

The MITE family *Heartbreaker* (*Hbr*): Molecular markers in maize

Alexandra M. Casa*, Cory Brouwer†, Alexander Nagel‡, Lianjiang Wang‡, Qiang Zhang*§, Stephen Kresovich*, and Susan R. Wessler*¶

‡Departments of Botany and Genetics, The University of Georgia, Athens, GA 30602; *The Institute for Genomic Diversity and Department of Plant Breeding, Cornell University, Ithaca, NY 14853; and †Pioneer Hi-Bred International, Inc., Johnston, IA 50131

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 28, 1998.

Contributed by Susan R. Wessler, July 7, 2000

Transposable elements are ubiquitous in plant genomes, where they frequently comprise the majority of genomic DNA. The maize genome, which is believed to be structurally representative of large plant genomes, contains single genes or small gene islands interspersed with much longer blocks of retrotransposons. Given this organization, it would be desirable to identify molecular markers preferentially located in genic regions. In this report, the features of a newly described family of miniature inverted repeat transposable elements (MITEs) (called *Heartbreaker*), including high copy number and polymorphism, stability, and preference for genic regions, have been exploited in the development of a class of molecular markers for maize. To this end, a modification of the AFLP procedure called transposon display was used to generate and display hundreds of genomic fragments anchored in *Hbr* elements. An average of 52 markers were amplified for each primer combination tested. In all, 213 polymorphic fragments were reliably scored and mapped in 100 recombinant inbred lines derived from a cross between the maize inbreds B73 × Mo17. In this mapping population, *Hbr* markers are distributed evenly across the 10 maize chromosomes. This procedure should be of general use in the development of markers for other MITE families in maize and in other plant and animal species where MITEs have been identified.

Bacterial, plant, and animal genomes are populated with numerous families of transposable elements (TEs) that possess a variety of remarkable capabilities. Since the discovery of TEs by McClintock (1), their biological properties have intrigued scientists while simultaneously providing the raw materials that have been fashioned into versatile experimental tools.

There are two broad classes of TEs, each with characteristic properties (2). For all Class 1 or retroelements, such as retrotransposons, short interspersed nuclear elements, and long interspersed nuclear elements, it is the element-encoded mRNA, and not the element itself, that forms the transposition intermediate. In contrast, Class 2 or DNA elements are characterized by short terminal inverted repeats (TIRs) and transposition occurs via a DNA intermediate. Class 2 elements are themselves divided into two groups. Autonomous elements like *Ac* and *Spm* encode all of the products necessary for their transposition in maize and in some other plant species. Nonautonomous elements like *Ds* and *dSpm* are usually deletion derivatives of autonomous elements and, as such, require the presence of autonomous elements for transposition (3). Most field and wild strains lack transposase activity for all Class 2 elements tested either because there is no autonomous element in the genome or because the autonomous element has been epigenetically silenced (4, 5).

The vast genetic array of Class 2 elements in maize has facilitated their exploitation as powerful genetic and molecular tools. McClintock first discovered the *Ds* element as a site of chromosome breakage in maize (6). This feature has been incorporated into the design of chromosomes that break in a predictable fashion for mosaic analysis (7). Similarly, the propensity of the *Ac* and *Ds*

elements to transpose preferentially to genetically linked sites led to a procedure to target insertions into genes linked to *Ac* or *Ds* elements. In this way, numerous mutant alleles were isolated for genetic (8, 9) and molecular (10–12) studies. The ability of several maize DNA elements to transpose preferentially into genic regions has been exploited for gene discovery and isolation by using both forward and reverse genetic strategies (13, 14).

In this study, we have exploited the unique properties of a recently described group of TEs called miniature inverted repeat transposable elements (MITEs) to develop a new class of molecular marker. MITEs were first discovered in association with the genes of several grass species, including maize (15, 16), rice (17), and barley (18). They are also abundant genomic components in nongrass species such as green pepper (19) and *Arabidopsis* (20, 21) and in several animal genomes including *Caenorhabditis elegans* (22, 23), insects (24), humans (22), and zebrafish (25). Although MITE families are numerous and diverse, all are distinguished by several structural features that are reminiscent of nonautonomous DNA elements (26). These include an absence of coding capacity and the presence of short TIRs (usually 10–14 bp). However, unlike most nonautonomous DNA elements, MITEs are small (≈100 to 500 bp) and have high copy number (≈1,000 to 15,000 per haploid genome) and a preference for insertion into 2- to 3-bp targets that are rich in A and T residues.

A recent study described the characterization of a maize MITE family called *Heartbreaker* (*Hbr*) (27). Unlike previously described MITEs, most of the 3,000–4,000 members of the *Hbr* family display over 90% sequence identity, suggesting that this family may have spread recently throughout the genome. Consistent with this view was the finding that *Hbr* insertion sites are highly polymorphic in maize and teosinte lines. Furthermore, *Hbr* insertion sites, like those of the much lower copy number *Ac* and *Mutator* elements, were found to be preferentially in the low copy regions of the maize genome. This result indicates a genic preference because low copy number regions may account for less than 20% of the maize genome.

Features of the *Hbr* family that made them of interest as molecular markers included their high copy number, DNA sequence identity, polymorphism, and genic preference. In this study, we have adapted the newly described technique of transposon display (TD) for use with the *Hbr* family. TD is a modification of the AFLP technique (28) that permits the simultaneous detection of many TEs from high copy number lines. Modified protocols have been used to analyze plant DNA

Abbreviations: LOD, logarithm of odds; TE, transposable element; TIR, terminal inverted repeat; MITE, miniature inverted repeat TE; RIL, recombinant inbred line; RFLP, restriction fragment length polymorphism; TD, transposon display; SSR, simple sequence repeat; cM, centimorgan.

§Present address: Monsanto Mystic Research, 62 Maritime Drive, Mystic, CT 06355.

¶To whom reprint requests should be addressed. E-mail: sue@dogwood.botany.uga.edu.

