the plant journal

SEB Society for Experimental Biology

The Plant Journal (2011) doi: 10.1111/j.1365-313X.2011.04867.x

TECHNICAL ADVANCE

Stable integration of an engineered megabase repeat array into the maize genome

Han Zhang[1], Bao H. Phan[2], Kai Wang[3], Barbara J. Artelt[2], Jiming Jiang[3], Wayne A. Parrott[2] and R. Kelly Dawe[1,4,*]

[1]*Department of Genetics, University of Georgia, Athens, GA 30602, USA,*

[2]*Department of Crop and Soil Sciences, University of Georgia, Athens, GA 30602, USA,*

[3]*Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA, and*

[4]*Department of Plant Biology, University of Georgia, Athens, GA 30602, USA*

Received 10 October 2011; revised 21 November 2011; accepted 22 November 2011.

[*]For correspondence (fax +1 706 542 1805; e-mail kelly@plantbio.uga.edu).

SUMMARY

Plant genome engineering as a practical matter will require stable introduction of long and complex segments of DNA sequence into plant genomes. Here we show that it is possible to synthetically engineer and introduce centromere-sized satellite repeat arrays into maize. We designed a synthetic repeat monomer of 156 bp that contains five DNA-binding motifs (LacO, TetO, Gal4, LexA, and CENPB), and extended it into tandem arrays using an overlapping PCR method similar to that commonly used in gene synthesis. The PCR products were then directly transformed into maize using biolistic transformation. We identified three resulting insertion sites (arrayed binding sites), the longest of which is at least 1100 kb. The LacI DNA-binding module is sufficient to efficiently tether YFP to the arrayed binding sites. We conclude that synthetic repeats can be delivered into plant cells by omitting passage through *Escherichia coli*, that they generally insert into one locus, and that great lengths may be achieved. It is anticipated that these experimental approaches will be useful for future applications in artificial chromosome design.

Keywords: tandem repeats, centromeres, artificial chromosomes, molecular tethering, biolistic transformation, maize.

INTRODUCTION

In recent years, the field of plant improvement has shifted from widespread use of single transgenes into discussions of multi-gene stacking (Halpin, 2005) and plant chromosome engineering (Purnick and Weiss, 2009). This interest has been paralleled by exciting new methods for gene synthesis, precise targeting of transgenes to specific loci (Urnov *et al.*, 2010) and manipulation of plant ploidy (Ravi and Chan, 2010). In *Escherichia coli* and other smaller-genome species, the horizon has further expanded as multiple authors consider the design or substantial re-design of entire genomes (Purnick and Weiss, 2009). Similar ideas may one day be feasible in plants, but at present we are limited by severe technical bottlenecks. Two of these are the inability to clone and transfer large DNA molecules into plants, particularly those with repetitive sequences, and the lack of an empirical framework to design artificial centromeres.

One of the first examples of genome engineering was the creation of artificial chromosomes in *Saccharomyces cerevisiae*. Soon after the small (125 bp) yeast centromeres were identified, they were combined with origins of replication and telomeric repeats and shown to be fully transmissible with large foreign DNA cargoes (Murray and Szostak, 1983). Similar approaches (referred to as 'bottom-up' strategies) have been used to create synthetic human artificial chromosomes (Harrington *et al.*, 1997; Ikeno *et al.*, 1998; Henning *et al.*, 1999). These small functional chromosomes may be linear or circular, and include centromeric alpha-satellite arrays of natural or synthetic origin (Ebersole *et al.*, 2000; Basu *et al.*, 2005; Nakano *et al.*, 2008). In some cases, the resulting mini-chromosomes were found to be substantially larger than the original input DNA, suggesting that the introduced sequences had undergone duplication and expansion as part of chromosome formation (Ikeno *et al.*, 1998).

Plant and animal centromeres are similar in size and content, being approximately 300–2000 kb in size and

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd

composed primarily of tandem repeat arrays and interspersed transposons (Birchler and Han, 2009). Given these similarities, several laboratories have attempted to create plant artificial chromosomes. The first published experiments showed that the simple strategy of transforming rice (*Oryza sativa*) centromeric bacterial artificial chromosomes (BACs) did not result in artificial chromosomes (Phan *et al.*, 2007). In a similar but more informative study, a second group identified maize (*Zea mays*) centromeric BACs, retrofitted them with origins of replication and telomeric repeats, and then transformed them into maize (Ananiev *et al.*, 2009). The results suggested that introduced centromeric BACs can initiate centromere assembly, although most of the minichromosomes appeared to have resulted from genome rearrangements. In a third study, centromeric BACs as small as 19 kb were assembled into a circular vector and shown to be both mitotically and meiotically heritable, with evidence that the chromosomes were transmitted as independent ring chromosomes (Carlson *et al.*, 2007). The latter study appeared to be the most efficient, but doubts remain regarding the size of the constructs and the mechanism of inheritance (Houben *et al.*, 2008).

The factors limiting the success of *de novo* centromere assembly in plants are both technical and biological. Major technical limitations lie in the methods of plant transformation and the cloning vehicles used. Previous studies have relied on existing centromere sequences that were cloned into BACs and introduced into plant genomes using biolistic transformation (Carlson *et al.*, 2007; Phan *et al.*, 2007; Ananiev *et al.*, 2009). It should also be possible to introduce large (up to 160 kb) centromeric BACs into plant genomes using *Agrobacterium* and specialized T-DNA vectors (Hamilton *et al.*, 1996; Liu *et al.*, 1999). However, given the instability of repeat arrays in *E. coli* and *Agrobacterium*, there are inherent limitations to what can be accomplished using BAC clones (Song *et al.*, 2001, 2003; Nakano *et al.*, 2005; Chang *et al.*, 2011). In addition, a major unaddressed biological limitation is the fact that centromeres are specified in a sequence-independent (epigenetic) fashion (Birchler and Han, 2009). Even if we could easily clone and manipulate centromere-sized sequences, we do not yet understand which sequences may be most useful, or whether any existing sequences can be used reliably.

