



## Methylation-Spanning Linker Libraries Link Gene-Rich Regions and Identify Epigenetic Boundaries in *Zea mays*

Yinan Yuan, Phillip J. SanMiguel and Jeffrey L. Bennetzen

*Genome Res.* 2002 12: 1345-1349

Access the most recent version at doi:[10.1101/gr.185902](https://doi.org/10.1101/gr.185902)

---

### References

This article cites 28 articles, 8 of which can be accessed free at:  
<http://genome.cshlp.org/content/12/9/1345.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, and the Cellecta logo, which consists of a green molecular structure and the word "CELLECTA" below it.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# Methylation-Spanning Linker Libraries Link Gene-Rich Regions and Identify Epigenetic Boundaries in *Zea mays*

Yinan Yuan,<sup>1</sup> Phillip J. SanMiguel,<sup>2</sup> and Jeffrey L. Bennetzen<sup>1,3</sup>

<sup>1</sup>Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA; <sup>2</sup>Genomics Center, Purdue University, West Lafayette, Indiana 47907, USA

Complex cereal genomes are largely composed of small gene-rich regions intermixed with 5 kb to 200 kb blocks of repetitive DNA. The repetitive DNA blocks are usually 5-methylated at 5'-CG-3' and 5'-CNG-3' cytosines in most or all adult tissues, while the genes are generally unmethylated at these sites. We have developed methylation-spanning linker library (MSLL) technology as a tool to span large methylated DNA blocks and thereby link unmethylated genic regions. MSLL clones contain insertions of large fragments that are size fractionated over gels after complete digestion of total genomic DNA with restriction enzymes that are sensitive to the 5-methylation of cytosines in 5'-CG-3' and 5'-CNG-3' sequences. Our data indicate that the end sequences of maize MSLL clones are greatly depleted in repetitive DNAs and enriched in genes relative to total genomic DNA. Combined with other gene-enrichment approaches, MSLL technology can efficiently generate fully-linked contiguous sequences in complex genomes that are resistant to shotgun sequencing.

Large grass genomes, including those of barley, maize, and wheat, are mostly comprised of 5–20 kb blocks of genes intermixed with repetitive DNA blocks that range in size from a few kb up to more than 100 kb (SanMiguel et al. 1996; Panstruga et al. 1998; Tikhonov et al. 1999; Dubcovsky et al. 2001; Wicker et al. 2001). In a few cases, including within tandem gene families or the rare unrelated gene cluster (Llaca and Messing 1998; Fu et al. 2001), gene-rich regions may extend for 50 kb or more. In most cases, however, gene-rich chromosome segments only contain 1–4 genes in a region of 20 kb or less. The repetitive DNAs found in the intermixed repeat blocks are usually nested insertions of a class of mobile DNAs called long terminal repeat- (LTR-) retrotransposons (SanMiguel et al. 1996; Llaca and Messing 1998; Panstruga et al. 1998; Kumar and Bennetzen 1999; Tikhonov et al. 1999; Dubcovsky et al. 2001; Fu et al. 2001; Wicker et al. 2001). These nested LTR-retrotransposons can make up well over 50% of total genomic DNA, with most of this DNA coming from only a handful of different element families that have copy numbers of several thousand per nuclear genome (SanMiguel and Bennetzen 1998; Kumar and Bennetzen 1999; Vicent et al. 1999; Meyers et al. 2001). The LTR-retrotransposons are relatively large in size (usually greater than 5 kb) and can have numerous copies that are < 99% identical within the same genome (W. Ramakrishna, P. SanMiguel and J. Bennetzen, unpubl. obs.). Hence shotgun sequencing of complex grass genomes would not yield information that can be converted into long contiguous sequences (Bennetzen et al. 2001). Nuclear genomes in higher plants contain extensive 5-methylation of cytosine residues, much of it associated with 5'-CG-3' and 5'-CNG-3' sequences (Gruenbaum et al. 1981a). In many animals and plants, cytosine methylation is associated with heterochromatic regions, where it apparently contrib-