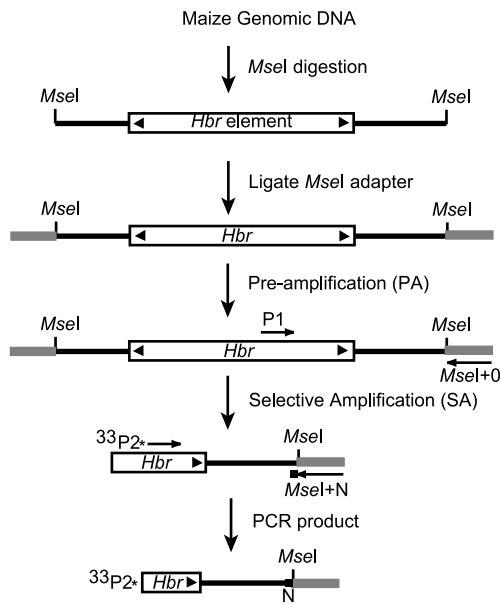


Fig. 1. Schematic of the *Hbr*-display protocol. The nested primers used in this study were P1 = *MseI* + 0/*Hbr*Int5-E and P2 = *MseI* + N/*Hbr*Int5-F (see *Materials and Methods*), where N is the selective base. P2 was labeled either with ^{33}P as shown or with a fluorescent tag. Black arrowheads represent the TIRs.

elements and retrotransposons including the *dTph1* family of petunia (29) and *BARE-1* of barley (30).

Materials and Methods

Plant Material and Genomic DNA Extraction. Two maize recombinant inbred populations were used in this study. One comprised 37 recombinant inbred lines (RILs) that were derived from a cross between inbreds CO159 × Tx303 (31). A larger population of 100 RILs was derived from a cross between inbreds B73 × Mo17 (32). *Hbr*-transformed and nontransformed rice strains (*Oryza sativa*) were obtained from C. Fauquet (Scripps Research Institute).

DNA was extracted from leaf tissue of single plants as described (33). The crude nucleic acid precipitates were suspended in TE (10 mM Tris-HCl, pH 8.0/1 mM EDTA), incubated with RNase at 37°C for 1 h, and quantified by fluorescence with a plate reader (Perkin-Elmer, model LS50B).

***Hbr* TD. DNA restriction and ligation of adapters.** Total genomic DNA (200–500 ng maize, 50 ng rice) was digested to completion for 3 h at 37°C in 40 μl containing 2 units *MseI* or *BfaI*, 5 mM DTT, 5 μg BSA, and 1× one-phor-all (OPA) buffer (100 mM Tris-acetate, pH 7.5/100 mM magnesium acetate/500 mM potassium acetate). Adapters (5'-GACGATGAGTCCTGAG and 5'-TACTCAGGACTCAT) were ligated to the digested DNAs by adding 10 μl of a mix containing 1× OPA buffer, 1.2 mM ATP, 5 mM DTT, 5 μg BSA, 50 pmol adapters, and 1 Weiss unit T4 DNA ligase and incubated for 3 h at 37°C. Because *MseI* and *BfaI* generate identical 3' overhangs, the same adapters were used in both ligations. Aliquots of the restriction/ligation reactions were visualized on 0.8% agarose gels to check the quality of DNA restriction and diluted 4-fold with 0.1× TE.

Preselective amplifications. PCRs were done by using a primer complementary to the adapters (*MseI* 5'-GACGATGAGTCCT-GAGTAA or *BfaI* 5'-GACGATGAGTCCTGAGTAA) and another primer (*Hbr*Int5-E) complementary to an internal *Hbr* element sequence (5'-GATTCTCCCCACAGCCAGATTC) (Fig. 1). For experiments done in radioactive format, reactions were performed in 50 μl containing 5 μl of the diluted restriction/ligation reactions, 12 pmol of each primer, 1× GeneAmp PCR buffer II

(Perkin-Elmer/ABI), 0.2 mM dNTPs, 2.5 mM MgCl_2 , and 1 unit AmpliTaq DNA polymerase (Perkin-Elmer/ABI). PCRs to be assayed in fluorescent format were done in 20 μl containing 3 μl of the diluted reactions, 8 pmol of each primer, 1× GeneAmp PCR buffer II, 0.2 mM dNTPs, 1.5 mM MgCl_2 , and 0.4 units AmpliTaq. These and subsequent reactions were carried out with a Robocycler Gradient Temperature Cycler (Stratagene), a Thermal Cycler Perkin-Elmer 480, or a PTC-100 Programmable Thermal Controller (MJ Research, Cambridge, MA). The temperature cycling parameters were as follows: 72°C/2 min; 94°C/3 min; 24 cycles of 94°C/30 sec, 59°C/30 sec, and 72°C/1 min, and a final cycle of 72°C/5 min. After visualizing aliquots of each PCR on 1.2% agarose gels stained with ethidium bromide, the remaining volumes were diluted 20-fold with 0.1× TE.

Selective amplification. Selective amplification for radioactive detection was performed in 20 μl containing 5 μl of the diluted preselective amplification products, 8 pmol of selective primer *MseI* + N, 1.25 pmol ^{33}P -labeled *Hbr*Int5-F (5'-GAGCCAGATTTCA-GAAAAGCTG), ×1 GeneAmp PCR buffer II, 0.2 mM dNTPs, 2.5 mM MgCl_2 , and 0.4 units AmpliTaq DNA polymerase. For the fluorescent assay, PCRs were as above except that either *MseI* + N or *BfaI* + N primers (4 pmol) were used in combination with 4 pmol of the *Hbr*Int5-F primer labeled with 6-FAM (Perkin-Elmer/ABI), and the MgCl_2 concentration was reduced to 1.5 mM. Temperature cycling used a “touchdown” protocol: 94°C/5 min, followed by 94°C/30 sec, 70°C/30 sec, and 72°C/1 min. In subsequent cycles, the annealing temperature was reduced from 69°C to 61°C in 1°C increments each cycle. Twenty-seven cycles were performed at the 61°C annealing temperature, followed by a final cycle of 72°C/5 min.