Recently, it has been reported in budding yeast and animal cell cultures that *de novo* centromere activity can be achieved by tethering kinetochore proteins to synthetic sequences (Kiermaier *et al.*, 2009; Lacefield *et al.*, 2009; Barnhart *et al.*, 2011; Gascoigne *et al.*, 2011; Mendiburo *et al.*, 2011). This approach takes advantage of the high-affinity interaction between DNA-binding proteins and their target sequences. A DNA-binding protein such as the lactose repressor protein (LacI) will recognize and bind to *lac* operator sequences (LacO) integrated into the genome. Kinetochore proteins fused to LacI can be tethered to LacO arrays to initiate the assembly of additional kinetochore proteins (Kiermaier *et al.*, 2009; Lacefield *et al.*, 2009; Barnhart *et al.*, 2011; Gascoigne *et al.*, 2011; Mendiburo *et al.*, 2011). These studies have shown that it is possible to artificially assemble nearly complete kinetochores that display most if not all of the segregation properties of natural kinetochores. In principle, the tethering approach combines the predictability of a sequence-based (genetic) specification with the ability to carefully control the process of kinetochore assembly.

This study addresses one of the key technical barriers to creating engineered centromeres in plants. Very long synthetic tandem arrays containing DNA-binding modules were generated by overlap extension PCR, and directly transformed into maize, where they were incorporated as single intact loci. The utility of the arrays was demonstrated by showing that the arrays efficiently recruit LacI–YFP fusion proteins to form large fluorescent spots that are easily visible by standard microscopy.

RESULTS

Long tandem repeats can be transformed into maize and remain intact

We designed a repeat monomer that contains the DNA recognition motifs for five DNA-binding proteins (Figure 1): LacO, TetO, LexA, Gal4 and the CENPB box, which binds to the CENPB DNA-binding protein found in human centromeres (see Experimental Procedures). Each monomer of the array was designed to be 156 bp, the size of the endogenous maize centromeric satellite CentC (Ananiev *et al.*, 1998). Arrays were synthesized by PCR using two long oligonucleotides (98 and 99 bp) that overlap at both ends. In a standard PCR reaction, the complementary oligos anneal and extend the corresponding strands, and then extend the repeating monomers indefinitely (Figure 2). The engineered repeat arrays produced in this way are referred to as arrayed binding sites (ABS).

Agarose gel electrophoresis of ABS amplification reactions showed a negative relationship between primer concentration and the size of the resulting PCR products: as the primer concentration decreased, PCR product sizes

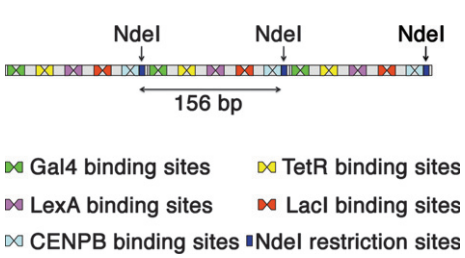

Figure 1. Design of the arrayed binding sites (ABS).
A trimer of the 156 bp monomer is shown. Binding sites for Gal4, TetR, LexA, LacI and CENPB, as well as an *Nde*I restriction site were present in each monomer. Outside the six constrained sequences, bases were chosen at random.

© 2011 The Authors

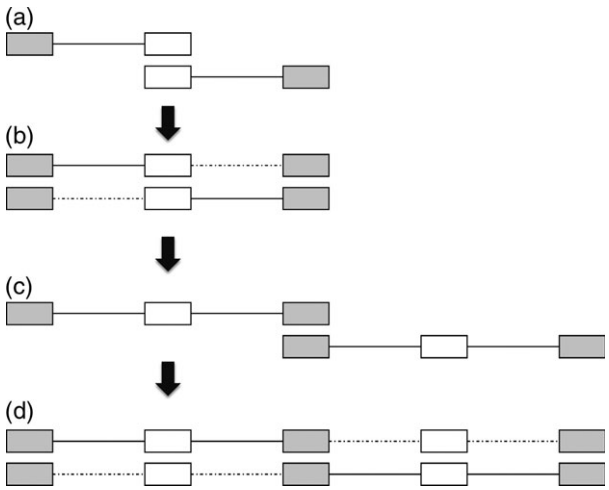

Figure 2. Scheme of overlap extension PCR.
Two long oligonucleotides (98 and 99 bp) overlapping at both the 5′ end (gray boxes) and the 3′ end (white boxes) were used in the PCR reaction to generate long repeat arrays. Upon denaturing, complementary oligos attach to each other at homologous regions (a, c) and serve as primers to extend the array (b, d). Subsequent cycles extend the array further.

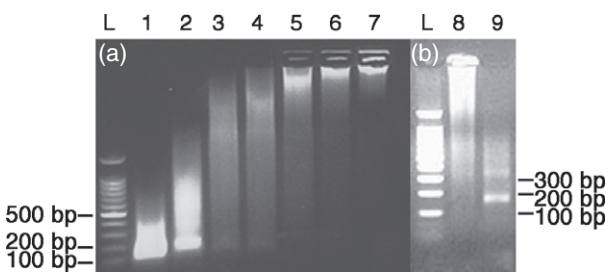

Figure 3. Synthesis of long repeat arrays with expected tandem structure.
(a) Gel electrophoresis of the overlap extension PCR products using a dilution series of primers. L, DNA ladder; lanes 1–6, PCR products using primer concentrations of 0.4, 0.2, 0.1, 0.04, 0.02 and 0.01 μM, respectively; lane 7, positive control using products of lane 6 as templates for a second round of PCR.
(b) Gel electrophoresis of the PCR products before (lane 8) and after (lane 9) restriction digest with *Nde*I. L, DNA ladder. The PCR products were cleaved into 156 bp monomers and 312 bp dimers.

increased (Figure 3a). PCR products rapidly reached sizes large enough to be excluded from the agarose matrix (Figure 3a, lane 6). We also used a contour-clamped homogeneous electric field (CHEF), which enables resolution of macromolecules as large as two megabases (Chu *et al.*, 1986), and again observed that the PCR products did not enter the gel matrix (data not shown). These PCR products are undoubtedly very large, but we do not know their actual size or structure; they may be partially circular or contain irregular multimers that prevent entry into agarose gels. Nevertheless, after digestion of the PCR products with *Nde*I, which cleaves once in each monomer, the PCR products were reduced to 156 bp monomers and 312 bp dimers, showing that the vast majority of sequences are tandemly arrayed as predicted (Figure 3b).

