoding a protein containing a nucleotide binding site and showing similarity to a lettuce resistance protein candidate (GenBank accession number AF017751; Meyers et al., 1998) and to an isolog of Arabidopsis disease resistance protein RPM1 (GenBank accession number U95973). Although the gene is actively transcribed, it con-

tains a frameshift mutation and therefore might represent a transcribed pseudogene, presumably arising from a very recent mutation.

Three members of a family of putative serine/threonine receptor-like kinases, similar to S-domain receptor-like protein kinases and to Arabidopsis *STE*-like receptor kinases (Walker, 1993), are also present and form a tight cluster spanning ~20 kb of DNA sequence. Nucleotide sequence identity among the three genes is >85%, and one of these elements (*STE Kinase A*) contains a single open reading frame capable of encoding an 820-amino acid protein. The other two members of the cluster are apparent pseudogenes, as indicated by the presence of frameshifts and stop codons prematurely interrupting the open reading frame. The function of both the Arabidopsis genes and the rice genes reported here remains to be established, but a role in signal transduction that possibly involves a pathogen response may be hypothesized on the basis of the structural features.

Components of signal transduction pathways, *MAPK*, *PIK*, and a *Myb*-like factor are also present, as are sequences similar to enzymes involved in oxidative burst (peroxidase) or in detoxification processes (dimethylaniline monooxygenases) (Hammond-Kosack and Jones, 1997; Yang et al., 1997).

The association between the *Adh1* region and the locus controlling the resistance to *M. grisea* race a (*Pi-a*) has been reported by Goto et al. (1981). According to these studies, the *Pi-a* locus is located 1 cM from *Adh1*, placing it directly in the interval sequenced. The double haploid line used for the construction of the BAC library derives from the crossing of rice varieties sensitive to race a of *M. grisea*. Therefore, if the *Pi-a* locus is included in the sequenced region, then it will be represented by its sensitive allele.

Interestingly, a homolog of *NifS* was found. *NifS*, one of the components of the major *Nif* gene cluster of *Azotobacter vinelandii*, is required for the activity of the bacterial nitrogenases (Kennedy and Dean, 1992). *NifS* homologs have been isolated from different organisms, including non-nitrogen-fixing bacteria (S. Chen et al., 1997) and higher eukaryotes, including mammals (Nakai et al., 1998), but to our knowledge no plant homolog has been reported.

Synteny in the *Adh2* Homologous Regions of Rice and Maize

The relationship between rice and maize genetic maps in the vicinity of *Adh1-Adh2* was established by mapping rice genes derived from this region in two different maize populations.

Thirteen different rice probes were used to map three homologous maize loci in bin 3 of maize chromosome 4. This group included maize loci homologous to the rice *Adh2*, *PIK*, and *MAPK* genes. The association of the *RZ53*-homologous locus of maize with *Adh2* is based on a published report (Ahn and Tanksley, 1993). The mapping results (Figure

1) indicate microcolinearity between the rice *Adh1-Adh2* region on chromosome 11 and the maize chromosome 4 segment around the *Adh2* locus. A notable exception to this synteny is seen at the *Adh1* locus, which is discussed later. Eight of the rice probes failed to hybridize with maize genomic DNA on DNA gel blots at high and medium stringency, and the extent of gene order conservation for these genes remains to be assessed. We expect that the homologs of these rice genes are present in maize, but the degree of sequence conservation is less. The lack of DNA sequence conservation of many genes in this region represents a factor that will limit studies based on utilizing genome colinearity between these two species.

In general, two classes of DNA sequences may be distinguished. One corresponds to conserved genes encoding basic metabolic functions. This class includes *Adh1*, which shows as much as 94 to 100% sequence conservation between rice and maize. The other class, which shows less sequence conservation, as measured by a lack of interspecific DNA hybridization, is likely to include genes with a more specialized function and genes that have been subjected to divergent selection. Genes structurally similar to disease resistance genes are an example of this latter class of DNA sequence (Meyers et al., 1998). Leister et al. (1998) postulates that reorganization of disease resistance genes during cereal evolution was rapid. Sequencing of the maize chromosome 4 region between the *Adh2* and the *MAPK* genes would allow further insights into the divergence of these disease resistance-like genes between maize and rice.