utes to the transcriptional inactivity of any sequences within the condensed chromatin. In maize, studies indicate that most of this cytosine methylation is associated with repetitive DNAs, including the LTR-retrotransposons. In adult tissues, most LTR-retrotransposons appear to be 100% methylated at all 5'-CG-3' and 5'-CNG-3' sites, while genes appear to be unmethylated at these same sites (Gruenbaum et al. 1981b; Antequera and Bird 1988; Bennetzen et al. 1994). This lack of genic methylation differs somewhat from that observed in mammals (Bird 1986; Frank et al. 1991), for instance, because the absence of methylation is even found in genes that are not expressed in the tissues that were the source of the DNA that was characterized (Bennetzen et al. 1994). There are likely to be exceptions to this general rule (Jacobsen and Meyerowitz 1997), but overall it appears that most genic regions can be separated from most LTR-retrotransposon blocks by this difference in DNA methylation. Perhaps most interesting of all, the size of the methylated DNA blocks has an upper limit of 200 kb or less (Springer 1992; Bennetzen et al. 1994), suggesting that an 'open' region of chromatin is needed at this spacing to allow some essential nuclear function, such as the initiation of DNA replication or chromosome folding.

Martienssen and coworkers have used the difference in DNA methylation between repetitive and genic DNA as a tool to efficiently sequence gene-rich regions of the maize genome by a shotgun approach (Rabinowicz et al. 1999). In their methyl filtration technology, the insertion of sheared fragments of total genomic DNA into a plasmid vector is followed by transformation of this library into an *Escherichia coli* strain that will not tolerate 5-methylation of cytosines in the cloned DNA. In maize, this approach yielded a greater than two-fold enrichment for genic sequences and at least a six-fold depletion of known LTR-retrotransposon sequences, relative to the same library inserted into a methylation-tolerant *E. coli* host (Rabinowicz et al. 1999; Meyers et al. 2001). Application of methyl filtration technology to the full maize genome should yield contiguous sequences (contigs) for the genic regions,

### <sup>3</sup>Corresponding author.

**E-MAIL** [maize@bilbo.bio.purdue.edu](mailto:maize@bilbo.bio.purdue.edu); **FAX (765) 496-1496**.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.185902>.

varying in size from a few kb up to a few dozen kb. However, this filtration technique does not localize the genic contigs relative to each other or to the maize genetic map. We have developed an approach that we call methylation-spanning linker library (MSLL) technology that overcomes this deficiency and also isolates the boundaries between methylated and unmethylated regions.

