Original Paper

Generation of stable engineered chromosomes in soybean

Yunfang Zhang[1,a], Asuka Itaya[1], Ping Fu[2], Suqin Zheng[1], Jacquie Hulm[2], Laureen Blahut-Beatty[1], Elizabeth-France Marillia[2,b], Michael Lindenbaum[2], Steven Fabijanski[2], Daina Simmonds[1,]*

[1] Eastern Cereal and Oilseed Research Centre, 960 Carling Avenue, Central Experimental Farm, Ottawa, ON K1A 0C6, Canada; [2] Agrisoma Biosciences Inc., 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada
* E-mail: daina.simmonds@agr.gc.ca   Tel & Fax: +1-613-759-1320

Abstract   A system for engineering plant chromosomes has been developed to facilitate the introduction of novel genes into the plant genome. The system is based on the establishment of a unique genetic locus within the ribosomal DNA (rDNA) region of the host chromosome to provide a permissive environment for expression of the introduced genes of interest (GOI). The genetic locus can exist within an independent, fully functional "minichromosome" (MC) or as a segment of a modified host chromosome (termed Engineered Trait Locus or ETL). The site-specific integration of transgenes to the rDNA locus isolates them from other endogenous genes, an advantage over conventional transformation in which foreign genes are inserted randomly into the host genome. Furthermore, MCs or ETLs can confer stability and high expression of the transgenes, as demonstrated in mammalian systems. To evaluate this system in plants, several MC and ETL lines have been generated in soybean, an important crop used worldwide for protein and oil consumption. The characterization of a soybean line containing an MC demonstrates that 1) the MC is stable over multiple generations as well as in field conditions, 2) maintaining the MC has no adverse phenotypic consequences, and 3) the MC can provide high-level expression of the introduced GOI.

Key words: Engineered chromosomes, artificial chromosomes, traits, stacking, soybean, field trial.

Transgenic technology has enabled a great range of improvements to be incorporated into crop plants by allowing access to novel and useful traits from a wide variety of genomes. However, conventional plant transformation faces some challenges and limitations. These include random insertion of the transgene into the host genome, often leading to disruption of endogenous gene(s); instability of the transgene due to positional effects or gene silencing; limitation in the number of genes that can be transferred simultaneously; and loss of unlinked transgenes in successive generations. A promising solution for these challenges is to employ engineered (or artificial) chromosomes, a technology successfully used in other organisms, including bacteria, yeast, and mammals (Hadlaczky 2001; Monaco and Larin 1994).

Engineered chromosomes can support the integration of a large amount of DNA and serve as exclusive repositories of introduced genes without affecting endogenous genes. Bacterial and yeast artificial chromosomes have been used extensively for three decades (Burke et al. 1987; Shizuya et al. 1992; Sternberg 1990), and mammalian artificial chromosomes were introduced fifteen years ago (Harrington et al. 1997). However, engineering of plant chromosome began only recently. In one example, maize cells transformed with a mixture of centromere specific sequences, telomere constituents and genes of interest (GOI) were shown to have assembled functional minichromosomes (MC) that could be transmitted through multiple generations of maize plants (Ananiev et al. 2009; Carlson et al. 2007). In another approach, an existing accessory chromosome of maize was truncated via targeted telomere-induced cleavage (Yu et al. 2007) and further modified to support targeted gene insertion technology (Birchler et al. 2008).

In an attempt to develop a universal approach applicable to a broad range of crop plants, we have adapted a technology, first used in mammalian cells, to generate heterochromatin-based MCs (Hadlaczky 2001). Satellite artificial chromosome (SATAC) technology (Hadlaczky 2001; Holló et al. 1996; Keresõ et al. 1996) or artificial chromosome engineering (ACE) technology

---

Transformation
Gene of interest, selectable marker
Homologous rDNA sequences for targeting

Target

Chromosome 13
with target rDNA

ETL chromosome

Co-integration
into rDNA locus
amplification
and ETL formation

Minichromosome

Cleavage generates MC
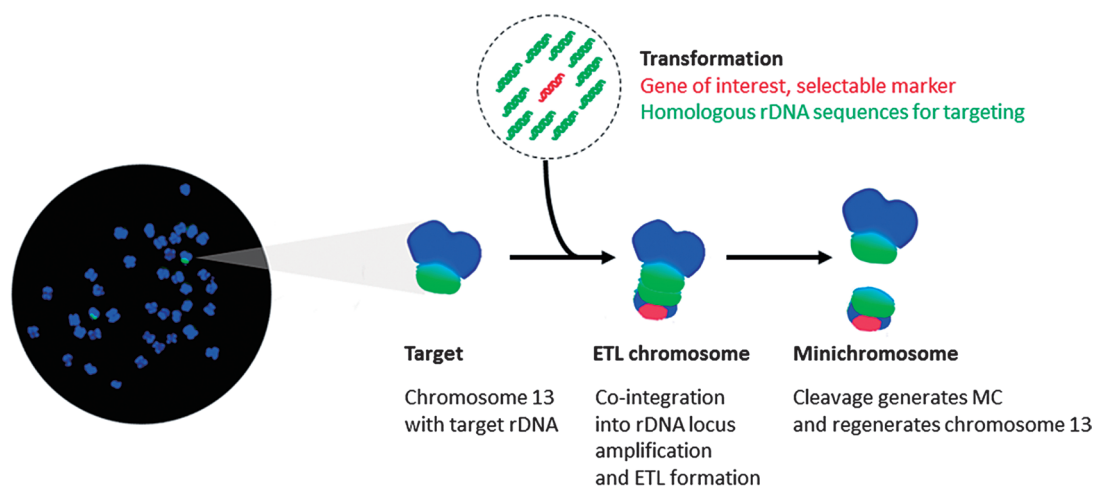and regenerates chromosome 13

Figure 1. MC and ETL generation. Transformation of a gene of interest (GOI) with excess amounts of targeting rDNA results in site-specific integration of GOI at rDNA locus of a target chromosome. Integration of foreign DNA into the rDNA arrays can trigger a large-scale amplification of the pericentric region resulting in formation of an engineered trait locus (ETL) that can be stably maintained on a resident host chromosome. In some cases, a dicentric chromosome is formed which ultimately resolves into a centric break-off product (minichromosome; MC) independent of the original host chromosome.

(Lindenbaum et al. 2004), makes use of a universal property of pericentric heterochromatic regions associated with ribosomal DNA (rDNA) gene loci to undergo large scale amplification upon integration of foreign DNA (Keresõ et al. 1996). Insertion of foreign DNA into rDNA regions has been potentiated by co-transformation with a large excess of homologous targeting DNA (such as a portion of the 45S rDNA gene) (Csonka et al. 2000; Lindenbaum et al. 2004). The resulting amplifications were comprised of large reiterated amplicon blocks each spanning 10 Mbp or more in size, and contained multiple copies of the foreign DNA in addition to the expected satellite markers and rDNA gene arrays. Due to the co-amplification of adjacent centromeric components, amplification often led to the generation of a dicentric chromosome as an intermediate, from which a stable break-off SATAC or MC could be generated, in addition to the regenerated host chromosome (Co et al. 2000; Telenius et al. 1999). The amplified chromosomal regions generated in this process have been stably maintained over multiple generations, in tissue culture (Telenius et al. 1999) or in transgenic animals (Co et al. 2000), and have provided an environment that supports consistent levels of gene expression. In addition, the amplified region can contain multiple recombination sites that allow multiple cycles of gene insertion into the pre-existing modified chromosome domain (Kennard et al. 2009a; Kennard et al. 2009b; Lindenbaum et al. 2004).