For simplicity, we opted to directly transform the PCR products into maize embryogenic cultures. Prior work had shown that smaller PCR products can be transformed into rice by co-bombarding with a plasmid that contains a selective marker (Kumar *et al.*, 2006). The amplified ABS products were mixed with plasmid pAHC25 (Christensen and Quail, 1996), which contains the BASTA resistance gene and the GUS reporter gene, and transformed into maize calli by microprojectile bombardment. A total of 59 individual transformants were BASTA-resistant, and 22 (37.3%) contained detectable quantities of the ABS array by PCR. From this set of 22 calli, 17 mature plants were recovered (28.8%).

We initially used Southern blots to assess the size of the integrated ABS loci. Genomic DNA was digested using multiple restriction enzymes that do not cleave in the array but do cleave within the co-bombarded plasmid (*Eco*RI, *Hin*dIII, *Bam*HI and *Kpn*I). The blots were then probed using the ABS monomer. Of the 17 recovered plants, five showed bright bands at molecular weights greater than 40 kb, the resolution of the agarose gel used (Figure 4a). One interesting exception is ABS3 (lane 8), which showed a large band of approximately 40 kb as well as a second distinct band of >20 kb. We also observed faint smaller bands in all lanes, which may represent either short independent arrays or subdomains of the larger arrays. Importantly, when these DNAs were digested with *Nde*I, all of the ABS-hybridizing sequences were cleaved into short pieces representing 156 bp monomers (Figure 4b).

We were able to identify progeny carrying the three longest arrays (ABS3, ABS4 and ABS7) using primers homologous to the pAHC25 marker plasmid, suggesting that at least one copy of the marker plasmid co-segregates with the arrays. Re-probing the Southern blot in Figure 4(a) using the GUS gene (data not shown) showed clear hybridization in the >40 kb molecular weight range. These results indicate that the pACH25 plasmid occasionally co-integrates with the PCR products; however, this is relatively rare, generally occurring at intervals >40 kb in the longest arrays.

Identification of three independent lines with ABS in different positions relative to centromeres

In order to determine the location of the ABS loci, fluorescence *in situ* hybridization (FISH) was performed on the three transgenic lines with the most abundant ABS signal. Chromosomes can be identified based on size, arm ratio and staining intensity of the centromere repeat CentC (Kato *et al.*, 2004). Examination of mitotic chromosome spreads showed that the arrays are integrated into single sites on the long arms of chromosomes 3, 4 and 7 (Figure 5), and thus are referred to as ABS3, ABS4 and ABS7. In ABS3 and ABS4, the tandem arrays are located at central positions of the long arms, while the array in ABS7 is located near the

© 2011 The Authors

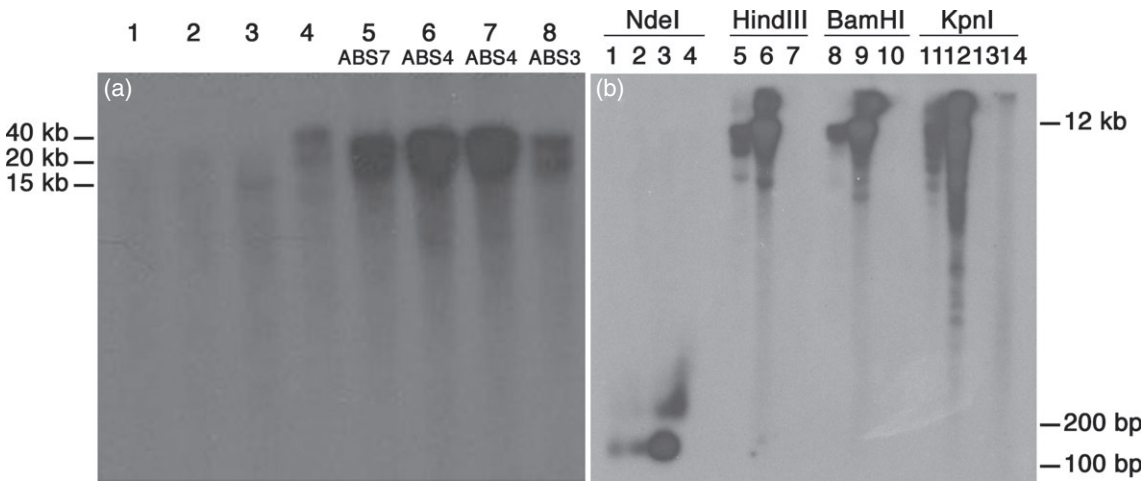

Figure 4. Southern blot analysis showing the size of the integrated ABS repeat arrays.
(a) Comparison of transformed lines. Genomic DNAs from wild-type (lane 1) and seven transgenic lines (lanes 2–8) were digested using *Eco*RI and probed with the ABS monomer. Five lines (lanes 4–8) showed bright bands at a molecular weight >40 kb. Three of these were studied further and renamed ABS3, ABS4 and ABS7 as shown (lanes 6 and 7 were later found to be clones of the same event).
(b) Demonstration that ABS arrays are composed of tandem repeats. Genomic DNAs from wild-type (lanes 4, 7, 10 and 13), ABS3 (lanes 2, 5, 8 and 11) and ABS4 (lanes 3, 6, 9 and 12) were digested using *Nde*I, *Hin*dIII, *Bam*HI and *Kpn*I and probed using ABS monomers. PCR products digested with *Nde*I (lane 1) and *Kpn*I (lane 14) were loaded as controls to indicate expected product sizes.

endogenous centromere. The cytological position of ABS7 was further refined by analyzing meiotic pachytene chromosomes, which provide higher resolution (Wang *et al.*, 2006). The results show that ABS7 is located in pericentromeric heterochromatin at position 7L.1 on the cytological map (approximately 10% of the distance to the telomere; see Figure 6).

To determine the physical lengths of the integrated ABS arrays, we used fiber-FISH technology, which can accurately estimate size from stretched DNA molecules (Figure 7 and Table 1). More than 20 fiber-FISH images were collected and measured for each ABS line. The lengths of the fiber-FISH signals (in μm) were converted to kb using the 3.21 kb μm^{-1} conversion rate derived from rice fibers (Cheng *et al.*, 2002). The data show that ABS3 comprises two satellite blocks of 28 and 53 kb that are separated by a 109 kb gap. The fact that ABS3 is divided into two separate regions, one of which is approximately 30 kb, may explain the existence of the second smaller band in the ABS3 lane of the Southern blot (Figure 4a). In contrast, the fiber-FISH signals from ABS4 and ABS7 lacked obvious intervening sequences, and measured 1146 and 315 kb, respectively. The dotted hybridization pattern is typical of fiber-FISH images and does not imply that the ABS arrays are interrupted at regular intervals. If we assume that the arrays are nearly continuous, ABS4 contains over 7000 tandemly arrayed copies of the synthetic 156 bp monomer.