Gel electrophoresis. For radioactive detection, 20 μl of loading-denaturing buffer (98% deionized formamide/10 mM EDTA, pH 8.0/0.025% xylene cyanol/0.025% bromophenol blue) was added to the PCR reactions. Samples were denatured at 95°C/5 min and placed on ice, and 3 μl of the mixture was immediately loaded on 6% denaturing (7.5 M urea) acrylamide-bisacrylamide (19:1) gels in ×1 TBE buffer (89 mM Tris/89 mM borate/2 mM EDTA). The size standard, a 30- to 330-bp AFLP DNA ladder (GIBCO/BRL), was also denatured, and 2 μl was loaded in separate gel lanes. After samples were electrophoresed (35 mA constant) for 2 h, the gel was transferred to filter paper, dried, and exposed to an x-ray film for 24 h.

For detection in fluorescent format, samples containing 0.3 μl of the PCR products, 0.1 μl GeneScan 500 XL [TAMRA] internal lane size standard (Perkin-Elmer/ABI), and 1.6 μl of loading buffer (4 deionized formamide/1 blue dextran) were denatured at 95°C for 5 min and placed on ice, and 0.8 μl of the mixture was immediately loaded on 5% denaturing (6 M urea) acrylamide-bisacrylamide (19:1) gels in 1× TBE. Samples were electrophoresed (3,000 V for 3 h at 51°C in 1× TBE) on an automated DNA sequencer (Model 377, Perkin-Elmer/ABI).

Recovery of Gel Bands. Thirty-eight DNA fragments were excised from radioactive gels, eluted in buffer (0.5 M NH_4Oac /10 mM MgCl_2 /0.1% SDS/1 mM EDTA, pH 8.0), precipitated with ethanol, and suspended in 10 μl TE. Fragments were amplified in 50 μl containing 5 μl DNA, 12 pmol of each primer (*MseI* + N and *Hbr*Int5-F), 1× GeneAmp PCR buffer II, 0.2 mM dNTPs, 2.5 mM MgCl_2 , and 1 unit of AmpliTaq DNA polymerase. PCR was performed by using the “touchdown” cycling protocol described above. Reactions were resolved in 1% agarose gels, fragments excised, purified (QIAquick, Qiagen, Chatsworth, CA), cloned (TA cloning kit, Invitrogen), and DNA sequence data obtained by using fluorescent dye terminator chemistry and automated DNA sequencers (Model 373A, Perkin-Elmer/ABI).

Mapping of *Hbr*-Insertion Sites. All mapping data were collected in fluorescent format. Electropherograms were analyzed and DNA

fragments sized by using GENESCAN Ver. 2.1 software (Perkin-Elmer/ABI). Peak scoring was verified for each DNA sample, and a fragment presence/absence matrix was generated for both mapping populations by using GENOTYPER Ver. 2.5 (Perkin-Elmer/ABI). The frequency of nonparental bands was calculated by dividing the absolute number of nonparental fragments by the total number of bands scored in all progeny. Nonparental bands included both fragments scored in either parent but absent in the progeny and bands scored in the progeny but absent in either parent.

Linkage analysis of *Hbr* markers was performed with MAPMAKER Ver. 3.0 (34). To include a locus in a linkage group, a minimum logarithm of odds (LOD) threshold of 3.0 and a recombination fraction (rf) of 0.40 were used. Because the B73 × Mo17 recombinant inbred population was constructed by four rounds of intermating before establishment of selfed lines, rf values calculated by MAPMAKER were corrected as described (35). Three-point analyses followed by multipoint analyses were done to determine the putative order between the loci. This analysis produced 19 linkage groups that were placed into the standard 10 linkage groups by comparisons to published maps by using the MAIZEDB (<http://www.agron.missouri.edu/>).

A χ^2 test for goodness of fit was used to determine whether the *Hbr* markers were evenly distributed among the 10 maize chromosomes. The expected number of markers per chromosome was estimated by multiplying the total number of markers observed by the proportion of the total genetic length of the map (in centimorgans) represented by each chromosome.

Results

Development of *Hbr* Display. TD is a modification of the AFLP procedure (28) where PCR products are derived from primers anchored in a restriction site (i.e., *Bfa*I or *Mse*I) and a transposable element rather than in two restriction sites (29). For this study, candidate primers in *Hbr* elements were designed based on a consensus sequence generated from 35 elements randomly isolated from the maize genome (27). Initial amplifications included *Mse*I-digested maize or rice DNAs, the adapter primer, and a primer complementary to the *Hbr* TIR (Figs. 1 and 4). Rice genomic DNA was isolated from untransformed plants and from transgenic plants containing several copies of the *Hbr* element. In this way, the rice genomic DNA served as both positive (transformed) and negative (untransformed) controls.

Use of the *Hbr* TIR primer led to the amplification of multiple products in all samples, including the untransformed rice control (data not shown). Although the *Hbr* family is not present in rice (27), the rice genome contains many MITE families, including some with *Hbr*-like TIRs (17, 36). To further promote specific amplification of *Hbr*-containing fragments, “nested” PCR was used in preselective amplification (PA) and selective amplification (SA) reactions (Fig. 1). For the PA reaction, a primer located 60 bp from the *Hbr* TIR (*Hbr*Int5-E) was used in combination with an adapter primer without an additional nucleotide (*Mse*I + 0). The subsequent SA reaction involved a more terminal primer (*Hbr*Int5-F, located 30 bp from the end of the element) in combination with an adapter primer containing an additional nucleotide (*Mse*I + N). Use of this protocol led to the amplification of products in both maize and transformed rice but not in untransformed rice (Fig. 2).

To test the reproducibility of this procedure, DNA was extracted three times from the same plant material and separate restriction and ligations, preselective, and selective amplifications were performed with each preparation. For all replicates, virtually identical banding patterns were observed in both radioactive and fluorescent detection systems.