In the work described here, this technology has been applied to soybean (*Glycine max*), a crop species of considerable economic value. Two types of engineered chromosomes have been generated. One is an MC that exists independently, alongside normal host chromosomes. The other, termed engineered trait locus

(ETL), resides within an rDNA array of the modified host chromosome (Figure 1). Our study demonstrates the generation of MCs and ETLs in embryogenic soybean cultures, and the subsequent regeneration of genetically modified plants. Transmission of the MC through several generations of plants grown in controlled environmental chambers and one field trial confirmed the stability of the MC and its capacity to support a high level of transgene expression. The application of engineered chromosome technology in gene stacking is discussed.

Materials and methods

Soybean embryogenic culture
Soybean embryogenic cultures were prepared and transformed as previously described (Simmonds 2003). Embryogenic proliferative cultures were initiated from cotyledons of soybean cv. X5 (Agriculture and Agri-Food Canada breeding line X2650-7-3-2) and We (Ontario Agricultural College, Westag 97).

Plasmid construction and fragment purification
Two plasmids, pABI034 and pABI035, containing a hygromycin phosphotransferase gene (*hptII*) driven by a CaMV *35S* promoter (*35S:hpt*) were constructed using standard cloning techniques (Sambrook et al. 2001). Plasmids pABI034 and pABI035 contain an *E. coli*-derived *attB* sequence and a lambda phage-derived *attP* sequence (Landy 1989), respectively (Figure 2), incorporated to enable future site-specific gene insertion. Fragments from pABI034 were used in bombardment experiments 1 and 3, and those from pABI035 were used in experiments 2 and 4 (Table 1). Two versions of targeting sequences, a 1.4 kb segment of the *Arabidopsis thaliana* 26S nuclear rDNA gene or a 3.4 kb segment of the soybean 26S nuclear rDNA gene, were cloned giving rise to plasmids
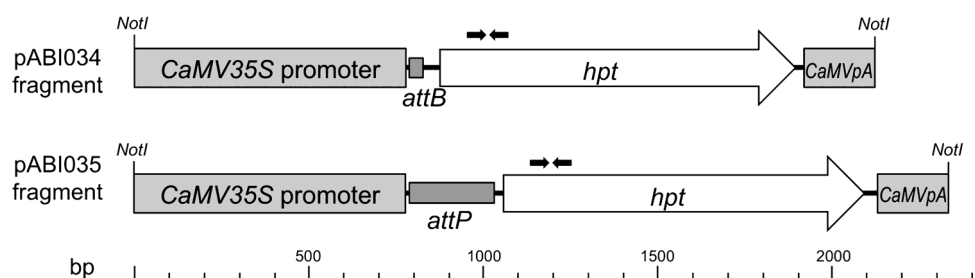
Figure 2. Gene of interest (GOI) used for transformation. DNA fragments containing a selectable marker, hygromycin phosphotransferase (*hpt*), driven by a CaMV *35S* promoter were used for transformation. Recombination sites *attB* or *attP* were incorporated to enable future site-specific gene insertion. pABI034 fragments were used in bombardment experiments 1 and 3, and pABI035 fragments were used in experiments 2 and 4 (Table 1). Arrows represent the position of primers used for qPCR or qRT-PCR analysis.

Table 1. Summary of four transformation experiments.

Experiment	26S rDNA	# of T0 events[a]	Events with *hpt* number >10	# of MC/ETL events, Name	*hpt* copy number [b] in MC/ETL
1	Arabidopsis	32	12	1, E1(mc)	25
2	Arabidopsis	34	10	0	
3	Soybean	31	13	1, E2(mc)	19
4	Soybean	23	7	1, E3(etl)	16

[a] Number of T0 events regenerated to plants. [b] Copy number at MC/ETL locus of hemizygous lines.

pJHD2-19A and pABI037, respectively. Construct details are provided as supplemental data. Transformation-quality plasmid DNA was prepared using the QIAprep Mini Prep kit (Qiagen Corp., Canada). The expression fragments, *35S:hpt* (containing either *attB* or *attP*), were excised via *Not*I restriction digest from pABI034 and pABI035, respectively. Targeting rDNA fragments were excised from pJHD2-19A and pABI037 via *Xho*I and *EcoR*I digest, respectively. The excised fragments were purified by extraction from agarose gel slices (QIAEX II Gel Extraction kit, Qiagen).

Biolistic transformation and plant regeneration
Bombardment of embryogenic soybean cultures followed a previously published procedure (Simmonds 2003). Three micrograms (3 µg) of gold particles (BioRad, Mississauga, Canada), 1 µm in diameter, were coated with a DNA mixture consisting of ca. 100 ng of the *35S:hpt* fragment and ca. 1300 ng of a targeting 26S rDNA fragment (1 : 10 molar ratio; Table 1). Tissue from 30-ml culture (approximately 150 mg fresh mass), was bombarded twice with approximately 1/15 of the total amount of coated gold particles per shot. Twelve cultures were bombarded for each experiment. Following bombardment, the embryogenic tissues were given a recovery period of 12–14 days. Selection was initiated at a low concentration of hygromycin (30 mg l⁻¹) and then moved to a higher concentration (55 mg l⁻¹). After 4 weeks of hygromycin selection, green colonies were first detected. The colonies (events) were collected weekly during the following 4–6 weeks. Individual events were cultured separately in multiwell plates and moved to larger containers for expansion. Following embryo maturation, desiccation and conversion, plants were regenerated. Typically 3 plants were potted per transgenic

event. In some cases root tips from converted embryos were harvested, blocked with nitrous oxide (Kato 1999), and processed for isolation of metaphase chromosomes for fluorescence in situ hybridization (FISH) analysis (see below). Leaf material from T0 plantlets was used for isolation of genomic DNA (gDNA) for PCR. T0 plants were maintained in growth chambers, and self-pollinated T1, T2, T3 and T4 seeds were used for transmission and segregation studies, as well as FISH analysis.

Names of transgenic events
The original names of transgenic events are simplified in this manuscript. The two events that produced the MC lines, 7B8X6B17 and 7B10W4C10, are referred to as E1(mc) and E2(mc) respectively. An ETL line, 7B11W10C3, is referred to as E3(etl). Non-MC/ETL lines, 7B9X10D3, 7B8W2A7, 7B10W4C5, 7B11X7D2, 7B8X11C1, 7B10X9C5, and 7B8X9A1 are referred to as NonMC/ETL (1), (2), (3), (4), (5), (6) and (7) respectively.

Genomic DNA purification and analysis
Genomic DNA samples were prepared from plant material using Qiagen DNeasy Plant Mini Kit (Qiagen) and were used for copy number analysis. Copy numbers were determined using real-time quantitative PCR (qPCR) and the values obtained were confirmed by Southern blot analysis (data not shown). A primer set and a probe for qPCR were designed for the *hpt* gene using an online service (Biosearch Technologies Inc., Novato, USA): forward: 5′-GGCGAAGAATCTCGTGCT TTCAG-3′; reverse: 5′-CATCGGCGCAGCTATTTACC-3′; probe: 5′FAM-CGCAGGACATATCCACGCCCTC-3′BHD.

Each qPCR reaction consisted of: 15 µl of 2X TaqMan ABI

Universal Master Mix (Life Technologies, Burlington, Canada), 1.8 µl each of 10 µM forward and reverse primers, 0.75 µl of 10 µM Probe, 0.6 µg of gDNA, and H$_2$O to a final volume of 30 µl. The reaction was performed as follows: Step 1, 95°C for 7 min; Step 2, 95°C for 45 s; Step 3, 63°C for 45 s; Step 4, plate read, then Steps 2–4 were repeated for a total of 45 cycles. For copy number standard, 0.6 µg of We or X5 genomic DNA [1.115 Mbp/1C (Arumuganathan and Earle 1991)] samples were spiked with 0, 2.7, 27, 54 and 81 pg of plasmid pABI034 (4.98 kbp) to provide the equivalent of 0, 1, 10, 20 and 30 copies, respectively. qPCR was performed with a CFX96 Real-Time PCR System and associated software (Bio-Rad, Mississauga, Canada).