All three ABS loci were reliably inherited over multiple generations. The fiber-FISH images in Figure 7 were taken from plants that were at least four generations removed from the original transformants. These data suggest that synthetic repeat arrays, like natural centromeric repeats and the longer maize knob repeat arrays (Adawy *et al.*, 2004), are stably replicated and transmitted.

Segregation and inheritance of tethered LacI–YFP in transgenic plants

To test whether the arrays function as tethering sites, we separately transformed maize with a LacI–YFP fusion construct whose expression was driven by the constitutive 35S promoter. We anticipated that if plants expressing LacI–YFP were crossed with ABS lines, the YFP signal would aggregate to produce strong and focused fluorescent signals. As expected, in plants hemizygous for LacI–YFP and ABS4, we observed either a single bright dot in the nuclei (Figure 8c, top row) or two closely paired signals, representing the G_2 phase of the cell cycle (Figure 8c, bottom row). Sibling plants that contained LacI–YFP but lacked ABS showed no fluorescent foci (Figure 8a). We also crossed hemizygous plants to produce a population segregating for LacI–YFP and zero, one or two copies of ABS4 (Figure 8b,d,f). Using FISH to determine the ABS genotype, we showed that, in a sample of 13 plants, the number of visible YFP foci corresponded perfectly to the number of ABS arrays present in the cell (Figure 8).

ABS4 is the longest of the arrays, being almost four times the size of ABS7 and more than 10 times the size of the combined ABS3 blocks. We found that fluorescent foci were always observed in LacI–YFP lines containing ABS4 (19 plants tested) and ABS7 (12 plants tested). However, visible YFP foci were observed in only one of nine plants containing the ABS3 array. We were able to directly compare ABS3 with ABS4 by crossing the two lines to create plants containing both arrays as well as LacI–YFP. In these lines, one bright dot

© 2011 The Authors

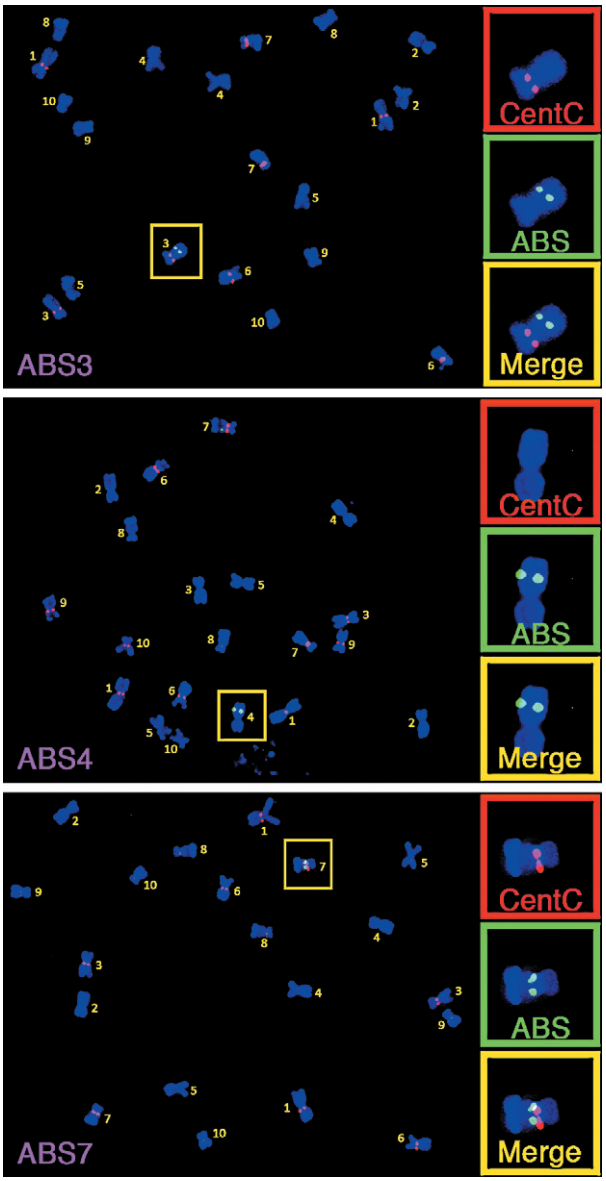

Figure 5. Localization of ABS arrays by fluorescence *in situ* hybridization.
Maize mitotic chromosome spreads were hybridized with endogenous centromeric satellite CentC (red) and synthetic ABS repeats (green). Each chromosome is identified by its number. Chromosomes containing ABS arrays are enlarged as insets to show their positions on the long arms of chromosome 3, 4 and 7. CentC quantity varies among chromosomes and lines, and the near absence of CentC signal on chromosome 4 has been reported previously (Kato *et al.*, 2004).

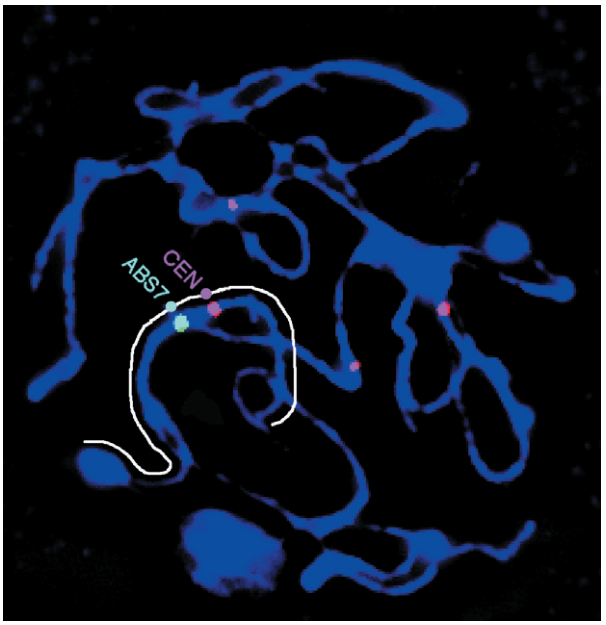

Figure 6. Localization of ABS7 in the pericentromeric heterochromatin of chromosome 7.
Maize pachytene chromosomes were hybridized with endogenous centromeric satellite CentC (red) and synthetic ABS repeats (green). Chromosome 7 was identified based on the large knob at the end of the long arm and is indicated by a white curve.