## RESULTS AND DISCUSSION

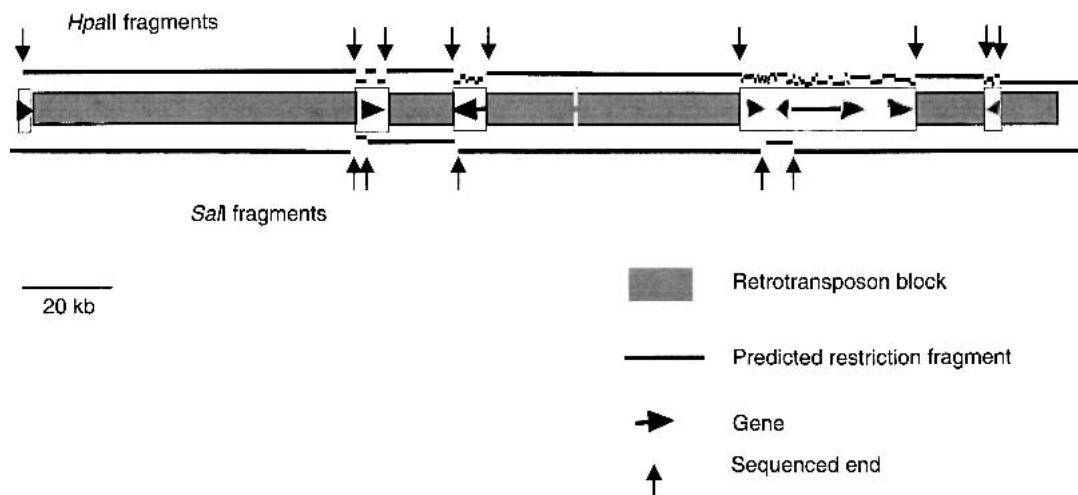
Figure 1 depicts the structure of a contiguous 225 kb of maize nuclear DNA that contains the *adh1-F* locus, as derived from the sequence of Tikhonov and coworkers (Tikhonov et al. 1999). Subsequent studies have shown that this is the standard structure for most or all gene-containing portions of the genome of maize and other large-genome cereals such as wheat and barley (Edwards et al. 1992; Llaca and Messing 1998; Dubcovsky et al. 2001; Fu et al. 2001; Wicker et al. 2001). These genic regions contain small islands of genes separated from each other by relatively large blocks of repetitive DNA. Because these repetitive DNA blocks are cytosine 5-methylated at all 5'-CG-3' and 5'-CNG-3' sites, they will not be digested by restriction enzymes that are inhibited by this type of DNA methylation. The lines in Figure 1 show the fragments that would be generated by complete digestion with *SalI* or *HpaII*, two enzymes that are cytosine-methylation sensitive.

Because *HpaII* has a 4 bp recognition/cleavage sequence, it should cut maize DNA, which has an approximate 50% GC content (Hake and Walbot 1980), an average of once every 4<sup>4</sup> (256) bp. However, from the analysis of bulk maize sequence data (Meyers et al. 2001), we found that the maize genome is about 53% AT, also observing that 5'-CG-3' and 5'-CNG-3' are somewhat underrepresented in the maize genome (a respective 4.6% and 5.3%, compared to the predicted 5.6%). In accord with the depletion of these bases, the actual frequency of *HpaII* sites that we detected in the bulk sequence data generated by Meyers et al. (2001) was once in every 305 bp. How-

ever, pulsed field gel analysis of *HpaII*-digested maize DNA indicates that the majority of the genome yields fragments that are larger than 50 kb (Springer 1992; Bennetzen et al. 1994). Hence isolation of *HpaII* fragments larger than a few kb should yield segments of DNA that contain internal methylated and repetitive sequences, while the ends are anchored in the unmethylated sequences associated with genic regions. Sequencing the ends of such fragments will mark adjacent genic regions, and provide clones that contain the sequences between those two genic regions. *SalI* can be used in the same manner, although its 6 bp specificity indicates that it will digest (on average) several kb away from the site at which a methylated DNA block begins.

To test this theory, we constructed three small BAC libraries. One library contained 9–14 kb *HpaII* fragments in the vector pBeloBAC11 (Kim et al. 1996). The fragments were generated by complete digestion of B73 maize genomic DNA with *HpaII*, followed by pulsed field gel electrophoresis of the digested DNA. Fragments of 9–14 kb were excised from the gel, half-filled (Zabarovsky and Allikmets 1986) with dCTP, and ligated into the vector digested with *Bam*HI and partially filled with dATP, dGTP and dTTP. The second two libraries were similarly constructed except that they used fragments from a complete *SalI* digestion, half-filled and ligated into *Bam*HI-digested and half-filled pBeloBAC11. The inserts in the *SalI* libraries were in the size ranges of 10–15 kb and 15–25 kb. The half-fill ligation approach and the choice of fragment size ranges that differed by less than two-fold were both designed to minimize the possibility of chimeric clones.