RNA isolation and purification

Plant tissue samples (ca. 100 mg) were ground into a fine powder in liquid nitrogen with a mortar and pestle. One milliliter (1 ml) of TRIzol reagent (Invitrogen, Burlington, Canada) was added to the powder inside a microcentrifuge tube and RNA was extracted as directed by the manufacturer. The crude RNA was further purified using an RNeasy Mini Kit (Qiagen), following the manufacturer's protocols. Contaminating gDNA was removed by on-column digestion with RNase-Free DNase (Qiagen). RNA quantity was determined by measurements of absorbance at 260 nm, and RNA purity was assessed by measurements of absorbance at 280 nm (260 nm /280 nm routinely ≈2.0) and 230 nm (260 nm /230 nm routinely ≈2.0).

Quantitative RT-PCR analysis

For each sample analyzed for *hpt* expression, 1 µg of total RNA was taken for cDNA synthesis using an iScript cDNA Synthesis Kit (Bio-Rad) according to manufacturer's instructions. Following completion of the reverse transcription (RT) reactions, serial dilutions for a standard curve were made as follows: cDNAs (3 µl of each RT product) were pooled and serially diluted with nuclease free water to obtain 2, 8, 32 and 128-fold dilutions. The *hpt* gene was amplified by the same primer set used for copy number analysis (see above). For the internal control, soybean *β*-actin gene was amplified using the following primers in a separate reaction: forward: 5′-TCC AAG GGG ACC TAA CGG AGA-3′; reverse: 5′-TGG GTC AAG AGC TGG ATG GTG-3′. qPCR reactions consisted of: 10 µl of 2X IQ SYBR Green supermix (Bio-Rad), 1 µl of forward and reverse primers (10 µM each), 1 µl of 10-fold diluted RT reaction mix (corresponding to 5 pg of input RNA) and water to a final volume of 20 µl. The reaction was carried out as follows: Step 1, 95°C for 10 min; Step 2, 95°C for 15 s; Step 3, 60°C for 15 s; Step 4, 72°C for 15 s, Step 5, plate read, then Steps 2–5 were repeated for a total of 45 cycles. After 45 cycles, the temperature was increased from 50 to 99°C at a rate of 0.2°C s^{-1}, and the fluorescence was measured continuously to construct the melting curves. All PCR reactions were run in duplicate for each sample. Relative expression levels of the *hpt* gene were presented as ratios to actin expression.

Fluorescence in situ hybridization (FISH) analysis

Root-tips from T1, T2, T3 and T4 germinated seeds were treated with nitrous oxide to block cells in metaphase (Kato 1999). Spreads of chromosomes at metaphase were hybridized with Alexa Fluor 488 Green or Texas red-labeled probes against soybean 18S rDNA or *hpt* respectively as previously described (Kato et al. 2006). Probes were prepared by incorporation of fluorescent nucleotides by nick translation, essentially as described (Kato et al. 2006).

Confined field trials

Trials were initiated on June 9, 2009 and carried out according to a randomized complete block design with 4-row plots in 5-meter rows. Each four-row plot included homozygous E1(mc)-T3 plants, one null line, one *iso*-line (X5) and 2 check-lines (AC-Colibri and Chikala). Nulls and parents (*iso*-lines) of the E1(mc) line served as comparisons to the test materials, while unmodified check line plots served as comparator to commercial lines. The blocks were replicated 4 times. The trial site was reproductively isolated from other *Glycine* species by a minimum of 10 m of fallow land.

Seed analysis

Seed from trial-grown material was subjected to compositional analysis; oil, carbohydrate and protein content were analyzed with an Infratec 1241 analyzer (FOSS NIRSystems, Eden Prairie, MN). Briefly, 500 g of seed (from field-grown wild type X5 or E1(mc)-T3 plants) were divided into 7 sub-samples which were analyzed independently as replicates; the data was analyzed using the two-tailed Students *t*-test.

Results

Minichromosome (MC) and Engineered Trait Loci (ETL) generation

The MC/ETL formation relies on the targeted introduction of foreign DNA sequences into the nuclear rDNA gene arrays found in the short arm pericentric regions of a subset of acrocentric chromosomes (Figure 1). To apply this technology to soybean, the GOI [hygromycin phosphotransferase (*hpt*), driven by a cauliflower mosaic virus *35S* promoter (*35S:hpt*)] was targeted to rDNA arrays by using ten times the molecular ratio of 26S rDNA. Four experiments were conducted to transform soybean with a mixture of the *35S:hpt* fragment and the targeting fragment [26S rDNA from *Arabidopsis* (Unfried and Gruendler 1990) or soybean] (Table 1). A range of 29 to 46 hygromycin-resistant events was recovered from each experiment and from these, T0 plants were regenerated from 23 to 34 events. This rate of plant recovery (67–94%) is comparable to typical biolistic transformation of soybean embryogenic cultures. qPCR analysis suggested that more than half of the events were within the range of 1–10 copies of integrated *35S:hpt* transgenes, and 30–40% of the events
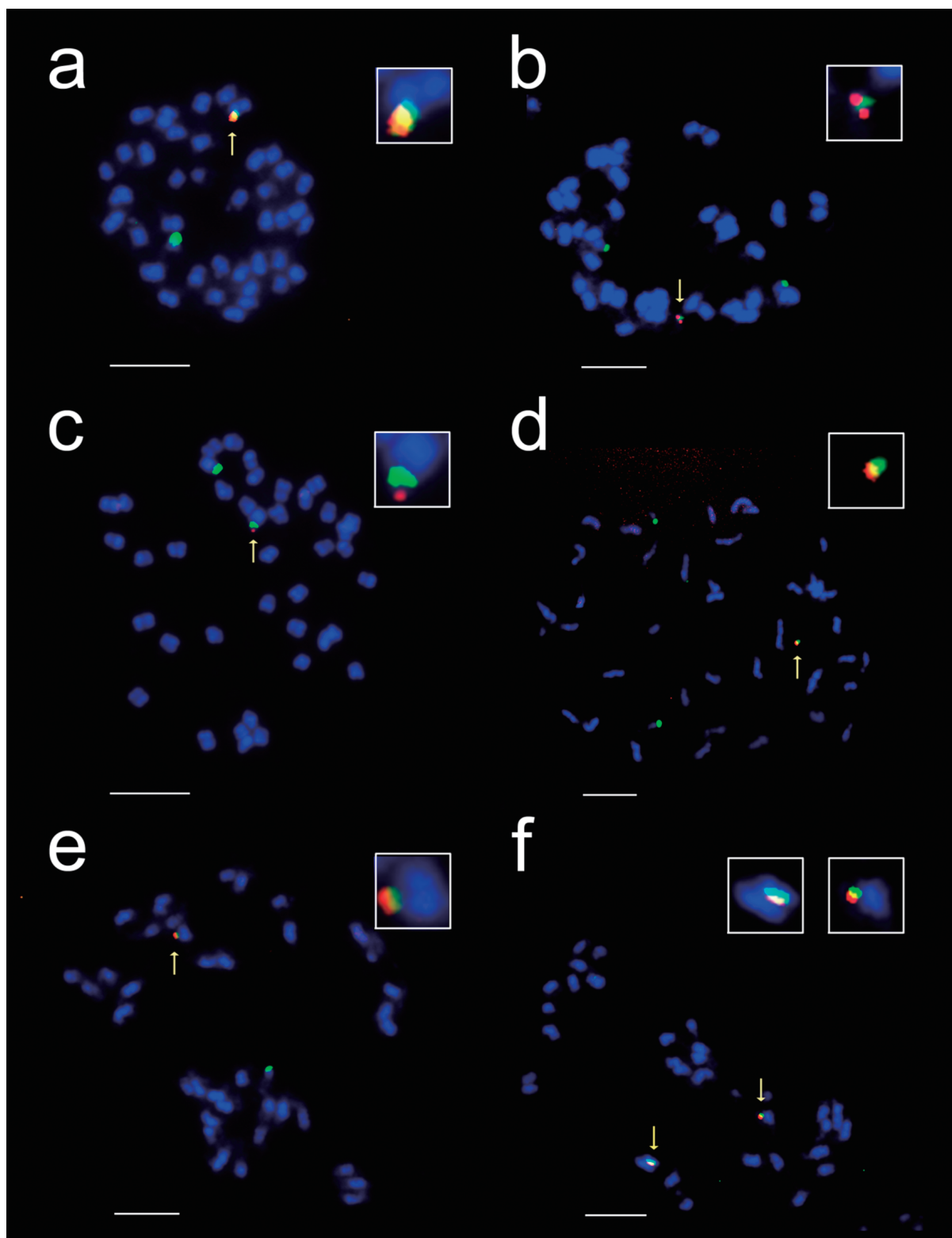
Figure 3.   FISH analysis of E1(mc), E2(mc) and E3(etl) T1 metaphasic chromosomes. Metaphasic chromosomes from pooled T1 root-tips were spread on slides and hybridized with a mixture of the *hpt* probe (Texas red) and 18S rDNA probe (Alexa Fluor 488 Green). FISH shows: (a) ETL chromosome (arrow) in E1(mc), (b) hemizygous MC (arrow) in E1(mc), (c) ETL chromosome (arrow) in E2(mc), (d) hemizygous MC (arrow) in E2(mc), (e) hemizygous and (f) homozygous ETL chromosomes (arrow) in E3(etl). Inserts show enlarged images of MCs or ETLs. Scale bar=5 μm.