As noted earlier, a significant exception to the colinearity between the rice chromosome 11 and the maize chromosome 4 occurs at the *Adh1* locus. Whereas in rice, *Adh1* and *Adh2* are only 35 kb apart, in maize and sorghum these two loci are on different chromosomes (*Adh1* is on maize chromosome 1 and sorghum linkage group C) (Paterson et al., 1995). Nakajima and co-workers observed a lack of colinearity of chromosome segments surrounding maize and rice *Adh1* loci (Tikhonov et al., 1999). A sequence-based analysis of the phylogenetic relationship shows that maize and rice *Adh1* as well as maize and rice *Adh2* are pairs of orthologs (results not shown). Gaut et al. (1999) proposed recently that *Adh* duplicated into *Adh1* and *Adh2* before the radiation of the grasses occurred ~65 million years ago. These observations are reconciled by proposing a translocation of the *Adh1* locus to a different chromosome in the lineage leading to maize and sorghum. The proposed translocation occurred before the divergence of maize and sorghum, as indicated by the overall colinearity observed in the *Adh1* regions of these species (Bennetzen et al., 1998; Tikhonov et al., 1999). The boundaries of the *Adh1* translocated region can be circumscribed to the small physical interval (34 kb) between the *Adh2* and the *RZ53* genes. Interestingly, the maize genome effectively doubled in size after its evolutionary separation from sorghum, but the *Adh1* and *Adh2* loci are still single copy in maize, as evidenced by hybridization patterns and genetic mapping (data not shown).

These data indicate that gene duplication, followed by divergent selection at different rates, and small translocations involving single genes play a role in the evolution of cereal genomes. On the background of overall colinearity, small rearrangements were also identified between maize and sorghum in the *Adh1* region, extending beyond the borders of the proposed *Adh1* translocation (Tikhonov et al., 1999). The frequency and molecular nature of such events will become clearer once larger segments of both rice and maize genomic DNA sequence become available. These evolutionary events also complicate the use of rice as a system for synteny-based gene isolation or identification in other grass genomes. Nonetheless, rice, with its relatively simple genome organization, remains attractive for comparative genome studies with grasses.

METHODS

Physical Mapping

The rice bacterial artificial chromosome (BAC) library from a double haploid line (YT14) derived from a cross between *Oryza sativa* ssp *japonica* cv Yashiro-mochi and Tsuyake was kindly provided by B. Valent and K.-S. Wu (DuPont Co., Wilmington, DE) (Wu et al., 1996). The BAC library was gridded onto fourteen 8 × 12-cm filters in a 4 × 4 pattern by using an HDR 96-Pin tool for Beckman Biomek 1000 (Beckman Instruments, Fullerton, CA). Filters were produced and processed under the conditions recommended by Olsen et al. (1993).

BAC85C11, BAC92H8, BAC166F9, and BAC196E1 were isolated by hybridization with a 1.3-kb rice alcohol dehydrogenase *Adh1* probe generated by amplification of rice genomic DNA with primers OSADH1.C.for (5'-GGAAGCCCATTTACCATT-3') and OSADH1.C.rev (5'-GCCCAGGATACACAGAAGA-3').

The hybridization probes were labeled with ³²P-dCTP by using the RadPrime labeling kit (Life Technologies, Rockville, MD) according to the manufacturer's instructions.

Hybridization was performed in a solution of 1 M NaCl, 50 mM Tris-HCl, pH 7.5, 1% SDS, and 5% dextran sulfate at 65°C overnight with a final wash in 0.1 × SSPE (1 × SSPE is 150 mM NaCl, 10 mM NaH₂PO₄, and 1 mM EDTA, pH 7.4), and 0.1% SDS at 65°C.

The following primers were used to test the isolated BAC clones for the presence of the *Adh2* locus: OSADH2.5.for (5'-GAGAGA-AAAGGCATCCATCC-3'), OSADH2.5.rev (5'-AGGGCGGTGTAGAGG-ATCTT-3'), OSADH2.CD.for (5'-GGTGTGTGTGTGGTTTCTGC-3'), OSADH2.CD.rev (5'-AGTCCACCGTTGGTCATCTC-3'), OSADH2.AG.for (5'-GAGTCTCCGCTGCGTCAT-3'), and OSADH2.AG.rev (5'-TCTCATCCATTTTTGCTTTCA-3'). The primers, which were designed on the basis of the rice *Adh2* locus sequence (GenBank accession number M36469), were used in all possible forward and reverse combinations.

BAC DNA was extracted by using an alkaline lysis procedure followed by cesium chloride gradient purification (Sambrook et al., 1989). The DNA was digested with NotI and HindIII restriction endonucleases, and the DNA fragments that were obtained were compared by using gel electrophoresis to establish the extent of overlap of different BAC clones. BAC end sequences were generated by direct dye terminator sequencing with M13 universal primers.