Both ends of 192 BAC clones from the *HpaII* library and 96 clones for each of the two *SalI* libraries were subjected to DNA sequence analysis. The length distribution of the obtained sequences was 100–759 bp, generating 410 kb of total sequence, with a mean read of 545 bp and a median read of 589 bp. Overall, just over 77% of the sequencing reactions yielded 100 bp or more of high quality (PHRED 20) (Ewing and Green 1998) sequence. These sequences were scored for



**Figure 1** Structure of the maize *adh1-F* region, an example of a genic segment of the maize genome. Narrow shaded boxes indicate blocks of methylated DNA consisting of nested retrotransposons (SanMiguel et al. 1996). The open boxes indicate gene islands, with genes depicted as arrows or arrowheads oriented in the predicted direction of their transcription and with a length indicative of their predicted initial transcript size. The lines indicate the predicted *HpaII* (top) or *SalI* (bottom) fragments that would be generated from digestion of maize genomic DNA if all of the *HpaII* and *SalI* sites in the retrotransposon blocks are 100% methylated at all 5'-CG-3' and 5'-CNG-3' sites, while these sites are 100% unmethylated in the gene islands. Vertical arrows indicate MSLL ends that would be sequenced. Data taken from Tikhonov (Tikhonov et al. 1999).

the presence of genes, LTR-retrotransposons, other repeats, and organellar DNA. The same analysis was performed on 167 sequences that we generated for the ends of *EcoRI* BACs from maize inbred B73 (<http://www.chori.org/bacpac>). Table 1 shows the summarized results for these four sets of data. The *HpaII* library yielded end sequences that exhibited homology to genes, retrotransposons, and chloroplast DNA for a respective 5%, 25% and 4% of the clones. The end sequences of the smaller *Sall* library yielded these same classes in a respective 14%, 23%, and 17% of clones, while the larger *Sall* library yielded a respective 18%, 18%, and 5% for these homologies. In sharp contrast, the *EcoRI* library yielded ends that were homologous to genes, retrotransposons, and chloroplast DNA for a respective 1%, 52%, and 1% of the time. The results for the *EcoRI* BAC ends are very similar to those seen in random sequencing of sheared fragments of the maize genome, which gave these same homologies in about 1%, 48%, and 1% of the sequenced clones (Rabinowicz et al. 1999; Meyers et al. 2001).

These results indicate that the MSLL clones are enriched for genes and deficient for LTR-retrotransposons at their ends. Even though the LTR-retrotransposons are underrepresented in the MSLL libraries, we were surprised that so many LTR-retrotransposon homologies were still detected. However, closer inspection of the sequence data indicated that just over 10% (14/136) of the sequence homologies to LTR-retrotransposons began within the first 10 bp of sequence for the MSLL clones. For the *EcoRI* library, about 55% (42/77) of the homologies to LTR-retrotransposons were found to begin in the first 10 bp of sequence. Hence many of the MSLL clones have ends that are outside, but very near, LTR-retrotransposon blocks. Because the average *HpaII* site will be less than 300 bp from the first methylated region, we expect that this close juxtaposition of the cleaved site and an LTR-retrotransposon block should be particularly frequent in *HpaII*-based MSLL clones.

Of the 292 *HpaII* sequences that gave 100 bp or more of PHRED 20 sequence, 106 sequences were found to have one or more additional *HpaII* sites within the sequences generated. Many of these *HpaII* sites (71%) are within annotated LTR-retrotransposons. Despite their extensive methylation at 5'-CG-3' and 5'-CNG-3' sites (Bennetzen et al. 1994), our sequence inspections have shown that the most abundant LTR-retrotransposons that make up over half of the maize genome (SanMiguel and Bennetzen 1998; Meyers et al. 2001) are actually enriched in *HpaII* sites relative to genes (one per 239 bp versus one per 492 bp for genes), largely because of the higher average GC content of the LTR-retrotransposons.