exceeded 10 copies (Table 1). Fluorescence in situ hybridization (FISH) analyses were conducted to identify MC/ETL. Events with 10 or more copies were analyzed

because of the limited sensitivity of FISH. Of the 42 high-copy events, 23 were analyzed conclusively by FISH while the remainder were not analyzed due to poor seed
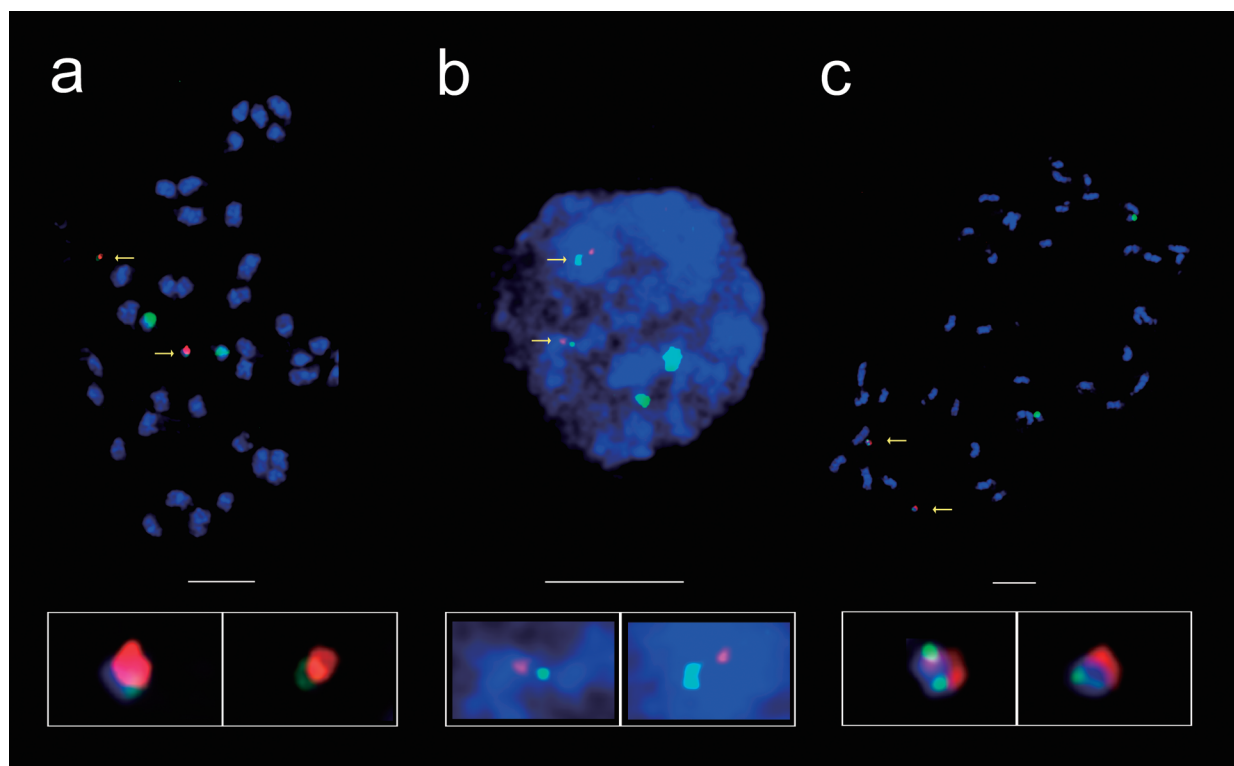
Figure 4.   FISH analysis of E1(mc) T3 and T4 chromosomes. Metaphase and interphase chromosomes from individual T3 (a–b) or T4 (c) homozygous root-tips were analyzed. The chromosomes were hybridized with a mixture of *hpt* probe (Texas red) and 18S rDNA probe (Alexa Fluor 488 Green). (a) T3 metaphasic chromosomes containing an MC pair (arrows). (b) T3 interphase nuclei showing the position of the MCs (arrows) and illustrating the spatial separation from the chromosome 13 pair. (c) T4 metaphasic chromosomes containing two MCs (arrows). Inserts show enlarged images of MCs. Scale bar=5 μm.

germination or plant sterility. Three events, observed at the T1 stage, contained amplified *35S:hpt* signals co-localized to endogenous rDNA-containing domains, indicating that MC/ETLs were generated. These three events are E1(mc), E2(mc) and E3(etl) (Table 1).

Event 1 with MC, E1(mc)

E1(mc) generated in experiment 1 in the X5 background contained approximately 25 copies of the *35S:hpt* gene (Table 1). FISH analysis showed large-scale amplification of *35S:hpt* sequences co-localized to the endogenous rDNA locus of chromosome 13, resulting in the formation of an ETL (Figure 3a). A novel MC was also observed in a small proportion (ca. 5%) of T1 metaphasic cells obtained from roots of pooled seedlings (Figure 3b). This MC contained co-localized endogenous rDNA and *35S:hpt* sequences and was maintained separately from the chromosome 13 pair. It is clear that the origin of the MC was a break-off product of the chromosome 13-derived ETL since 18S rDNA sequences and *35S:hpt* inserts co-localized, and 18S rDNA genes are known to be exclusive to the short arms of soybean chromosome 13 (Griffor et al. 1991; Skorupska et al. 1989). In metaphasic cells containing the MC, the chromosome 13 pair showed no residual *35S:hpt* staining; no insertions of targeting rDNA or *35S:hpt* signal were visible on other
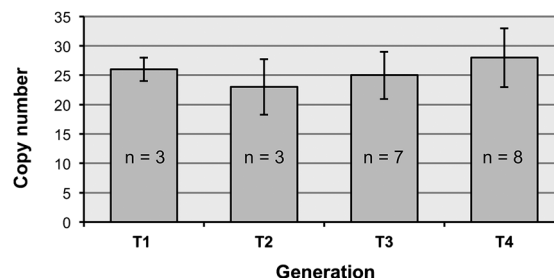

Figure 5.   *Hpt* copy number in the E1(mc) event over multiple generations. DNA was purified from leaf punches of homozygous E1(mc) plants (T1, T2, T3 and T4) and *hpt* copy number determined by qPCR. For T1 samples, zygosity was confirmed by FISH analysis. Bars show the average copy number per haploid genome; error bars indicate standard deviations of the means. The numbers inside the bars indicate the number of plants used in the analysis.

host chromosomes.