Mapping Strategy. TD was developed with several goals in mind. The first objective was to use TD as a rapid screen for newly

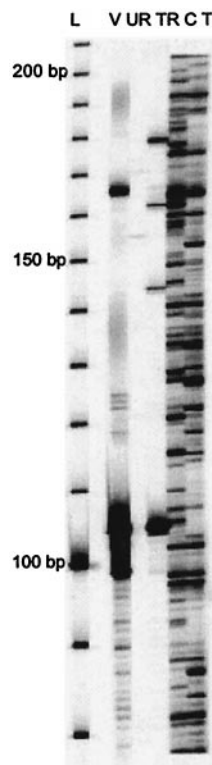


Fig. 2. Autoradiograph of *Hbr* display with rice and maize DNAs. *Hbr* display by using primers *Mse*I + A and *Hbr*Int5-F for selective amplification (see *Materials and Methods*). L, 30- to 330-bp molecular weight ladder; V, vector containing *Hbr* element; UR, untransformed rice; TR, *Hbr*-transformed rice; C and T, parental inbred lines CO159 and Tx303, respectively.

transposed MITEs. The second was to rapidly determine the map positions of hundreds of *Hbr* elements and, in the process, develop a new class of molecular marker. Recombinant inbred mapping populations were chosen for *Hbr*-TD because high frequencies of nonparental bands in the RILs might indicate new transpositions. However, if nonparental bands were rare, then polymorphic parental bands could be mapped after segregation analysis in the RILs.

PCR Products Are Anchored in *Hbr* Elements. *Hbr* display was initially tested in a small mapping population derived from a cross between maize inbreds CO159 × Tx303 (31). An example of TD from the parents and 4 of the 38 RILs in this population is shown in Fig. 3A. To confirm that the fragments were anchored in *Hbr* elements, 38 bands were recovered, reamplified, cloned, and sequenced. Bands were chosen from each of the four *Mse*I selective amplifications (*Hbr*Int5-F/*Mse*I + A, G, C, and T) and included 30 polymorphic fragments and 4 pairs of comigrating (monomorphic) parental bands.

Similar to the partial DNA sequences shown in Fig. 4, all 38 clones contained the *Hbr* TIR sequence adjacent to the *Hbr*Int5-F primer, thus confirming that all were anchored in *Hbr* elements. The DNA sequences of three comigrating fragments from CO159, each amplified with a different selective primer set (*Hbr*Int5-F/*Mse*I + A, G or T), had unique flanking sequences and contained the appropriate 3' selective base (data not shown). These data demonstrate that the addition of specific nucleotides to the 3' end of the *Mse*I primer resulted in the amplification of different *Hbr*-containing fragment sets. In contrast, the DNA sequences were identical for three of four monomorphic fragment pairs assayed. The fourth pair differed both in sequence and length (by 1 bp), indicating that comigrating fragments had not been resolved and isolated.

Genomic Distribution of *Hbr*-Anchored Fragments. To determine the genomic distribution of *Hbr* elements and to assess the utility of TD fragments as a new molecular marker class, display fragments were

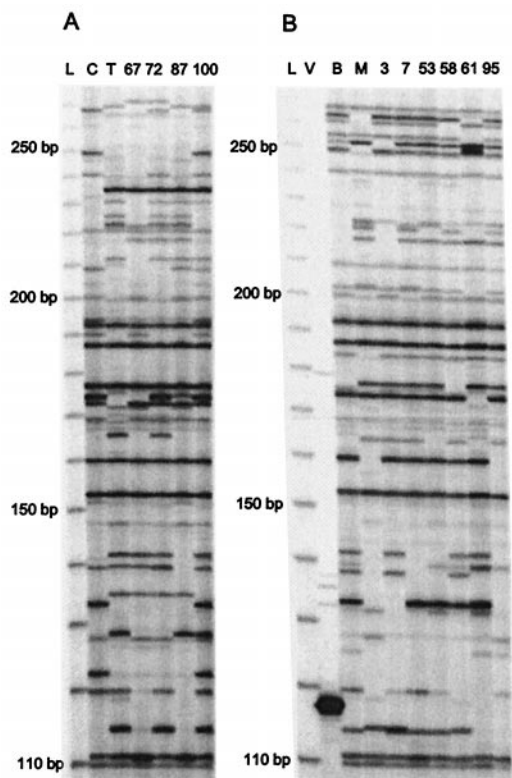


Fig. 3. Autoradiograph of *Hbr* display by using two mapping populations. (A) CO159 × Tx303, where C and T are the parental inbred lines, and 67, 72, 87, and 100 are select RILs. (B) B73 × Mo17, where B and M are the parental inbred lines, and 3, 7, 53, 58, 61, and 95 are select RILs. L, 30–330 bp size standard; V, vector containing *Hbr* element.

mapped in a much larger population of 100 RILs derived from a cross between the maize inbreds B73 × Mo17 (32). This analysis was done by using semiautomated fluorescence-based detection because the sensitivity, band resolution, and sizing precision were judged to be superior to the radioactive format.

A total of 418 fragments (65–450 bp) from 8 primer combinations (223 from *Hbr5-F/MseI* + N and 195 from *Hbr5-F/BfaI* + N) were unambiguously scored in the parental lines. The number of fragments amplified and the polymorphisms detected for each primer combination are summarized in Table 1. Among all primer combinations, the number of amplified fragments ranged from 42 (*HbrInt5-F/BfaI* + C) to 77 (*HbrInt5-F/MseI* + A). Overall, 252 of the 418 fragments, or 60.3%, were polymorphic and could be assigned to specific maize chromosomes by determining linkage relationships to 282 previously mapped restriction fragment length polymorphism (RFLP) markers (<http://www.intl-pag.org/pag/7/abstracts/pag7605.html>). Of these 252 polymorphic markers, 213

Table 1. Polymorphism detected in the B73 × Mo17 mapping population

Primer/enzyme combination	Number of amplified fragments			% polymorphic
	Monomorphic	Polymorphic	Total	
<i>Hbr-BfaI</i> +A	26	38	64	59.4
<i>Hbr-BfaI</i> +C	16	26	42	61.9
<i>Hbr-BfaI</i> +G	16	27	43	62.8
<i>Hbr-BfaI</i> +T	23	23	46	50.0
Total	81	114	195	58.5
<i>Hbr-MseI</i> +A	37	40	77	51.9
<i>Hbr-MseI</i> +C	13	33	46	71.7
<i>Hbr-MseI</i> +G	15	28	43	65.1
<i>Hbr-MseI</i> +T	20	37	57	64.9
Total	85	138	223	61.9

(111 from Mo17 and 102 from B73) were assigned to chromosomes (Fig. 5). These will be referred to as “*Hbr* markers.” Results from χ^2 tests showed that the *Hbr* markers were evenly distributed in the maize genome [$\chi^2 = 2.5$, not significant (n.s.)]. When considered separately, both the *BfaI* ($\chi^2 = 4.8$, n.s.) and *MseI* ($\chi^2 = 6.3$, n.s.) markers were also evenly distributed. The *Hbr* markers, as a whole, mapped with an average LOD score of 43.6. The LOD scores for individual markers were greater than 8.0, except for one marker (*mHbrMT96*) that mapped to the telomeric region of chromosome 4 (LOD = 3.75).