Various crude predictions for the maize genome suggest that around 5%–15% of the total nuclear DNA is composed of genes, while about 50%–80% is composed of LTR-

retrotransposons (SanMiguel and Bennetzen 1998; Meyers et al. 2001). Part of the rationale for using an internal control with *EcoRI* clone ends was to balance our criteria for gene and LTR-retrotransposon identification. We expect that the true frequencies of both genic and LTR-retrotransposon sequences in our data sets are higher than the conservative numbers that we apply, but we used the same criteria for both MSLL and *EcoRI* clones. Because the predicted distance between the cleaved *HpaII* site and the LTR-retrotransposon block should average less than 300 bp, we expect that *HpaII* BAC end sites will rarely be within the peptide-encoding portion of a gene, and thus rarely identified as genic.

The frequency of chloroplast DNA homologies in our MSLL clones was about as expected, given that these libraries contained relatively small inserts. In the *Sall* libraries, for instance, the same chloroplast fragments were seen over and over again, and these were the rare chloroplast *Sall* fragments that were of the appropriate size to be found in these libraries. Libraries made with bigger *Sall* fragments, larger than 50 kb, should not have any organellar DNA fragments. These chloroplast DNA fragments were useful, however, in that they exhibited the expected chloroplast sequence homology at both ends, suggesting that the libraries did not have many chimeric clones.

The results indicate that MSLL technology can be used to link adjacent genic regions, while providing the intervening repetitive/methylated DNA block on a clone that is available for any subsequent analysis. In order to be comprehensive, several complete digestion libraries would need to be made across a large range of DNA sizes and with a variety of restriction enzymes. *HpaII* would be especially efficient for spanning small repetitive DNA blocks (those less than 15 kb or so), whereas *Sall*, *PstI*, *SmaI*, *SstII* or other methylation-sensitive enzymes with a 6 bp specificity could best characterize large methylated blocks. End sequences of these BACs would link and order all unmethylated regions. Combined with the sequences of these unmethylated regions by methyl filtration shotgun sequencing (Rabinowicz et al. 1999), the MSLL data would permit the assembly of full chromosome contigs.

The sequences of the MSLL BAC ends identify the boundaries between unmethylated DNA (e.g., the cleaved *HpaII* site) and methylated DNA (e.g., the first *HpaII* site in the BAC end sequence, hence the first methylated *HpaII* site). It is not known how these “epigenetic boundaries” are composed in plants, how they are established, or what effects they may have on adjacent genes. The MSLL technology provides comprehensive access to these regions, making them available for more detailed study.

Although our experiments were focused on characterization of the maize genome, MSLL technology should be

**Table 1. BAC End Sequence Compositions**

Library name	Total # of sequences <sup>a</sup>	Known retrotransposon (%) <sup>b</sup>	Other repetitive DNA (%) <sup>c</sup>	Known gene (%) <sup>d</sup>	Chloroplast DNA (%) <sup>e</sup>	Ribosomal DNA (%) <sup>f</sup>
<i>HpaII</i> BACs	292	72 (25%)	30 (10%)	16 (5.5%)	12 (4%)	2 (0.7%)
<i>Sall</i> BACs (10–15 kb)	147	34 (23%)	13 (8.8%)	20 (14%)	25 (17%)	2 (1.4%)
<i>Sall</i> BACs (15–25 kb)	163	30 (18%)	16 (10%)	29 (18%)	8 (5%)	0 (0%)
<i>EcoRI</i> BACs	149	77 (52%)	21 (14%)	2 (1.3%)	2 (1.3%)	2 (1.3%)

<sup>a</sup>Sequences with high quality bases  $\geq 100$

<sup>b</sup>TBLASTX, <sup>c</sup>e<sup>1</sup>BLASTN and <sup>d</sup>BLASTX cut off  $E < 10^{-5}$ .

equally useful for application to any genome with a structure similar to that of maize. These similarly accessible genomes would certainly include barley, wheat, and numerous other vascular plants, but could also include many animals, protists, or fungi with complex genomes and DNA methylation that is enriched in repetitive DNAs.