BAC167C6, BAC178G5, and BAC62F3 were successively isolated

by screening pooled BAC clones with polymerase chain reaction (PCR) primers designed from the end sequences of, respectively, BAC85C11, BAC167C6, and BAC178G5.

DNA Sequencing and Assembly

A set of minimally overlapping clones consisting of BAC85C11, BAC178G5, and BAC62F3 were chosen for complete sequencing (Figure 1B). Smaller insert size subclones were sequenced to link BAC85C11 to the *Adh2* locus (p*Adh2A*, p*Adh2B*, and p*Adh2C* from BAC166F9) and to BAC178G5 (pBAC85link from BAC167C6).

BAC85C11, BAC178G5, and BAC62F3 were sequenced by using a shotgun approach (Bodenteich et al., 1993). Cesium chloride-purified BAC DNA was sheared by nebulization (Roe et al., 1996). End repair was performed by using Pfu DNA polymerase (Stratagene, La Jolla, CA) treatment according to the manufacturer's directions. DNA fragments were size-fractionated and cloned into the SmaI site of pUC18 (Amersham Pharmacia Biotech, Piscataway, NJ). After transformation into *Escherichia coli* DH10B electrocompetent cells (Life Technologies, Rockville, MD), recombinant clones were randomly picked. DNA templates for sequencing were isolated by using a 96-well alkaline lysis miniprep kit (Advanced Genetic Technologies Corp, Gaithersburg, MD). Sequencing reactions were performed by using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit with FS AmpliTaq DNA polymerase (PE Applied Biosystems, Foster City, CA) and analyzed on ABI 377 (PE Applied Biosystems) sequencing gels.

The sequence data were assembled by using PHRED/PHRAP software (Green, 1996).

Contigs were extended and joined by two successive rounds of primer walking. Primers were designed with the program PRIMO (Li et al., 1997) in the version for Sun (Sun Microsystems, Inc., Palo Alto, CA). Remaining gaps were filled by sequencing PCR products bridging the ends of the existing contigs. Regions without adequate sequencing depth or with assembly ambiguities were subcloned and resequenced.

Clones p*Adh2A*, p*Adh2B*, p*Adh2C*, and pBAC85link were sequenced by using the Prism Primer Island Transposition kit (PE Applied Biosystems).

Sequence Analysis

Homology searches against public and private (DuPont) databases were used to identify candidate genes in the region. The final sequence of the region was divided into 3-kb overlapping fragments and searched for nucleic or protein homologies by using the BLAST program (Altschul et al., 1997). The GENSCAN program (Burge and Karlin, 1997) was used for gene predictions. Programs from the Genetics Computer Group (Madison, WI) were used for sequence comparison and to identify motifs in the sequenced region. The DOTTER program (Sonnhammer and Durbin, 1995) was used to identify and classify repeat families and miniature inverted repeat transposable elements (MITEs) (Wessler et al., 1995). Lasergene software (DNASar, Inc., Madison, WI) was used for sequence similarity analysis.

Comparative Mapping

Eight expressed sequence tag (EST) clones and five genomic subclones containing gene regions of interest were used as restriction

fragment length polymorphism probes on maize genomic DNA to determine the conservation and map position of these clones and subclones in maize. The ESTs included rr1.pk0002.b4 (*Adh2*), rls2.pk0022.f10 (*PIK*), and rlr24.pk0099.c2 (*MAPK*). High-stringency hybridization and washing conditions were used (final wash in 0.1 × SSPE and 0.1% SDS at 65°C). Probes that did not hybridize at high stringency were hybridized at lower stringency, but this did not improve the results. Maize populations ALEB9 (86 individuals; pooled F₃ progeny from an R67 × P38 cross) and DRAG2 (86 individuals; pooled F₃ progeny from an ED0 × MW0 cross) were used for mapping purposes. The resulting segregation data were used to place the genes on the maize genetic map with Mapmaker 3.0b (Lander et al., 1987). The order of markers on maize chromosome 4 was inferred on the basis of their relative positions with respect to umc31.

Genbank Accession Numbers

The GenBank accession number for the sequence DUPR11 reported here is AF172282.

ACKNOWLEDGMENTS

We thank Barbara Valent and Kunsheng Wu for sharing the rice BAC library, Mike Hanafey and Romeo Hubner for bioinformatics assistance, Michele Morgante for many stimulating discussions, Maureen Dolan for accommodating our sequencing needs and for comments on the manuscript, Bruce Roe for his advice on genomic sequencing, and Barbara Mazur for her support. We also thank the anonymous reviewers for suggesting improvements to the manuscript.

Received October 18, 1999; accepted January 17, 2000.

REFERENCES

- Ahn, S., and Tanksley, S.D. (1993). Comparative linkage maps of the rice and the maize genomes. *Proc. Natl. Acad. Sci. USA* **90**, 7980–7984.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Arumuganathan, K., and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218.
- Bennetzen, J.L., SanMiguel, P., Chen, M., Tikhonov, A., Francki, M., and Avramova, Z. (1998). Grass genomes. *Proc. Natl. Acad. Sci. USA* **95**, 1975–1978.
- Bevan, M., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**, 485–488.
- Bodenteich, A., Chisoe, S., Wang, Y.F., and Roe, B.A. (1993). Shot-gun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In *Automated*

- DNA Sequencing and Analysis Techniques, J.C. Venter, ed (London: Academic Press), pp. 42–50.
- Burge, C., and Karlin, S.** (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94.
- Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.S., Zhang, H., Wing, R.A., and Bennetzen, J.L.** (1997). Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. *Proc. Natl. Acad. Sci. USA* **94**, 3431–3435.
- Chen, S., Zheng, L., Dean, D.R., and Zalkin, H.** (1997). Role of NifS in maturation of glutamine phosphoribosylpyrophosphate amidotransferase. *J. Bacteriol.* **179**, 7587–7590.
- Devos, K.M., Chao, S., Li, Q.Y., Simonetti, M.C., and Gale, M.D.** (1994). Relationship between chromosome 9 of maize and wheat homeologous group 7 chromosomes. *Genetics* **138**, 1287–1292.
- Gale, M.D., and Devos, K.M.** (1998). Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**, 1971–1974.
- Gaut, B.S., Peek, A.S., Morton, B.R., and Clegg, M.T.** (1999). Patterns of genetic diversification within the *Adh1* gene family in the grasses (Poaceae). *Mol. Biol. Evol.* **16**, 1086–1097.
- Gierl, A., Saedler, A., and Peterson, P.A.** (1989). Maize transposable elements. *Annu. Rev. Genet.* **23**, 71–85.
- Goto, I., Jaw, Y.L., and Baluch, A.A.** (1981). Genetic studies on resistance of rice plant to blast fungus. IV. Linkage analysis of four genes, *Pi-a*, *Pi-k*, *Pi-z* and *Pi-i*. *Ann. Phytopathol. Soc. Jpn.* **47**, 252–254.
- Grandbastien, M.A.** (1992). Retroelements in higher plants. *Trends Genet.* **8**, 103–108.
- Green, P.** (1996). Towards completely automated sequence assembly. DOE Human Genome Program Contractor-Grantee Workshop V, 157. (Washington, DC: U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research).
- Hammond-Kosack, K.E., and Jones, J.D.G.** (1997). Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 1–39.
- Hartl, D.L., Lohe, A.R., and Lozovskaya, E.R.** (1997). Modern thoughts on an ancient mariner: Function, evolution, regulation. *Annu. Rev. Genet.* **31**, 337–358.
- Harushima, Y., et al.** (1998). A high-density rice genetic linkage map with 2275 markers using a single F₂ population. *Genetics* **148**, 479–494.
- Kennedy, C., and Dean, D.** (1992). The *nifU*, *nifS* and *nifV* gene products are required for activity of all three nitrogenases of *Azotobacter vinelandii*. *Mol. Gen. Genet.* **231**, 494–498.
- Kilian, A., Kudrna, D.A., Kleinhofs, A., Yano, M., Kurata, N., Steffenson, B., and Sasaki, T.** (1995). Rice–barley synteny and its application to saturation mapping of the barley *Rpg1* region. *Nucleic Acids Res.* **23**, 2729–2733.
- Kurata, N., Moore, G., Nagamya, Y., Foote, T., Yano, M., Minobe, Y., and Gale, M.** (1994). Conservation of genome structure between rice and wheat. *Bio/Technology* **12**, 276–278.
- Lagerkrantz, U., Ellegren, H., and Andersson, L.** (1993). The abundance of various polymorphic microsatellite motifs differs between plants and animals. *Nucleic Acids Res.* **21**, 1111–1115.
- Lander, E., Green, P., Abrahamson, J., Barlow, A., Daley, M., Lincoln, S., and Newburg, L.** (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.
- Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A., and Schulze-Lefert, P.** (1998). Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. USA* **95**, 370–375.
- Li, P., Kupfer, K.C., Davies, C.J., Burbee, D., Evans, G.A., and Garner, H.R.** (1997). PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing. *Genomics* **40**, 476–485.
- Manninen, I., and Schulman, A.H.** (1993). *BARE-1*, a *Copia*-like retroelement in barley (*Hordeum vulgare* L.). *Plant Mol. Biol.* **22**, 829–846.
- Matsuoka, Y., and Tsunewaki, K.** (1997). Presence of wheat retrotransposons in Gramineae species and the origin of wheat retrotransposon families. *Genes Genet. Syst.* **72**, 335–343.
- Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S., and Michelmore, R.W.** (1998). Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* **11**, 1833–1846.
- Moore, G., Devos, K.M., Wang, Z., and Gale, M.D.** (1995). Cereal genome evolution. Grasses, line up and form a circle. *Curr. Biol.* **5**, 737–739.
- Nacken, W.K.F., Piotrowiak, R., Saedler, H., and Sommer, H.** (1991). The transposable element *Tam1* from *Antirrhinum majus* shows structural homology to the maize transposon *En/Spm* and has no sequence specificity of insertion. *Mol. Gen. Genet.* **228**, 201–208.
- Nakai, Y., Yoshihara, Y., Hayashi, H., and Kagamiyama, H.** (1998). cDNA cloning and characterization of mouse *Nifs*-like protein, *m-Nfs1*: Mitochondrial localization of eukaryotic NifS-like proteins. *FEBS Lett.* **433**, 143–148.
- Nei, M., Gu, X., and Sitnikova, T.** (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**, 7799–7806.
- Olsen, A.S., Combs, J., Garcia, E., Elliott, J., Amemiya, C., de Jong, P., and Threadgill, G.** (1993). Automated production of high density cosmid and YAC colony filters using a robotic workstation. *BioTechniques* **14**, 116–123.
- Paterson, A.H., Lin, Y.R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S.C., Stansel, J.W., and Irvine, J.E.** (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714–1718.
- Pereira, A., Cuypers, H., Gierl, A., Schwarz-Sommer, Z., and Saedler, H.** (1986). Molecular analysis of the *En/Spm* transposable element system of *Zea mays*. *EMBO J.* **5**, 835–841.
- Roe, B.A., Crabtree, J.S., and Khan, A.S.** (1996). DNA Isolation and Sequencing. (New York: John Wiley and Sons).
- Sambrook, J., Fritsch, E.F., and Maniatis, T.** (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L.** (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765–768.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L.** (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**, 43–45.

- Sasaki, T.** (1998). The rice genome project in Japan. *Proc. Natl. Acad. Sci. USA* **95**, 2027–2028.
- Smyth, D.R., Kalitsis, P., Joseph, J.L., and Sentry, J.W.** (1989). Plant retrotransposon from *Lilium henryi* related to *Ty3* of yeast and the *Gypsy* group of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **86**, 5013–5019.
- Sonnhammer, E.L., and Durbin, R.** (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**, GC1–GC10.
- Thomson, K.G., Thomas, J.E., and Dietzgen, R.G.** (1998). Retrotransposon-like sequences integrated into the genome of pineapple, *Ananas comosus*. *Plant Mol. Biol.* **38**, 461–465.
- Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L., and Avramova, Z.** (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* **96**, 7409–7414.
- Walker, J.C.** (1993). Receptor-like protein kinase genes of *Arabidopsis thaliana*. *Plant J.* **3**, 451–456.
- Wang, G.-L., Song, W.-Y., Ruan, D.-L., Sideris, S., and Ronald, P.C.** (1996). The cloned gene, *Xa21*, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. *Mol. Plant-Microbe Interact.* **9**, 850–855.
- Weber, J., and May, P.E.** (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396.
- Wessler, S.R., Bureau, T.E., and White, S.E.** (1995). LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814–821.
- Wu, K.S., Martinez, C., Lentini, Z., Thome, J., Chumley, F.G., Scolnik, P.A., and Valent, B.** (1996). Cloning a blast resistance gene by chromosome walking. In *Rice Genetics III. Proceedings of the Third International Rice Genetics Symposium*, October 16–20, 1995, G.S. Khush, ed (Manila, The Philippines: International Rice Research Institute), pp. 669–674.
- Yang, Y., Shah, J., and Klessig, D.F.** (1997). Signal perception and transduction in plant defense responses. *Genes Dev.* **11**, 1621–1639.
- Zhang, J., Rosenberg, H.F., and Nei, M.** (1988). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**, 3708–3713.