(presumably ABS4) was observed in all cells, and occasionally a second much weaker dot, barely at the limit of detection, was also visible (data not shown). It is possible that more LacO repeats are necessary to visualize YFP in maize than in Arabidopsis, where 120 copies of the *lacO* sequence were sufficient to visualize LacI fused to GFP (Kato and Lam, 2001). However, other factors are also probably involved, such as the expression level of the 35S promoter in maize, the relative brightness of the fluorescent protein used, and the transmission of fluorescence through intact maize roots.

DISCUSSION

Our goal was to produce synthetic repeat arrays resembling natural centromeres, and introduce the arrays into the maize genome. At the time the study was initiated, LacO and TetO repeat arrays were available as plasmid clones (Robinett *et al.*, 1996; Michaelis *et al.*, 1997); however, these arrays (<6 kb) are far shorter than natural centromeres, and are known to be unstable in *E. coli*, and lack the nucleosome-sized repeat structure typical of satellite repeats. We used a simple PCR method involving double-overlapping primers to extend the tandem repeat structure indefinitely within the confines of a PCR reaction. The large PCR products were then coated onto metal spheres together with a marker plasmid and delivered into maize embryogenic cultures. The ABS4 array described here (Figure 7) is the longest segment of foreign DNA yet incorporated into a higher plant, and, at 1.1 Mb, is the size of many functional centromere regions (Murata *et al.*, 2008; Yan *et al.*, 2008; Wolfgruber *et al.*, 2009).

Previous studies have shown that biolistic transformation can be used to introduce nearly any type of DNA into plant

© 2011 The Authors

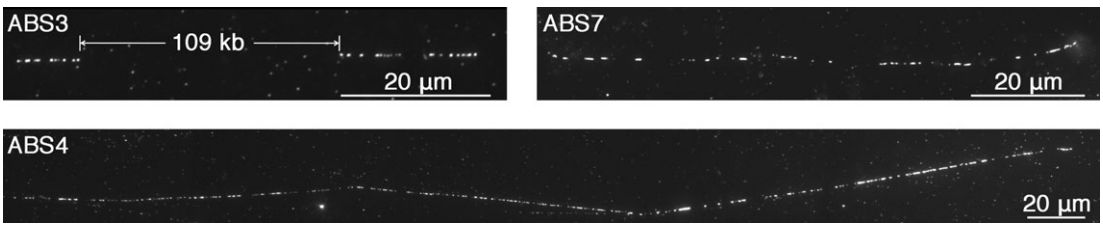

Figure 7. Size estimation of ABS arrays by fiber-FISH.
Stretched DNA molecules from the ABS lines were hybridized with synthetic ABS repeat probes (white). ABS4 and ABS7 are continuous or nearly continuous, and ABS3 comprises two arrays separated by a 109 kb gap.

Table 1 Size estimates of ABS arrays by fiber-FISH

	Mean length (µm)	SD	*n*	Estimated length (kb)
ABS4	357.04	55.47	22	1146.1
ABS7	98.11	9.17	24	314.93
ABS3				
Insert 1	8.76	1.01	31	28.12
Gap	34.1	2.31	31	109.46
Insert 2	16.67	2.09	31	53.51
Total for ABS3	59.53	3.66	31	191.09

Fiber-FISH images were measured in µm and converted to kb using a 3.21 kb μm^{-1} conversion rate.

genomes, but it often results in rearranged or duplicated insertions that sometimes include interspersed genomic sequences (Sautter *et al.*, 1999; Friebe *et al.*, 2001). In our case, it is difficult to interpret the events that occurred upon transformation; however, consistent with previous studies, we found that the PCR products co-integrated with marker plasmids, and that there is a low level of host genome rearrangement upon insertion (see ABS3, Figure 7). More importantly, our data show that the great majority of integrated sequences are derived from ABS arrays, and that these long synthetic molecules survived biolistic transformation to become heritable features of the genome. While basic considerations of size and repeat character would have made this study impossible using *Agrobacterium*, at present there appears to be no inherent limit to the size or type of DNA molecule that can be transformed into plants using biolistic methods.

The value of long arrays of centromere-like repeats lies in their use in artificial chromosome technology. One may question the need for synthetic repeats when the natural centromere repeats (alpha-satellites) in human cells are sufficient to form functional artificial chromosomes (Harrington *et al.*, 1997; Ebersole *et al.*, 2000), and similar strategies in plants have met with at least partial success (Carlson *et al.*, 2007; Ananiev *et al.*, 2009). However, an artificial chromosome strategy based on its success in cultured human cells has inherent weaknesses for higher plants. Unlike plants, human alpha-satellites have evolved in concert with a specialized binding protein called CENPB. Human CENPB appears to function primarily to establish or