## METHODS

### Preparation of High-Molecular-Weight (HMW) DNA From Maize

Maize inbred B73 seeds were kindly provided by Dr. Chris Staiger (Purdue University). HMW DNA was extracted from the leaves of 10-day-old seedlings, as previously described (Liu and Whittier 1994). The final nuclear pellet was embedded in an equal volume of 1.5% low-melting-point agarose. Plugs containing 4–5 µg of DNA were treated with lysis buffer (1% sodium lauryl sarcosine, 0.1 mg/ml proteinase K, 0.1% ascorbic acid, 0.5M EDTA pH 9.1) in 50 ml volume for 48 h at 50°C, with one change of lysis buffer after 24 h.

### Digestion, Size Selection of HMW DNA, and BAC Library Construction

Before digestion, the agarose plugs were washed at 50°C in several volumes of washing buffer containing 1mM phenylmethylsulfonyl fluoride (PMSF). 6–8 plugs were equilibrated for 30 min to 1 h in 400 µl *Sall* or *HpaII* buffer. Digestion was performed in 200 µl volumes with 60 units of restriction enzymes at 37°C for 12–14 h to achieve complete digestion. Digested DNA plugs were size fractionated on the CHEF-DR1 system (Bio-Rad) and visualized under DARK READER (Clare Chemical Research). Different size fragments were cut from the gel and recovered by GElase (Epicentre Technologies). *Sall* fragments were half filled with dTTP and dCTP, and ligated into the pBeloBAC11 vector digested with *BamHI* and partially filled with dATP and dGTP. *HpaII* fragments were partially filled with dCTP and cloned into the same vector digested with *BamHI* and partially filled with dATP, dGTP, and dTTP. Ligations were transformed into ElectroMAX DH10B-competent cells (Life Technologies).

### BAC End Sequencing and Analysis

The BAC DNA templates were prepared following a modification of the standard procedure (Kelley et al. 1999). In brief, a 96-well block containing 1.3 ml per well of LB medium with chloramphenicol was incubated at 37°C for 16 h. 100 µl of the overnight culture was transferred to four 96-well blocks containing 1.3 ml per well of LB medium with chloramphenicol, and grown at 37°C for 14 hr. The BAC DNA was isolated from each block using the Qiagen R.E.A.L. prep 96 system following the manufacturer's instructions. The final DNA pellet containing the pooled DNA from each of the four identical cultures was dissolved in 40 µl of water. Sequencing reactions were set up with 10 µl template, 6 µl 5 X ABI (Perkin Elmer), 4 µl Big Dye (Perkin Elmer), 1 µl DMSO and 0.1 µl forward or reverse universal primer in a final reaction volume of 21.1 µl. The sequence traces were transferred to a Sun E450 server and bases were called using phred (Ewing and Green 1998). Vector sequences were masked by CROSS\_MATCH. BLAST (Altschul et al. 1997) was employed to compare all the trimmed sequences with the public sequence database as both nucleotides and predicted amino acid translations. Retrotransposons

were detected additionally by CROSS\_MATCH and TBLASTX against a set of 103 known retroelements.