The total length of the genetic map was 1,092 cM. Addition of *Hbr* markers to the maize RFLP map increased the genetic map length by 150 cM and reduced the overall distance between markers. Previously, there were 51 regions of the RFLP map where the recombination fraction was greater than 0.3. Inclusion of the *Hbr* markers reduced the number of these regions to 37.

No recombination was detected among 24 sets of markers (comprising a total of 50 markers or 23.5% of the 213 *Hbr* markers mapped) (Fig. 5). In most instances (79%), complete linkage was observed between a pair of *MseI* and *BfaI* markers. This result would be expected if these markers include the same *Hbr* insertion site. In a minority of cases, complete linkage involved either two *BfaI* markers (8.3%) or one *MseI* and two *BfaI* markers (12.5%). Because these markers were not sequenced, it is not known whether linkage in these cases was because of the same *Hbr* insertion or to tightly linked elements.

Segregation ratios that departed from the Mendelian expectation of 1:1 were detected for 49 *Hbr* markers (19.5%; 26 in Mo17, 23 in B73). Thirty-nine distorted markers were mapped (Fig. 5). In general, these markers were not evenly distributed across all maize chromosomes, but were clustered on chromosomes 1, 2, 4, and 8. Finally, 39 *Hbr* markers (15.5%) were not mapped either because of low LOD scores (<3) or because linkage to the framework RFLP markers was not detected.

To determine whether polymorphic bands that comigrated in the two mapping populations (B73 × Mo17 and CO159 × Tx303) also comapped, the map positions of comigrating *Hbr* markers that were polymorphic in both mapping populations were compared. For this experiment, markers were mapped from only the two primer combinations (*HbrInt5-F/MseI* + C and *HbrInt5-F/MseI* + G) that yielded the highest number of polymorphic fragments in the B73 × Mo17 mapping population (Table 1). Fourteen *Hbr* markers were common between the two mapping populations. The chromosomal locations matched for 13 markers (93%), whereas 1 mapped to a different location.

Nonparental Inheritance. TD bands present in one or more progeny but absent in both parental inbreds or monomorphic between parents but missing in the progeny were defined as displaying nonparental inheritance. Seven RILs from the B73 × Mo17 mapping population yielded a large number of markers showing

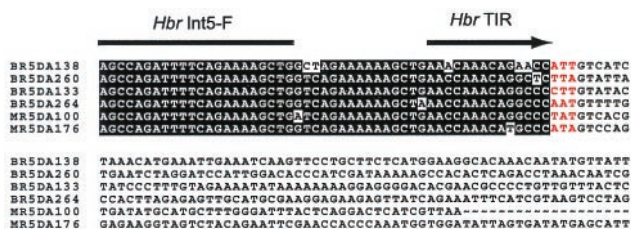


Fig. 4. DNA sequences of 6 of 38 markers recovered from *Hbr*-display gels. The positions of the primer (*HbrInt5-F*) and the *Hbr* TIR are indicated. Shown in red are the 3-bp host sequences presumably duplicated on *Hbr* insertion.