Event 2 with MC, E2(mc)

Thirteen events from experiment 3 were identified as containing *35S:hpt* copy numbers greater than 10. Of these events, 7 were analyzed by FISH conclusively. One event generated in the We background, E2(mc), with approximately 19 copies of *35S:hpt* transgenes was shown to contain an ETL (Table 1 and Figure 3c). Moreover, similar to the E1(mc) T1 seedlings, a proportion of

Table 2. *Hpt* gene expression in T2 seedlings of MC/ETL and non-MC/ETL events

Events	*hpt* copy number[a]	Expression[b]	SD[c] (n)
E1(mc)	25	3.1875	0.825 (6)
E2(mc)	19	0.0927	0.083 (6)
E3(etl)	16	0.0030	0.003 (6)
NonMC/ETL(1)	5	1.3408	0.213 (3)
NonMC/ETL(2)	17–21	0.9814	0.752 (2)
NonMC/ETL(3)	24–30	0.0336	0.018 (3)
NonMC/ETL(4)	20–25	0.0048	0.000 (3)
NonMC/ETL(5)	25	0.0037	0.001 (3)
NonMC/ETL(6)	ND[d]	0.0009	0.001 (3)
NonMC/ETL(7)	25	0.0003	0.001 (2)

[a] Copy number per haploid genome. [b] Expression level of *hpt* (normalized by internal *actin* expression). [c] SD, Standard deviation; n, number of biological replicates. [d] Not determined.

metaphasic cells contained an MC (Figure 3d).

Event 3 with ETL, E3(etl)

Of the 23 events isolated from experiment 4, 7 showed a copy number of 10 or higher. Of these, 4 were analyzed by FISH and one event, E3(etl), generated in the We background, had approximately 16 copies of *35S:hpt* transgenes and contained an ETL (Table 1 and Figure 3e). Unlike the E1(mc) and E2(mc) events, only ETLs were observed; no MCs were detected. FISH analysis of individual T1 progeny revealed presence of a pair of ETL chromosomes and absence of other chromosomes with rDNA labeling, suggesting the generation of a homozygous ETL line (Figure 3f).

Characterization of E1(mc)

E1(mc) was chosen for further analysis because of its high-expression level of the *hpt* gene (see next section) and its excellent fertility. The heterogeneous distribution and relative rarity of MC in T1 chromosome spreads of E1(mc) and E2(mc) events suggested that the formation of the dicentric chromosome and break-off occurred only in a sub-population of cells in the initial T0 event. Spontaneous generation of MC from chromosome 13 in un-transformed soybean was not observed (data not shown). Since the initial material analyzed by FISH was from pooled T1 seeds, it was not possible to determine whether the MC had arisen from a break off from an ETL chromosome occurring at a late stage in the development of T1 seeds, or whether it represented a rare T0 plant event that was transmitted through the germline. T2 progeny of the transgenic line E1(mc) were screened by PCR to determine the genotype of each T1 plant. Of the 18 T1 plants analyzed, 6 were homozygous for the *35S:hpt*, 9 were hemizygous and 3 were null.

Metaphasic chromosomes of homozygous T3 plants of E1(mc), (E1(mc)-T3), were analyzed by FISH. The analysis was carried out on individuals rather than a pooled population. Several of the T3 individuals showed an increased proportion of MCs, with as many as 80% of metaphase spreads containing a pair of independent

MCs (Figure 4a). The remainder contained either a single MC and an ETL-like chromosome or two ETL-like chromosomes (data not shown). Independent MCs were also apparent in interphase nuclei (Figure 4b). qPCR analysis performed on DNA obtained from multiple replicate plant samples showed that over 4 generations, the *35S:hpt* copy number of the E1(mc) line remained fairly constant (Figure 5), confirming that no additional large-scale amplification was occurring in generations following the initial amplification event and that, in the absence of selection pressure, the amplified regions remained stable.

MC/ETL gene expression

Expression levels of the *hpt* gene were analyzed in MC/ETL and non-MC/ETL lines. The first true leaves and stems of 7 days-old T2 seedlings were used for analysis. Among the 10 events tested, E1(mc) exhibited the highest level of *hpt* expression (Table 2), indicating that an MC could support high-level expression of a transgene(s). Expression of *hpt* was also analyzed in various tissues from E1(mc) T3 plants grown in a controlled environment cabinet and in the field (Table 3). *Hpt* expression of the E1(mc) line was higher than that of a homozygous non-MC/ETL(5) with similar copy number (approximately 25 copies of *35S:hpt* per haploid genome) in every tissue examined (Table 3). The result indicates that an MC can provide stable and high expression of a transgene through generations as well as in different growth conditions.

Field trial of E1(mc) line

The confined field trial of the E1(mc) line demonstrated that the T3 plants and T4 seeds had the same agronomic properties as the parental line in seed vigor, flower color, lodging, absence of nutrient deficiency symptoms and other properties listed in Table 4 but not in yield. Both the null line and the E1(mc) line showed a lower average seed yield per plot when compared to the parental line; however this was in fact due to the adventitious predation by groundhogs which disproportionately

Table 3. Expression analysis of *hpt* gene in tissues of E1(mc) T3 plants grown in an environmentally controlled growth cabinet and in the field. A non-MC/ETL line with similar transgene copy numbers of 25 per haploid genome was analyzed for comparison.

| | E1(mc)-T3 | | | | NonMC/ETL(5)-T2 | |
| | Growth cabinet | | Field | | Growth cabinet | |
Tissue	Expression[a]	SD (n)[b]	Expression	SD (n)	Expression	SD (n)
V6[c] leaf	5.003	1.35 (6)	3.473	2.81 (10)	0.018	0.01 (5)
R2[d] flower	0.660	0.23 (6)	1.061	0.54 (8)	0.016	0.01 (5)
R5 seed	0.113	0.01 (5)	0.279	0.25 (3)	0.006	0.003 (5)
R6 seed	0.458	0.09 (5)	0.387	0.22 (8)	0.004	0.001 (5)
R6 leaf	2.716	0.32 (5)	ND[e]	ND	0.490	0.22 (5)

[a] Expression level of *hpt* (normalized by internal *actin* expression). [b] SD, Standard deviation; n, number of biological replicates. [c] V6 leaf, young leaf at the 6th node of a vegetative plant. [d] R2 flower, flower of plant at full bloom; R5 seed–seed at early development; R6 seed, full green seed; R6 leaf, mature green leaf. [e] Not determined.

Table 4. Agronomic data for 2009 field trial of E1(mc)-T3 plants. The agronomic data is shown for E1(mc), parental wild type X5 and null lines grown in small plot randomized trials.

	X5 parent	Null (T3)	E1(mc)-T3
Flowering (days)[a]	43	43	43
Plant height (cm)	74–80	72–77	76–80
Pest incidence[b]	aphids at V6	aphids at V6 groundhogs at R5-6	aphids at V6 groundhogs at R5-6
Maturity (days)	126	126	126
Shattering[c]	2	2	2
Seed weight (g/100 seeds)	10.18	8.87	10.88
Per plot seed weight (g)	1786±134	1161±443	1300±172

[a] Days required for 50% of plants to be in flower. [b] V6, vegetative plant at the 6th internode stage; R5-6, early to late reproductive stage. [c] For mature dry pods, a score of 1 is no pod shattering and 9 is 100% shattering.

affected both of these lines (data not shown). We have observed that seed yield of the E1(mc) line in a growth chamber is comparable to that of the X5 parental line (data not shown). Additional field trials in a wider range of locations over a number of years would be necessary to obtain a definitive picture of the E1(mc) line performance, however this preliminary study confirmed that the MC did not confer deleterious phenotypic effects.