## ACKNOWLEDGMENTS

We thank the US National Science Foundation for support of this project (grants 9975618 and 9975793).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Altschul, S.F., Madden, T.L., Schäffer, 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. *Nucl. Acids Res.* **25**: 3389–3402.
- Antequera, F. and Bird, A.P. 1988. Unmethylated CpG islands associated with genes in higher plant DNA. *EMBO J.* **7**: 2295–2299.
- Bennetzen, J.L., Schrick, K., Springer, P.S., and SanMiguel, P. 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* **37**: 565–576.
- Bennetzen, J.L., Chandler, V.L., and Schnable, P. 2001. National Science Foundation-Sponsored Workshop Report. Maize Genome Sequencing Project. *Plant Physiol.* **127**: 1572–1578.
- Bird, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209–213.
- Dubcovsky, J., Ramakrishna, W., SanMiguel, P.J., Busso, C.S., Yan, L., Shiloff, B.A., and Bennetzen J.L. 2001. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol.* **125**: 1342–1353.
- Edwards, K.J., Thompson, H., Edwards, D., de Saizieu, A., Carline, S., Thompson, J.A., Greenland, A.J., Evers, M., and Schuch, W. 1992. Construction and characterization of a yeast artificial chromosome library containing three haploid maize genome equivalents. *Plant Molecular Biology.* **19**: 299–308.
- Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**: 186–194.
- Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A., and Cedar, H. 1991. Demethylation of CpG islands in embryonic cells. *Nature* **351**: 239–241.
- Fu, H., Park, W., Yan, X., Zheng, Z., Shen, B., and Dooner, H.K. 2001. The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. *Proc. Natl. Acad. Sci.* **98**: 8903–8908.
- Gruenbaum, Y., Naveh-Many, T., Cedar, H., and Razin, A. 1981a. Sequence specificity of methylation in higher plant DNA. *Nature* **292**: 860–862.
- Gruenbaum, Y., Stein, R., Cedar, H., and Razin, A. 1981b. Methylation of CpG sequences in eukaryotic DNA. *FEBS Lett.* **124**: 67–71.
- Hake, S. and Walbot, V. 1980. The genome of *Zea mays*, its organization and homology to related grasses. *Chromosoma* **79**: 251–270.
- Jacobsen, S.E. and Meyerowitz, E.M. 1997. Hypermethylated SUPERMAN epigenetic alleles in Arabidopsis. *Science* **277**: 1100–1103.
- Kelley, J.M., Field, C.E., Craven, M.B., Bocskai D., Kim U.-J., Rounsley, S.D., and Adams, M.D. 1999. High throughput direct sequencing of BAC clones. *Nucleic Acids Res.* **27**: 1539–1546.
- Kim, U.-J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.-L., Simon, M.I., and Shizuya, H. 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**: 213–218.
- Kumar, A. and Bennetzen, J.L. 1999. Plant retrotransposons. *Annu. Rev. Genet.* **34**: 479–532.
- Liu, Y.-G. and Whittier, R.F. 1994. Rapid preparation of megabase plant DNA from nuclei in agarose plugs and microbeads. *Nucleic Acids Res.* **22**: 2168–2169.
- Llaca, V. and Messing, J. 1998. Amplicons of maize zein genes are conserved within genetic but expanded and constricted in intergenic regions. *Plant J.* **15**: 211–220.
- Meyers, B.C., Tingey, S.V., and Morgante, M. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* **11**: 1660–1676.

- Panstruga, R., Buschges, R., Piffanelli, P., and Schulze-Lefert, P. 1998. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Res.* **26**: 1056–1062.
- Rabinowicz, P.D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L.D., Stein, L., McCombie, W.R., and Martienssen, R.A. 1999. Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genet.* **23**: 305–308.
- SanMiguel, P. and Bennetzen, J.L. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* **82**: 37–44.
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- Springer, P.S. 1992. Genomic organization of *Zea mays* and its close relatives, Ph.D. Thesis, Purdue University.
- 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.* **96**: 7409–7414.
- Vicient, C.M., Suoniemi, A., Anamthawat-Jónsson, K., Tanskanen, J., Beharav, A., Nevo, E., and Schulman, A.H. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**: 1769–1784.
- Wicker, T.N., Albar, L., Feuillet, C., Schlagenhauf, E., and Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J.* **26**: 307–316.
- Zabarovsky, E.R. and Allikmets, R.L. 1986. An improved technique for the efficient construction of gene libraries by partial filling-in of cohesive ends. *Gene* **42**: 11.

## WEB SITE REFERENCES

<http://www.chori.org/bacpac>; BACPAC Resources home page for Pieter de Jong's lab at the Children's Hospital Oakland Research Institute.

Received February 14, 2002; accepted in revised form July 17, 2002.