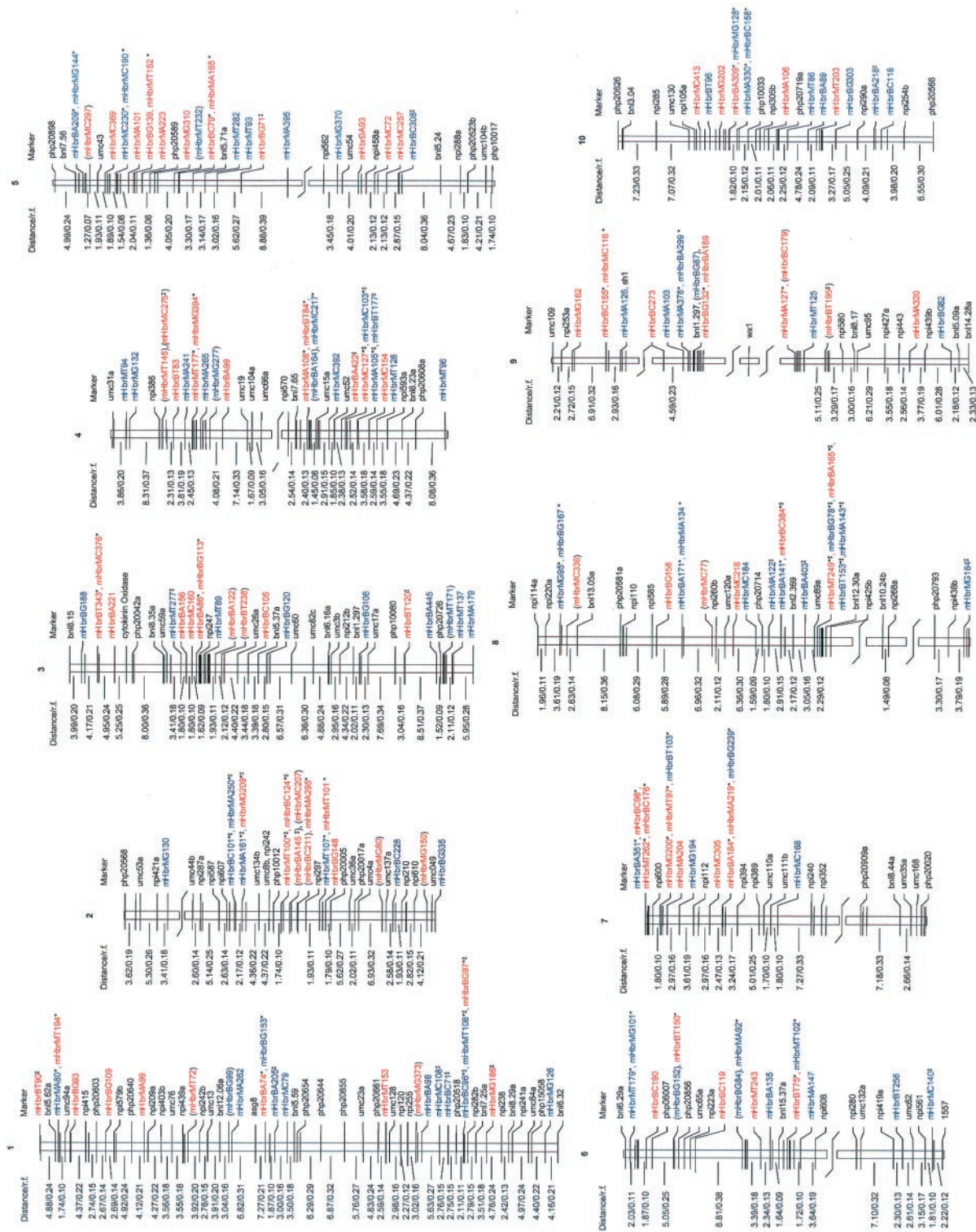


Fig. 5. Genetic map of maize with *Hbr* and RFLP markers. Two hundred and thirteen *Hbr* markers were assigned to the ten maize chromosomes by using a previously established map of 282 RFLP markers (not all RFLP markers are shown). The total genetic length of the map was 1,092 cM. *Hbr* markers are in blue (from Mo17) and red (from B73) and red (from B73) and red (from Mo17). *Hbr* markers in parentheses were completely linked to other *Hbr* markers (only one is shown). Those marked with an asterisk were separated by less than 1.2 cM. Markers showing segregation distortion are indicated by a double-plus sign. For each *Hbr* marker, the following information has been provided: DNA element and respective family from which the marker derived (mHbr = MiteHbrtbreaker); the restriction enzyme used (M and B for fragments generated by *Mse*I and *Bfa*I, respectively); the selective base used (A, C, G, or T), and the size of the fragment in base pairs. Numbers to the left of each chromosome represent distances (in centimorgans) and recombination fractions, respectively.

nonparental inheritance for almost every primer pair combination tested (data not shown). The TD of one of these RILs is shown (Fig. 3B, line 61). These same seven RILs also showed excessive nonparental bands when analyzed for the segregation of either RFLP or SSR markers (J. Arbuckle, personal communication), indicating that outcrossing probably occurred at some point during their development. If the aberrant lines are not considered, only 31 fragments of 12,471 data points showed nonparental inheritance (0.25%) for all 8 primer combinations tested.

Discussion

In this study, the unique features of *Hbr* elements, including their high copy number and polymorphism, were exploited in the development of a new class of molecular marker for maize. Over 200 *Hbr* markers were generated and mapped to loci that were evenly distributed throughout the maize genome. Most importantly, most of these markers are probably near genes given the genic preference of sequences flanking randomly chosen *Hbr* elements (27).

Given the widespread occurrence of MITE families with similar characteristics in plant and animal species, this study was undertaken to evaluate the utility of MITE markers and to develop a set of protocols that could be routinely adapted to other families. The applicability of this methodology to other organisms is discussed below, along with an assessment of how *Hbr* markers compare with other marker systems in maize.

Applicability to Other MITE Families. Unlike the universal primers that are available for AFLP, TD requires specific primers for each element family. The *Hbr* family was chosen as the prototype for evaluating MITE markers because element consensus sequences that are required for TD could be easily deduced from the *Hbr* sequences obtained in a previous study (27). Unfortunately, these primers are of little use for other MITE families in maize or even in closely related grasses such as sorghum, which has no identifiable *Hbr* elements (27).

Despite these limitations, we have successfully applied TD to seven other MITE families: three in maize (*Hb2*, *mPIF*, and *Tourist*) (Z. Magbanua, X. Zhang, N. Jiang, and S.R.W., unpublished data) and four in rice (*Gaijin*, *Tourist*, *Olo*, *Ditto*) (A.N. and S.R.W., unpublished data). In all cases, family-specific consensus sequences were easily derived from large numbers of elements downloaded from public databases. The availability of even a modest database for a particular organism is usually sufficient to identify MITE families with the desirable characteristics for TD, namely, high copy number and high within-family sequence identity. For organisms for which little or no sequence is available, biochemical approaches, such as fractionation by hybridization rate, may provide a source of MITE-enriched DNA.

***Hbr*-Markers vs. Other Maize Marker Systems. Level of polymorphism.** Over 60% of the *Hbr*-anchored fragments (generated by using two restriction enzymes) were polymorphic in the B73 × Mo17 RILs. This level of polymorphism compares favorably to that observed for SSRs (53–69%) (37, 38) and RFLPs (50–80%) (39, 40). It is higher than values reported for AFLP markers assayed in maize mapping populations (26–41%) (41, 42).

The level of polymorphism of MITE markers is a reflection of many factors, some that are species specific (e.g., the extent of restriction site polymorphism) and others that are MITE family specific. In the latter case, the extent of polymorphism is a reflection of when each family spread through the population. That is, families that are still active or recently active will display higher levels of polymorphism than families active in the more distant past. However, because high sequence identity also correlates with recent amplification, it is anticipated that most families displaying high sequence identity will also be highly polymorphic in mapping populations.

Mapping and distribution of markers. The ability to map markers with confidence is a function of the reproducibility of the protocol and the ability to unambiguously score segregating bands. In this regard, *Hbr* markers were highly robust as 84% (213:252) of the polymorphic markers could be assigned to 1 of the 10 maize chromosomes with a high level of confidence. However, like the parent AFLP technique, the ability to score MITE markers is determined, in part, by the number of amplified fragments displayed in each lane. Adding selective bases to the restriction site primer can reduce this in turn. For the 3,000 to 4,000 members of the *Hbr* family, addition of one selective base was found to be optimal. In contrast, two selective bases were required to clearly resolve subsets of the ≈12,000 members of the maize *Hb2* family (Z. Magbanua and S.R.W., unpublished data).

The distribution of markers in the genome has important implications for the general applicability and utility of the marker class (43). Randomly distributed markers are desirable as they provide for maximum genome coverage. Both the *MseI*- and *BfaI*-derived *Hbr*-markers were evenly distributed, both among and within maize chromosomes (Fig. 5). In contrast, AFLP fragments produced by enzymes with AT-rich recognition sequences (i.e., *EcoRI* and *MseI*) tend to cluster in centromeric and pericentromeric regions, whereas markers generated with enzymes having recognition sequences rich in G and C residues show a more random genomic distribution (41, 42, 44). Centromeric regions of plant chromosomes consist of repetitive sequences enriched in A and T residues and are thought to be relatively gene poor (45, 46). That *Hbr* elements, unlike many other repeat families, do not cluster in centromeric regions is consistent with the previously observed *Hbr* insertional preference for nonrepetitive genic regions of maize (27).

The distribution of *Hbr* markers may not be characteristic or even typical of other MITE families. In *C. elegans*, where MITEs account for 1–2% of the genome, each of four MITE families was found to have a distinct distribution (*Cele1*, *Cele2*, *Cele14*, and *Cele42*) (23). Although *Cele2* elements were evenly distributed on all autosomes, *Cele14* elements clustered near the ends of all six chromosomes. Preliminary analysis of the maize *Hb2* family indicates that almost half of the 550 mapped elements were distributed throughout the 10 maize chromosomes, whereas the other half were in several large clusters (Z. Magbanua and S.R.W., unpublished data).

A comparison of MITE distribution in maize and *C. elegans* needs to be made with caution. The distributions of the *Cele* family members were derived from the whole genome sequence and as such account for virtually all family members. In contrast, distributions derived from TD are restricted to polymorphic family members. Whether this subset of elements is representative of the entire family is not known at this time.

Segregation distortion and nonparental inheritance. Segregation distortion (deviation from the expected 1:1 Mendelian ratio) is commonly observed when mapping molecular markers in plants (47). In this study, 19.5% of *Hbr* markers in the B73 × Mo17 RILs showed distorted segregation. For the majority of these markers, however, linked RFLP markers also showed non-Mendelian segregation (C.B., unpublished observation).

Much less common than segregation distortion is the occurrence of nonparental bands. The frequency of nonparental *Hbr*-anchored fragments was low, ranging from 0.2% to 2.5% depending on the population and enzyme/primer combination assayed. Similar frequencies have been observed for SSR and RFLP markers in maize. For example, in sets of maize “triplets” (i.e., two parental lines and the derived hybrid), 2% to 5% of SSRs and 3% of RFLP loci exhibited alleles of nonparental origin (48). For the framework RFLP markers in the B73 × Mo17 population, nonparental bands occurred at a frequency of 0.23% (C.B., unpublished data). To our knowledge, similar data for AFLPs in maize have not been reported.

The nonparental *Hbr* fragments in the two mapping popula-

tions could originate from a number of sources, including residual heterozygosity in one or both of the parental lines, pollen contamination (outcrossing events), sequence variation at flanking restriction sites or the internal *Hbr* primer binding site, or genomic rearrangement. Residual heterozygosity is commonly observed in inbred lines of maize (31). Therefore, it is possible that some of the variation originally present in the parental lines was lost over subsequent generations of inbreeding but maintained in selected progeny. It is also probable that mutations have occurred within a subset of the restriction sites. Based on data from 311 expressed sequence tag loci, base substitutions occur once every 80 bp on average in inbred lines of maize (49). This polymorphism level is 10-fold higher than for humans, where base substitutions average 1 in 1,000 bp (50).

An additional source of variation for transposon-anchored markers is, of course, transposition. The fact that nonparental *Hbr* markers occurred at approximately the same frequency as other molecular markers indicates that the *Hbr* family is not active in these genetic backgrounds. This result is not surprising, because no strains have been reported to date in plants or animals that support MITE transposition. However, if one of the parents did harbor mobile elements, activity would be easy to discern by the appearance of an unusually high number of new bands or the loss of parental bands in the progeny.

Concluding Remarks. In summary, we have developed a class of molecular marker that should prove useful in a number of applications in maize and in other organisms where MITEs are found.

Hbr markers are highly polymorphic among inbred lines and evenly distributed in the maize genome and, like AFLP markers, large numbers can be generated easily and displayed in a semiautomated fashion. The data are extremely reproducible both within and between mapping populations, and fragments are easily recovered for possible conversion into sequence-based markers. The technique is cost effective, as limited numbers of primers (both generic and MITE specific) are required. Although MITE primers are not universal like AFLP primers, it should not be difficult to identify primers after database searches. It should also be possible to increase the number of markers displayed per lane by multiplexing the products of two or three MITE families that have each been amplified with primers containing distinct fluorescent tags.

Given the plethora of available marker systems, the major advantage of *Hbr* markers, and perhaps most MITE markers, is a preference for genic regions (18, 27). The ability to rapidly screen large numbers of MITE markers may expedite chromosome walks and facilitate map-based cloning protocols in the larger genomes of agriculturally important plants like maize, wheat, and barley.

We thank Ben Burr (Brookhaven Labs, Long Island) for providing the CO159 × Tx303 family and assisting with the mapping of *Hbr* markers in this population, Michael Lee (Iowa State University, Ames, IA) for providing the B73 × Mo17 family, Susan McCouch, Molly Jahn, and Sharon Mitchell for critical reading of the manuscript, and Jim Register and John Arbuckle for their assistance and access to unpublished information. This study was supported by grants from National Institutes of Health (to S.R.W.) and Pioneer (to S.K. and S.R.W.).

- McClintock, B. (1948) *Carnegie Inst. Wash. Year Book* **47**, 155–169.
- Finnegan, D. J. (1989) *Trends Genet.* **5**, 103–107.
- Wessler, S. R. (1988) *Science* **242**, 399–405.
- Fedoroff, N. V. & Chandler, V. (1994) in *Homologous Recombination and Gene Silencing in Plants*, ed. Paskowski, J. (Kluwer, Dordrecht, The Netherlands), pp. 349–385.
- Gutiérrez-Nava, M. L., Warren, C. A., Leon, P. & Walbot, V. (1998) *Genetics* **149**, 329–346.
- McClintock, B. (1947) *Carnegie Inst. Wash. Year Book* **46**, 146–152.
- Dawe, R. K. & Freeling, M. (1990) *Dev. Biol.* **142**, 233–245.
- Brink, R. A. & Williams, E. (1973) *Genetics* **73**, 273–296.
- Kermicle, J. L. (1980) *Science* **108**, 1457–1459.
- Athma, P., Grotewold, E. & Peterson, T. (1992) *Genetics* **131**, 199–209.
- Moreno, M. A., Chen, J., Greenblatt, I. & Dellaporta, S. L. (1992) *Genetics* **131**, 939–956.
- Weil, C. F., Marillonnet, S., Burr, B. & Wessler, S. R. (1992) *Genetics* **130**, 175–185.
- Bensen, R. J., Johal, G. S., Crane, V. C., Tossberg, J. T., Schnable, P. S., Meeley, R. B. & Briggs, S. P. (1995) *Plant Cell* **7**, 75–84.
- Das, L. & Martienssen, R. (1995) *Plant Cell* **7**, 287–294.
- Bureau, T. E. & Wessler, S. R. (1992) *Plant Cell* **4**, 1283–1294.
- Bureau, T. E. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1411–1415.
- Bureau, T. E., Ronald, P. C. & Wessler, S. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8524–8529.
- Shirasu, K., Schulman, A. H., Lahaye, T. & Schulze-Lefert, P. (2000) *Genome Res.* **10**, 908–915.
- Pozueta-Romero, J., Houlne, G. & Schantz, R. (1996) *Gene* **171**, 147–153.
- Casacuberta, E., Casacuberta, J. M., Puigdomenech, P. & Monfort, A. (1998) *Plant J.* **16**, 79–85.
- Surzycki, S. A. & Belknap, W. R. (1999) *J. Mol. Evol.* **48**, 684–691.
- Oosumi, T., Belknap, W. R. & Garlick, B. (1995) *Nature (London)* **378**, 672.
- Surzycki, S. A. & Belknap, W. R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 245–249.
- Tu, Z. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 7475–7480.
- Izsvak, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H. & Hackett, P. B. (1999) *J. Mol. Evol.* **48**, 13–21.
- Wessler, S. R., Bureau, T. E. & White, S. E. (1995) *Curr. Opin. Genet. Dev.* **5**, 814–821.
- Zhang, Q., Arbuckle, J. & Wessler, S. R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1160–1165.
- Vos, P., Hogers, R., Bleeker, M., Reijmans, M., van de Lee, T., Hornes, M., Frijters, B. A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. (1995) *Nucleic Acids Res.* **23**, 4407–4414.
- Van den Broeck, D., Maes, T., Sauer, M., J., Z., De Keukeleire, P., D'Hauw, M., Van Montagu, M. & Gerats, T. (1998) *Plant J.* **13**, 121–129.
- Kalendar, R., T., G., Regina, M., Suoniemi, A. & Schulman, A. (1999) *Theor. Appl. Genet.* **98**, 704–711.
- Burr, B., Burr, F. A., Thompson, K. H., Albertson, M. C. & Stuber, C. W. (1988) *Genetics* **118**, 519–526.
- Austin, O. F. & Lee, M. (1996) *Theor. Appl. Genet.* **92**, 817–826.
- Ferreira, M. E. & Grattapaglia, D. (1995) *Introducao ao uso de marcadores RAPD e RFLP em analise genetica* (Embrapa-Cenargen Documento 20) (Lumma Consultoria Projetos e Informatica, Brasilia, Brasil), pp. 125–130.
- Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. & Newburg, L. (1987) *Genomics* **1**, 174–181.
- Liu, S. C., Kowalski, S. P., Lan, T. H., Feldmann, K. A. & Paterson, A. H. (1996) *Genetics* **142**, 247–258.
- Tarchini, R., Biddle, P., Wineland, R., Tingey, S. & Rafalski, A. (2000) *Plant Cell* **12**, 381–391.
- Taramino, G. & Tingey, S. (1996) *Genome* **39**, 277–287.
- Senior, M. L., Chin, E. C. L., Lee, M., Smith, J. S. C. & Stuber, C. W. (1996) *Crop Sci.* **36**, 1676–1683.
- Beavis, W. D. & Grand, D. (1991) *Theor. Appl. Genet.* **82**, 636–644.
- Gardiner, J. M., Coe, E. H., Melia-Hancock, S., Hoisington, D. A. & Chao, S. (1993) *Genetics* **134**, 917–930.
- Castiglioni, P., Ajmone-Marsan, P., van Wijk, R. & Motto, M. (1999) *Theor. Appl. Genet.* **99**, 425–431.
- Vuylsteke, M., Mank, R., Antonise, R., Bastiaans, E., Senior, M. L., Stuber, C. W., Melchinger, A. E., Lubberstedt, T., Xia, X. C., Stam, P., et al. (1999) *Theor. Appl. Genet.* **99**, 921–935.
- Waugh, R., McLean, K., Flavell, A. J., Pearce, S. R., Kumar, A., Thomas, B. T. & Powell, W. (1997) *Mol. Gen. Genet.* **253**, 687–694.
- Young, W. P., Schupp, J. M. & Keim, P. (1999) *Theor. Appl. Genet.* **99**, 785–790.
- Dong, F., Miller, J. T., Jackson, S. A., Wong, G.-L., Ronald, P. C. & Jiang, J. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8135–8140.
- Miller, J. T., Dong, F., Jackson, S. A., Song, J. & Jiang, J. (1998) *Genetics* **150**, 1615–1623.
- Jenczewski, E., Gherardi, M., Bonnin, I., Prosperi, J. M., Olivieri, I. & Huguet, T. (1997) *Theor. Appl. Genet.* **94**, 682–691.
- Chin, E. C. L., Senior, M. L., Shu, H. & Smith, J. S. C. (1996) *Genome* **39**, 866–873.
- Bhattaramakki, D., A., C., Dolan, M., Register, J., Tingey, S. & Rafalski, A. (2000) *Maize Genetics Cooperation Newsletter* **74**, 54.
- Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al. (1998) *Science* **280**, 1077–1082.