Analysis of field-grown T4 seeds showed no statistical differences between E1(mc) and the X5 parental line with regard to protein, sugar, oil and complex carbohydrate composition (Figure 6).

The stability of MC in field-grown plants was evaluated in roots from field-derived E1(mc)-T4 plantlets. A representative example of FISH analysis of T4 material shows that the size and staining pattern of T4 MCs (Figure 4c) are comparable to those of T3 (Figure 4a).

Discussion

In this study, MCs and ETLs were successfully generated in soybean by a modification of techniques previously used to create mammalian artificial chromosomes (Csonka et al. 2000; Lindenbaum et al. 2004). The analysis of E1(mc) demonstrates that MCs can provide stable and high-level gene expression, and that MCs can be maintained without adverse phenotypic consequences. The MC structural stability was demonstrated by the
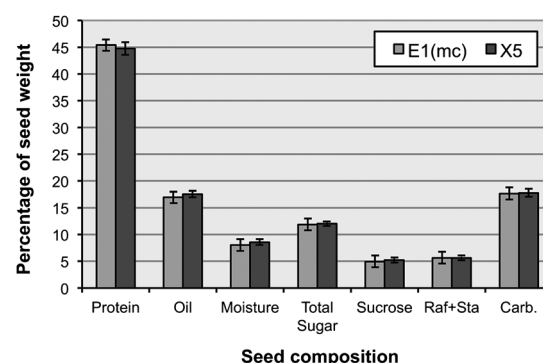
Figure 6. Analysis of T4 seeds from line E1(mc) field trial. T4 seeds from field-grown E1(mc) and parental X5 wild-type controls were analysed for composition as a percent of total seed weight. Bars show the average of 7 seed batches for each analysis. The error bars indicate standard deviations of the means. Raf+Sta: raffinose+stachyose, Carb.: complex carbohydrates. No significant difference was found between the E1(mc) and the X5 data for each category (two tailed Student *t*-test).

consistency of the *35S:hpt* copy number over the course of multiple generations (up to T4), including a generation maintained under field conditions. This suggests that the amplicon structure is not subject to additional rounds of amplification in subsequent generations. This is consistent with a mechanism by which the insertion of foreign DNA elicits a one-time amplification event after which the amplicon structure is stabilized. It is also consistent with the known properties of plant rDNA loci which, despite their highly repetitive nature, have been shown to be recombination repressed (Copenhaver et al.

1995).

Arabidopsis (1.5 kb) or soybean (2.6 kb) 26S rDNAs were tested as targeting DNA in this study. While both MCs and ETLs were obtained by using *Arabidopsis* or soybean 26S rDNA segments, more work will be required to determine whether the longer sequence and/or greater overall homology of the soybean targeting sequence offers any advantage in MC or ETL formation.

Interestingly, an MC was observed only in a small percentage of T1 metaphasic cells in the E1(mc) line. However by the T3 generation, the proportion of metaphasic cells containing homozygous MCs increased to nearly 80% in individual plantlets. The remaining 20% of metaphasic cells showed ETLs on chromosome 13. It is difficult to explain the persistent, albeit infrequent presence of these ETL-like chromosomes in T3 and T4 generations. One possibility is that the MCs are quasi-stable and are the result of break-offs occurring at each generation during early development. However, based on the high proportion of MCs in T3 and subsequent generations, it is more likely that the MCs are stable entities that can undergo meiosis and are transmitted through the germ line. The 20% metaphasic cells showing ETL-like chromosomes may be arising from fusions between the MCs and original chromosome 13, given the homologies between the MCs and rDNA locus of chromosome 13 and possible lack of telomere regeneration on MC (unpublished observation) (McClintock 1941). Analysis of MC and ETL segregation during germ cell differentiation may provide additional insights into MC and ETL transmission. Furthermore, it is possible that the ETL-like chromosomes in the E1(mc) line may ultimately be eliminated following multiple generations, as has been the case for stabilizing variegation (McClintock 1941).

An important property of MC/ETL loci in mammalian cells is their capacity to support high-level expression of introduced heterologous genes (Kennard et al. 2009b). While polymerase II-type promoters are not normally found in rDNA arrays, the rDNA intergenic spacer region of plants has been shown to contain a highly active polymerase II-type promoter/enhancer activity (Borisjuk et al. 2000; Schlögelhofer et al. 2002), which may contribute to the expression levels observed for genes within MCs or ETLs. Analysis of the E1(mc) suggests that plant MC/ETLs also have the potential to support high-level expression; however, generation and analysis of more MC/ETL plants is required to confirm such an advantage of MC/ETLs over conventional transformation.

One of the most important issues in current plant biotechnology is the question of how to introduce multigenic traits (gene stacking) into a plant species. While gene stacking can be done through conventional transgenic approaches, technical challenges remain. For example, genes of a multigenic trait can be introduced individually and then integrated via crossing of the individual events, a time consuming process that results in the transgenes being spread amongst multiple chromosomes, each segregating independently. Alternatively, multiple genes can be introduced into a host plant simultaneously, however, the outcome is unpredictable as lines with different combinations of transgenes and phenotypes will be generated (Zhu et al. 2008). To avoid segregation of transgenes, up to 9 genes have been cloned into a single vector for plant transformation (Wu et al. 2005). Unfortunately, construction of such vectors can be laborious and difficult, and transformation efficiency generally declines as the size of the foreign DNA increases (Simmonds et al., unpublished data). Engineered chromosomes can provide a more effective method for gene stacking by overcoming the challenges described above. We have recently created soybean ETL lines with a larger DNA fragment (10.4 kbp) containing multiple GOIs and confirmed expression of GOIs and the corresponding phenotypes/traits (unpublished data). To facilitate site-specific insertion and stacking of multiple genes at MC/ETL loci, our soybean MC/ETL lines have been engineered with lambda phage- and *E. coli*-derived *att* recombination sites (Landy 1989) (Figure 2) that allow site-specific unidirectional gene insertion (Kennard et al. 2009a; Kennard et al. 2009b; Lindenbaum et al. 2004). Site-specific integration of a transgene, mediated by recombination, has been demonstrated for soybean (Li et al. 2009). The combination of MC/ETL platforms with targeted gene insertion sites should significantly simplify and accelerate production of crops with valuable multigenic traits.
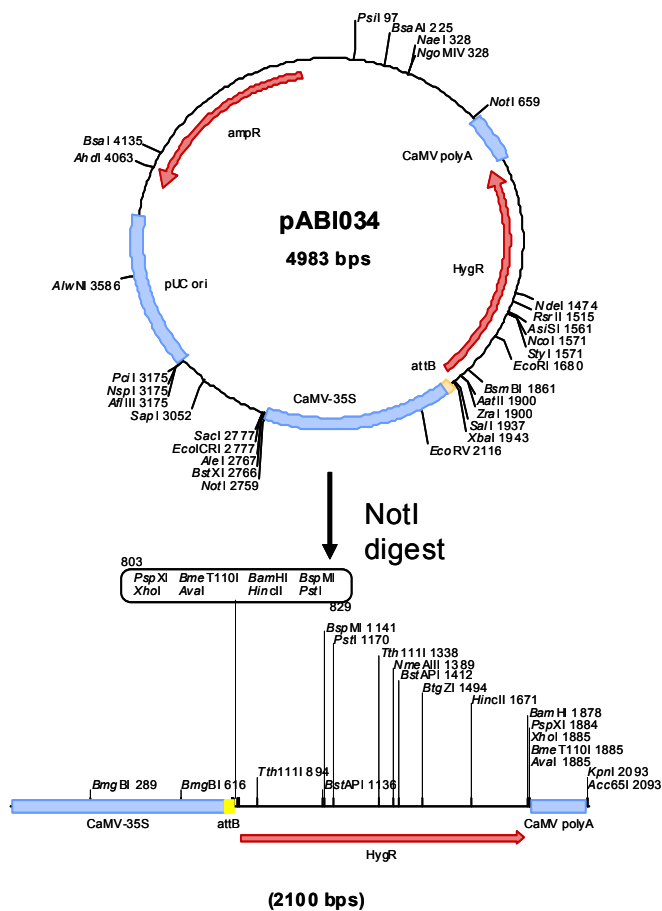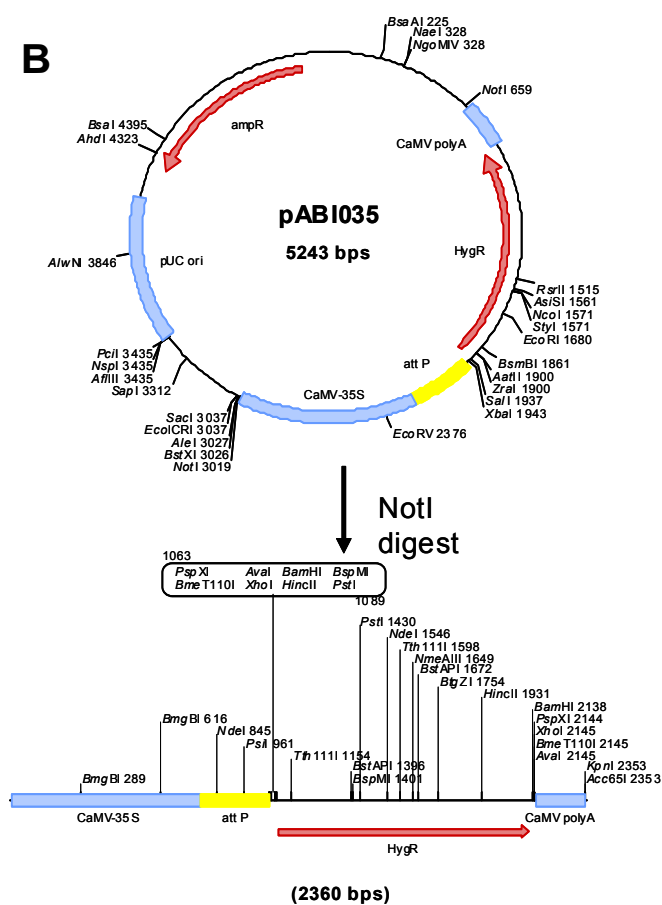
Acknowledgements

References

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

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208–218

Birchler JA, Yu W, Han F (2008) Plant engineered minichromosomes and artificial chromosome platforms. *Cytogenet Genome Res* 120: 228–232

Borisjuk N, Borisjuk L, Komarnytsky S, Timeva S, Hemleben V, Gleba Y, Raskin I (2000) Tobacco ribosomal DNA spacer element stimulates amplification and expression of heterologous genes. *Nat Biotechnol* 18: 1303–1306

Burke D, Carle G, Olson M (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome

vectors. *Science* 236: 806–812

Carlson SR, Rudgers GW, Zieler H, Mach JM, Luo S, Grunden E, Krol C, Copenhaver GP, Preuss D (2007) Meiotic Transmission of an In Vitro Assembled Autonomous Maize Minichromosome. *PLoS Genet* 3: e179

Co D, Borowski A, Leung J, van der Kaa J, Hengst S, Platenburg G, Pieper F, Perez C, Jirik F, Drayer J (2000) Generation of Transgenic Mice and Germline Transmission of a Mammalian Artificial Chromosome Introduced into Embryos by Pronuclear Microinjection. *Chromosome Res* 8: 183–191

Copenhaver GP, Doelling JH, Gens JS, Pikaard CS (1995) Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in *Arabidopsis thaliana*. *Plant J* 7: 273–286

Csonka E, Cserpan I, Fodor K, Hollo G, Katona R, Kereso J, Praznovszky T, Szakal B, Telenius A, deJong G, et al. (2000) Novel generation of human satellite DNA-based artificial chromosomes in mammalian cells. *J Cell Sci* 113: 3207–3216

Griffor MC, Vodkin LO, Singh RJ, Hymowitz T (1991) Fluorescent in situ hybridization to soybean metaphase chromosomes. *Plant Mol Biol* 17: 101–109

Hadlaczky G (2001) Satellite DNA-based artificial chromosomes for use in gene therapy. *Curr Opin Mol Ther* 3: 125–132

Harrington JJ, Bokkelen GV, Mays RW, Gustashaw K, Willard HF (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* 15: 345–355

Holló G, Keresõ J, Praznovszky T, Cserpán I, Fodor K, Katona R, Csonka E, Fátyol K, Szeles A, Szalay A, et al. (1996) Evidence for a megareplicon covering megabases of centromeric chromosome segments. *Chromosome Res* 4: 240–247

Kato A (1999) Air drying method using nitrous oxide for chromosome counting in maize. *Biotech Histochem* 74: 160–166

Kato A, Albert P, Vega J, Birchler J (2006) Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotech Histochem* 81: 71–78

Kennard ML, Goosney DL, Monteith D, Roe S, Fischer D, Mott J (2009a) Auditioning of CHO host cell lines using the artificial chromosome expression (ACE) technology. *Biotechnol Bioeng* 104: 526–539

Kennard ML, Goosney DL, Monteith D, Zhang L, Moffat M, Fischer D, Mott J (2009b) The generation of stable, high MAb expressing CHO cell lines based on the artificial chromosome expression (ACE) technology. *Biotechnol Bioeng* 104: 540–553

Keresõ J, Praznovszky T, Cserpán I, Fodor K, Katona R, Csonka E, Fátyol K, Holló G, Szeles A, Ross A, et al. (1996) De novo chromosome formations by large-scale amplification of the centromeric region of mouse chromosomes. *Chromosome Res* 4: 226–239

Landy A (1989) Dynamic, Structural, and Regulatory Aspects of lambda Site-Specific Recombination. *Annu Rev Biochem* 58: 913–949

Li Z, Xing A, Moon BP, McCardell RP, Mills K, Falco SC (2009) Site-Specific Integration of Transgenes in Soybean via Recombinase-Mediated DNA Cassette Exchange. *Plant Physiol* 151: 1087–1095

Lindenbaum M, Perkins E, Csonka E, Fleming E, Garcia L, Greene A, Gung L, Hadlaczky G, Lee E, Leung J, et al. (2004) A mammalian artificial chromosome engineering system (ACE System) applicable to biopharmaceutical protein production, transgenesis and gene-based cell therapy. *Nucl Acids Res* 32: e172

McClintock B (1941) The stability of broken ends of chromosomes in zea mays. *Genetics* 26: 234–282

Monaco AP, Larin Z (1994) YACs, BACs, PACs and MACs: Artificial chromosomes as research tools. *Trends Biotechnol* 12: 280–286

Sambrook J, MacCallum P, Russell D (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York

Schlögelhofer P, Nizhynska V, Feik N, Chambon C, Potuschak T, Wanzenböck E-M, Schweizer D, Bachmair A (2002) The upstream Sal repeat-containing segment of *Arabidopsis thaliana* ribosomal DNA intergenic region (IGR) enhances the activity of adjacent protein-coding genes. *Plant Mol Biol* 49: 653–663