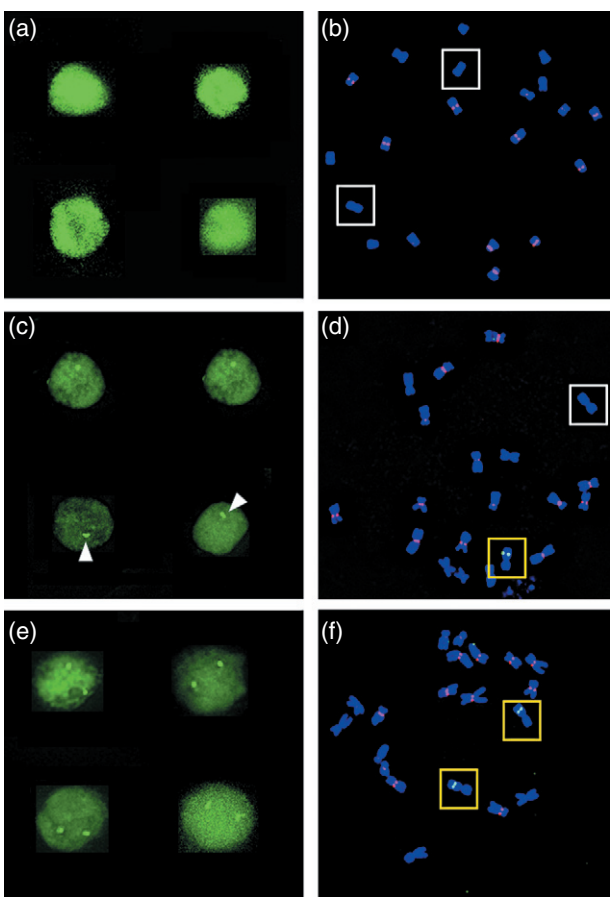

Figure 8. Tethered LacI–YFP in ABS4 segregating lines.
(a,c,e) YFP fluorescence in ABS4 segregating lines. The green spots represent localization of LacI–YFP proteins in ABS4 arrays, as confirmed by genotyping and FISH. The two closely paired dots in the bottom row of (c) (arrowheads) represent sister chromatids at the G_2 phase of the cell cycle.
(b, d, f) FISH on ABS4 segregating lines. Maize mitotic spreads were hybridized with endogenous centromeric satellite CentC (red) and synthetic ABS repeats (green). Each chromosome 4 is boxed in white (without ABS) or yellow (with ABS). The number of YFP spots matches the number of ABS loci in the line: no ABS (a, b), ABS hemizygous (c, d, one spot) and ABS homozygous (e, f, two spots).

maintain centromere identity by binding to a CENPB-binding motif in functional centromeric domains (Masumoto *et al.*, 1989; Muro *et al.*, 1992). If the motif is removed, artificial

© 2011 The Authors

centromeres do not form (Ohzeki *et al.*, 2002). Plants, which lack any evidence of a similar DNA-to-protein genetic specification mechanism, appear to rely heavily on epigenetic processes to ensure centromere assembly (Birchler and Han, 2009).

Recently, a series of papers have illustrated the utility of using protein tethering to confer a genetic specification mechanism to centromeric regions. In yeast, a microtubule binding protein (Dam1) was tethered to a DNA-binding array and shown to promote segregation of an otherwise acentric plasmid (Kiermaier *et al.*, 2009; Lacefield *et al.*, 2009). Similarly, tethering of the human kinetochore proteins CENPC or CENPN was sufficient to assemble outer kinetochore proteins over an integrated repeat array (Gascoigne *et al.*, 2011). Other studies went further by directly tethering Centromeric Histone H3 (CENH3) or its deposition factor Holliday Junction Recognition Protein (HJURP), and demonstrating the ectopic assembly of centromeric nucleosomes and fully functional kinetochores over the introduced repeat arrays (Barnhart *et al.*, 2011; Mendiburo *et al.*, 2011).

Using the technical advances described here, it is now possible to perform similar tests in plants. The strategy we describe for producing repeat arrays is simple and should make it possible to rapidly re-design the monomer, for instance to better match known nucleosome-positioning motifs in plant centromeres (Gent *et al.*, 2011). In the short-term, such tethering arrays will help us better understand the process of plant kinetochore assembly, which is not well understood at present. In the long-term, our goal is to create self-sustaining synthetic centromeres that can be used in constructing plant artificial chromosomes, an area of active research with great promise (Purnick and Weiss, 2009).

EXPERIMENTAL PROCEDURES

ABS preparation, transformation and detection

The sequence of the ABS array was designed to match the GC content of the maize CentC array. The sequence contains an *Nde*I site, the LacO operator 5′-AATTGTGAGCGGCTCACAATT-3′ (Simons *et al.*, 1984; Lanzer and Bujard, 1988), the TetO operator 5′-TCCCTATCAGTGATAGAGA-3′ (Lutz and Bujard, 1997), the LexA upstream activation sequence (UAS) 5′-TACTGTATATATATACA GTA-3′ (Berg, 1988), the Gal4 UAS 5′-CGGAGGACTGTCCTCCG-3′ (Giniger *et al.*, 1985; Reece and Ptashne, 1993) and the CENPB box 5′-TTTCGTTGGAAACGGGA-3′ (Masumoto *et al.*, 1989), which binds to a known DNA-binding domain (Tanaka *et al.*, 2001). Outside the six constrained sequences, bases were chosen at random. Overlap extension PCR was performed using primers syn3 for (5′-GTACCGCTAGTCCCTATCAGTGATAGAGATGCAGTCGCCGAAT ACTGTATATATATACAGTAAGCCTTCTACTGGAATTGTGAGCGGCT CACAATTAGT-3′) and syn3 rev (5′- CTGATAGGGACTAGCG GTACT CCGGAGGACAGTCCTCCGAGAAGGCATATGTCCCGTTTCCAACGA AAATTTCATCCTACTAATTGTGAGCCGCTCAC-3′). PCR was performed using the BD Advantage™ 2 PCR kit (Clontech, http://www.clontech.com/) in a reaction containing 0.01 μM primers, 100 μM dNTPs, 1 μl BD Advantage™ 2 PCR buffer and 0.1 μl BD Advantage™ 2 PCR polymerase mix in a 10 μl reaction. PCR reactions performed using other kits and Taq polymerases were not as effective. PCR was performed using the following parameters: initial denaturing at 95°C for 2 min, 35 amplification cycles of 20 sec at 94°C, 1 min at 52°C and 1 min at 72°C, followed by a final incubation at 72°C for 8 min. The resulting PCR products are expected to have gaps, nicks and A overhangs. Most of the products are probably hybrid molecules with parts of multiple PCR products entwined. To correct these presumed errors, PCR products were treated with *Pfu*I polymerase, which should fill gaps, correct errors (by its 3′ → 5′ exonuclease activity) and remove the A overhangs. Finally, we treated the sample with DNA ligase to seal the nicks and join the molecules together.

Plasmid pAHC25 (Christensen and Quail, 1996), which contains the β-glucuronidase (GUS) reporter gene and the selectable BAR gene, which confers resistance to the herbicide BASTA (Bayer CropScience, http://www.bayercropscience.com/), was mixed with the PCR products at a molar ratio of 1:6, and directly shot into maize by microprojectile bombardment as described previously (Phan *et al.*, 2007). Target maize calli were subsequently selected for BASTA resistance, and then screened for ABS by PCR using primers syn3for_short (5′-GTACCGCTAGTCCCTATCAGTGAT-3′) and syn3-rev_short (5′-TCCGGAGGACAGTCCTCC-3′). Southern blot was performed as described by Phan *et al.* (2007) using probes specific for a single monomer of ABS (156 bp).