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci USA* 89: 8794–8797

Simmonds D (2003) Genetic Transformation of Soybean with Biolistics. In: Jackson JF, Linskens HF (eds) *Genetic Transformation of Plants*. Springer Verlag, Berlin, pp 159–174

Skorupska H, Albertsen MC, Langholz KD, Palmer RG (1989) Detection of ribosomal RNA genes in soybean, Glycine max (L.) Merr., by in situ hybridization. *Genome* 32: 1091–1095

Sternberg N (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc Natl Acad Sci USA* 87: 103–107

Telenius H, Szeles A, Keresö J, Csonka E, Praznovszky T, Imreh S, Maxwell A, Perez C, Drayer J, Hadlaczky G (1999) Stability of a Functional Murine Satellite DNA-based Artificial Chromosome Across Mammalian Species. *Chromosome Res* 7: 3–7

Unfried I, Gruendler P (1990) Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucl Acids Res* 18: 4011

Wu G, Truksa M, Datla N, Vrinten P, Bauer J, Zank T, Cirpus P, Heinz E, Qiu X (2005) Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat Biotechnol* 23: 1013–1017

Yu W, Han F, Gao Z, Vega JM, Birchler JA (2007) Construction and behavior of engineered minichromosomes in maize. *Proc Natl Acad Sci USA* 104: 8924–8929

Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA* 105: 18232–18237
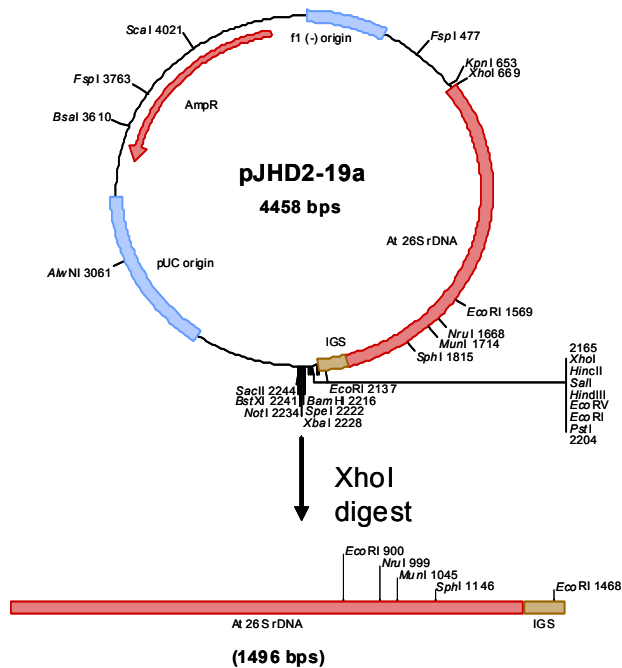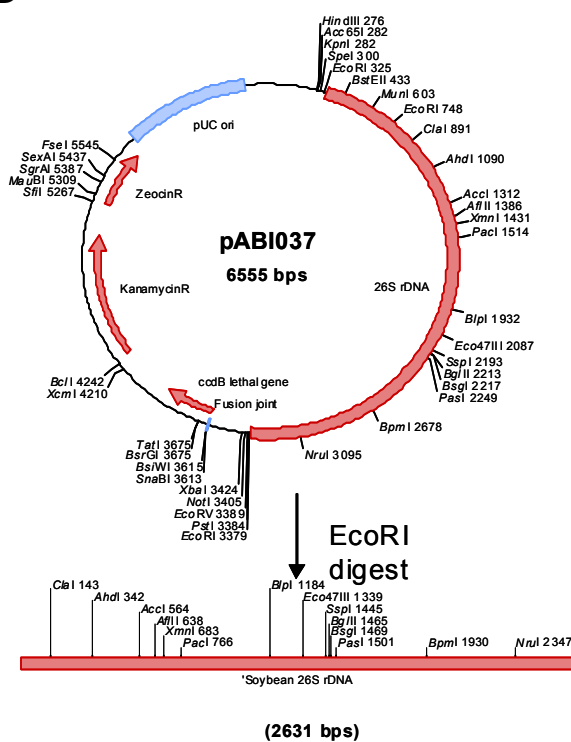
A

pABI034
4983 bps

NotI digest

(2100 bps)

B

pABI035
5243 bps

NotI digest

(2360 bps)

C

pJHD2-19a
4458 bps

XhoI digest

(1496 bps)

D

pABI037
6555 bps

EcoRI digest

(2631 bps)

Molecule: pABI034, 4983 bps DNA Circular

Start	End	Orientation	Name	Comment
1	440		pBluescript SK+ backbone	
441	135	C	f1 (+) origin	From pBLuescript
653	657		Multiple cloning site	From pBLuescript
657	460	C	B-galactosidase a-fragment	From pBLuescript
875	673	C	CaMV 35S polyA	Genbank accession V00140
1928	903	C	Hygromycin resistance	Genbank accession V01499.1
1989	1957	C	attB	Genbank accession J01638
2758	1989	C	CaMV-35S promoter	Genbank accession E01311.1
2761	2782		Multiple cloning site	From pBLuescript
2838	2760	C	B-galactosidase a-fragment	From pBLuescript
2960	2839	C	P-Lac	From pBLuescript
3180	3840		pUC origin	From pBLuescript
4855	3998	C	Amp Resistance	From pBLuescript
4856	4956		pBluescript SK+ backbone	

Molecule: pABI035, 5243 bps DNA Circular

Start	End	Orientation	Name	Comment
1	440		pBluescript SK+ backbone	
441	135	C	f1 (+) origin	From pBLuescript
653	657		Multiple cloning site	From pBLuescript
657	460	C	B-galactosidase a-fragment	From pBLuescript
875	673	C	CaMV 35S polyA	Genbank accession V00140
1928	903	C	Hygromycin resistance	Genbank accession V01499.1
2243	1962	C	attP	Genbank accession M23841
3017	2247	C	CaMV-35S promoter	Genbank accession E01311.1
3021	3042		Multiple cloning site	From pBLuescript
3098	3021	C	B-galactosidase a-fragment	From pBLuescript
3220	3099	C	P-Lac	From pBLuescript
3440	4107		pUC origin	From pBLuescript
5115	4258	C	Amp Resistance	From pBLuescript
5116	5243		pBluescript SK+ backbone	

Molecule: pJHD2-19a, 4458 bps DNA Circular

Start	End	Orientation	Name	Accession
1	21		pBluescript SK- backbone	
21	327		f1 (-) origin	From pBLuescript
653	669		Multiple cloning site	From pBLuescript
669	460	C	B-galactosidase a-fragment	From pBLuescript
670	2165		Arabidopsis 26S rDNA gene fragment	Genbank accession X52320.1
2166	2257		Multiple cloning site	From pBLuescript
2313	2167	C	B-galactosidase a-fragment	From pBLuescript
2435	2314	C	P-Lac	From pBLuescript
2655	3322		pUC origin	From pBLuescript
4330	3473	C	Amp Resistance	From pBLuescript
4331	4458		pBluescript backbone	

Molecule: pABI037, 6555 bps DNA Circular

Start	End	Orientation	Name	Accession
95	216		Lac promoter	From pCR-XL-TOPO
205	221		M13 reverse primer	From pCR-XL-TOPO
248	336		Multiple cloning site	From pCR-XL-TOPO
337	3372		Soybean 26S rDNA fragment	Genbank accession AK286061
3372	3435		Multiple cloning site	From pCR-XL-TOPO
3469	3484		M13 forward primer	From pCR-XL-TOPO
3613	3621		Fusion joint	From pCR-XL-TOPO
3622	3924		ccdB lethal gene	From pCR-XL-TOPO
4273	5067		Kanamycin resistance	From pCR-XL-TOPO
5274	5648		Zeocin resistance	From pCR-XL-TOPO
5716	6429		pUC origin	From pCR-XL-TOPO