FISH and fiber-FISH

FISH was performed on root tips as described previously (Kato *et al.*, 2004), except that fluorescently labeled oligonucleotides were used to detect the ABS arrays. Immature tassels were harvested and fixed in 1× buffer A (80 mM KCl, 20 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 15 mM PIPES buffer, pH 7.0), 4% paraformaldehyde and 0.1% Triton X-100 for 2–3 h before dissecting. Pachytene chromosomes were prepared as described previously (Shi and Dawe, 2006), and FISH was performed as for root tips. Fiber-FISH was performed as described by Jackson *et al.* (1998).

LacI–YFP construct preparation, transformation and detection

The LacI gene without a stop codon was cloned into the pENTR/D vector (Invitrogen, http://www.invitrogen.com/) from *E. coli* genomic DNA (kindly provided by Sidney Kushner) using primers LacI-NLS-F2 (5′-CACCATGAAGAAGAAGAGAAAGGTGGTGAAACCAG TAACGTTATACG-3′) and LacI-R2 (5′-CTGCCCGCTTTCCAGTCGG-GAA-3′). The forward primer introduces the nuclear localization sequence KKKRKV (from the SV40 large T-antigen) as used previously (Raikhel, 1992). The sequence-confirmed LacI sequence was then recombined into pEarleygate vector pEG101 (an Invitrogen Gateway vector derivative; Earley *et al.*, 2006) to fuse the LacI protein to YFP. The final 35S promoter–LacI–YFP fusion construct was introduced into *Agrobacterium tumefaciens* strain C58C1 (a gift from Richard Meagher), and then transformed into maize Hill lines by the Iowa State University Plant Transformation Facility (http://www.agron.iastate.edu/ptf/). The transgenes were screened in the progeny using primers Gus2F (5′-AGTGAAGGGCCAACAGTTCCTG AT-3′) and Gus2R (5′-AAATATTCCCGTGCACTTGCGGAC-3′) for ABS detection and 35S1F (5′-GGCCCTAACAAGCCCACCAA-3′) and 35S1R (5′-GGGCAATGGAATCCGAGGAG-3′) for LacI–YFP detection.

Root tips approximately 1 cm long were cut from 3-day-old seedlings and fixed on the slides by taping the edges of the cover slips onto the slides. PBS was added into the space between the cover slips and slides. The slides were then observed using a Zeiss Axio Imager microscope (http://www.zeiss.com/) and a 63× objective oil lens. Images were taken using Slidebook software (Intelligent Imaging Innovations, https://www.intelligent-imaging.com/).

© 2011 The Authors

ACKNOWLEDGMENTS

We thank Amy Luce for helping to prepare the LacI–YFP fusion constructs, Sarah Covert (Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA, USA) for providing access to the CHEF gel apparatus, Sidney Kushner for supplying *E. coli* genomic DNA, and Richard Meagher for providing the *Agrobacterium tumefaciens* strain. This work was supported by grants 0421671 and 092270 from the US National Science Foundation.

REFERENCES

Adawy, S.S., Stupar, R.M. and Jiang, J. (2004) Fluorescence *in situ* hybridization analysis reveals multiple loci of knob-associated DNA elements in one-knob and knobless maize lines. *J. Histochem. Cytochem.* **52**, 1113–1116.

Ananiev, E.V., Phillips, R.L. and Rines, H.W. (1998) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc. Natl Acad. Sci. USA*, **95**, 13073–13078.

Ananiev, E.V., Wu, C., Chamberlin, M.A., Svitashev, S., Schwartz, C., Gordon-Kamm, W. and Tingey, S. (2009) Artificial chromosome formation in maize (*Zea mays* L.). *Chromosoma*, **118**, 157–177.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E. and Foltz, D.R. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional *de novo* kinetochore. *J. Cell Biol.* **194**, 229–243.

Basu, J., Stromberg, G., Compitello, G., Willard, H.F. and Van Bokkelen, G. (2005) Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. *Nucleic Acids Res.* **33**, 587–596.

Berg, O.G. (1988) Selection of DNA binding sites by regulatory proteins: the LexA protein and the arginine repressor use different strategies for functional specificity. *Nucleic Acids Res.* **16**, 5089–5105.

Birchler, J.A. and Han, F. (2009) Maize centromeres: structure, function, epigenetics. *Annu. Rev. Genet.* **43**, 287–303.

Carlson, S.R., Rudgers, G.W., Zieler, H., Mach, J.M., Luo, S., Grunden, E., Krol, C., Copenhaver, G.P. and Preuss, D. (2007) Meiotic transmission of an *in vitro*-assembled autonomous maize minichromosome. *PLoS Genet.* **3**, 1965–1974.

Chang, Y.L., Chuang, H.W., Meksem, K., Wu, F.C., Chang, C.Y., Zhang, M. and Zhang, H.B. (2011) Characterization of a plant-transformation-ready large-insert BIBAC library of Arabidopsis and bombardment transformation of a large-insert BIBAC of the library into tobacco. *Genome*, **54**, 437–447.

Cheng, Z., Buell, C.R., Wing, R.A. and Jiang, J. (2002) Resolution of fluorescence *in-situ* hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Res.* **10**, 379–387.

Christensen, A.H. and Quail, P.H. (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res.* **5**, 213–218.

Chu, G., Vollrath, D. and Davis, R.W. (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science*, **234**, 1582–1585.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629.

Ebersole, T.A., Ross, A., Clark, E., McGill, N., Schindelhauer, D., Cooke, H. and Grimes, B. (2000) Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. *Hum. Mol. Genet.* **9**, 1623–1631.

Friebe, B., Jackson, S.A., Zhang, P., Chen, W.P., Phillips, R.L., Muthukrishnan, S. and Gill, B.S. (2001) High-resolution structural analysis of biolistic transgene integration into the genome of wheat. *Theor. Appl. Genet.* **103**, 56–62.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T. and Cheeseman, I.M. (2011) Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell*, **145**, 410–422.

Gent, J.I., Schneider, K.L., Topp, C.N., Rodriguez, C., Presting, G.G. and Dawe, R.K. (2011) Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. *Epigenetics Chromatin*, **4**, 3.

Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, **40**, 767–774.

Halpin, C. (2005) Gene stacking in transgenic plants – the challenge for 21st century plant biotechnology. *Plant Biotechnol. J.* **3**, 141–155.

Hamilton, C.M., Frary, A., Lewis, C. and Tanksley, S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl Acad. Sci. USA*, **93**, 9975–9979.

Harrington, J.J., Bokkelen, G.V., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of *de novo* centromeres and construction of first-generation human artificial chromosomes. *Nat. Genet.* **15**, 345–355.

Henning, K.A., Novotny, E.A., Compton, S.T., Guan, X.Y., Liu, P.P. and Ashlock, M.A. (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. *Proc. Natl Acad. Sci. USA*, **96**, 592–597.

Houben, A., Dawe, R.K., Jiang, J. and Schubert, I. (2008) Engineered plant minichromosomes: a bottom-up success? *Plant Cell*, **20**, 8–10.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H. and Masumoto, H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotechnol.* **16**, 431–439.

Jackson, S.A., Wang, M.L., Goodman, H.M. and Jiang, J. (1998) Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. *Genome*, **41**, 566–572.

Kato, N. and Lam, E. (2001) Detection of chromosomes tagged with green fluorescent protein in live *Arabidopsis thaliana* plants. *Genome Biol.* **2**, RESEARCH0045.

Kato, A., Lamb, J.C. and Birchler, J.A. (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl Acad. Sci. USA*, **101**, 13554–13559.

Kiermaier, E., Woehrer, S., Peng, Y., Mechtler, K. and Westermann, S. (2009) A Dam1-based artificial kinetochore is sufficient to promote chromosome segregation in budding yeast. *Nat. Cell Biol.* **11**, 1109–1115.

Kumar, S., Arul, L., Talwar, D. and Raina, S.K. (2006) PCR amplification of minimal gene expression cassette: an alternative, low cost and easy approach to 'clean DNA' transformation. *Curr. Sci.* **91**, 930–934.

Lacefield, S., Lau, D.T. and Murray, A.W. (2009) Recruiting a microtubule-binding complex to DNA directs chromosome segregation in budding yeast. *Nat. Cell Biol.* **11**, 1116–1120.

Lanzer, M. and Bujard, H. (1988) Promoters largely determine the efficiency of repressor action. *Proc. Natl Acad. Sci. USA*, **85**, 8973–8977.

Liu, Y.G., Shirano, Y., Fukaki, H., Yanai, Y., Tasaka, M., Tabata, S. and Shibata, D. (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. *Proc. Natl Acad. Sci. USA*, **96**, 6535–6540.

Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* **109**, 1963–1973.

Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A. and Heun, P. (2011) *Drosophila* CENH3 is sufficient for centromere formation. *Science*, **334**, 686–690.

Michaelis, C., Ciosk, R. and Nasmyth, K. (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, **91**, 35–45.

Murata, M., Yokota, E., Shibata, F. and Kashihara, K. (2008) Functional analysis of the Arabidopsis centromere by T-DNA insertion-induced centromere breakage. *Proc. Natl Acad. Sci. USA*, **105**, 7511–7516.

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M. and Okazaki, T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.* **116**, 585–596.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature*, **305**, 189–193.

Nakano, A., Suzuki, G., Yamamoto, M., Turnbull, K., Rahman, S. and Mukai, Y. (2005) Rearrangements of large-insert T-DNAs in transgenic rice. *Mol. Genet. Genomics*, **273**, 123–129.

Nakano, M., Cardinale, S., Noskov, V.N., Gassmann, R., Vagnarelli, P., Kandels-Lewis, S., Larionov, V., Earnshaw, W.C. and Masumoto, H. (2008)

© 2011 The Authors

Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev. Cell*, **14**, 507–522.

Ohzeki, J., Nakano, M., Okada, T. and Masumoto, H. (2002) CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA. *J. Cell Biol.* **159**, 765–775.

Phan, B.H., Jin, W., Topp, C.N., Zhong, C.X., Jiang, J., Dawe, R.K. and Parrott, W.A. (2007) Transformation of rice with long DNA-segments consisting of random genomic DNA or centromere-specific DNA. *Transgenic Res.* **16**, 341–351.

Purnick, P.E. and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422.

Raikhel, N. (1992) Nuclear targeting in plants. *Plant Physiol.* **100**, 1627–1632.

Ravi, M. and Chan, S.W. (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature*, **464**, 615–618.

Reece, R.J. and Ptashne, M. (1993) Determinants of binding-site specificity among yeast C6 zinc cluster proteins. *Science*, **261**, 909–911.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A. and Belmont, A.S. (1996) *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700.

Sautter, C., Uze, M. and Potrykus, I. (1999) Single-stranded DNA in the genetic transformation of wheat (*Triticum aestivum* L.): transformation frequency and integration pattern. *Theor. Appl. Genet.* **99**, 487–495.

Shi, J. and Dawe, R.K. (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics*, **173**, 1571–1583.

Simons, A., Tils, D., von Wilcken-Bergmann, B. and Muller-Hill, B. (1984) Possible ideal *lac* operator: *Escherichia coli lac* operator-like sequences from eukaryotic genomes lack the central G.C pair. *Proc. Natl Acad. Sci. USA*, **81**, 1624–1628.

Song, J., Dong, F., Lilly, J.W., Stupar, R.M. and Jiang, J. (2001) Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA sequences. *Genome*, **44**, 463–469.

Song, J., Bradeen, J.M., Naess, S.K., Helgeson, J.P. and Jiang, J. (2003) BIBAC and TAC clones containing potato genomic DNA fragments larger than 100 kb are not stable in *Agrobacterium*. *Theor. Appl. Genet.* **107**, 958–964.

Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, M., Iwahara, J., Okazaki, T. and Yokoyama, S. (2001) Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J.* **20**, 6612–6618.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646.

Wang, C.J., Harper, L. and Cande, W.Z. (2006) High-resolution single-copy gene fluorescence *in situ* hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *Plant Cell*, **18**, 529–544.

Wolfgruber, T.K., Sharma, A., Schneider, K.L. *et al.* (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PLoS Genet.* **5**, e1000743.

Yan, H., Talbert, P.B., Lee, H.R., Jett, J., Henikoff, S., Chen, F. and Jiang, J. (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol.* **6**, e286.

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd,