



(86) Date de dépôt PCT/PCT Filing Date: 2006/09/07
(87) Date publication PCT/PCT Publication Date: 2007/03/15
(45) Date de délivrance/Issue Date: 2014/12/16
(85) Entrée phase nationale/National Entry: 2008/03/07
(86) N° demande PCT/PCT Application No.: US 2006/034669
(87) N° publication PCT/PCT Publication No.: 2007/030510
(30) Priorité/Priority: 2005/09/08 (US60/715,976)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01),
A01H 5/10 (2006.01)

(72) Inventeurs/Inventors:
ZIELER, HELGE, US;
RUDGERS, GARY W., US;
PREUSS, DAPHNE, US;
COPENHAVER, GREGORY P., US;
PAULY, MICHAEL H., US

(73) Propriétaire/Owner:
CHROMATIN INC., US

(74) Agent: SMART & BIGGAR

(54) Titre : PLANTES MODIFIEES PAR DES MINI-CHROMOSOMES
(54) Title: PLANTS MODIFIED WITH MINI-CHROMOSOMES

(57) **Abrégé/Abstract:**

The invention is generally related to methods of generating plants transformed with novel autonomous mini-chromosomes. Mini-chromosomes with novel compositions and structures are used to transform plants cells which are in turn used to generate the plant. Methods for generating the plant include methods for delivering the mini-chromosome into plant cell to transform the cell, methods for selecting the transformed cell, and methods for isolating plants transformed with the mini-chromosome. Plants generated in the present invention contain novel genes introduced into their genome by integration into existing chromosomes.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 March 2007 (15.03.2007)

PCT

(10) International Publication Number
WO 2007/030510 A3

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/10* (2006.01)

(21) International Application Number:
PCT/US2006/034669

(22) International Filing Date:
7 September 2006 (07.09.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/715,976 8 September 2005 (08.09.2005) US

(71) Applicant (for all designated States except US): **CHROMATIN INC.** [US/US]; 2201 W. Campbell Park Drive, Chicago, IL 60612 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ZIELER, Helge** [US/US]; 910 W. VanBuren Street - #194, Chicago, IL 60607 (US). **RUDGERS, Gary, W.** [US/US]; 635 W Buckingham Place, Apt. 3E, Chicago, IL 60657 (US). **PREUSS, Daphne** [US/US]; 5807 S. Dorchester, Apt. #2E, Chicago, IL 60637 (US). **COPENHAVER, Gregory, P.** [US/US]; 112 Windover Drive, Chapel Hill, NC 27514 (US). **PAULY, Michael, H.** [US/US]; 2514 Via Pisa, Del Mar, CA 92014 (US).

(74) Agents: **SINTICH, Sharon, M.** et al.; **MARSHALL, GERSTEIN & BORUN**, 233 S. Wacker Drive, Suite 6300, Sears Tower, Chicago, IL 60606-6357 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
18 May 2007

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PLANTS MODIFIED WITH MINI-CHROMOSOMES

(57) Abstract: The invention is generally related to methods of generating plants transformed with novel autonomous mini-chromosomes. Mini-chromosomes with novel compositions and structures are used to transform plants cells which are in turn used to generate the plant. Methods for generating the plant include methods for delivering the mini-chromosome into plant cell to transform the cell, methods for selecting the transformed cell, and methods for isolating plants transformed with the mini-chromosome. Plants generated in the present invention contain novel genes introduced into their genome by integration into existing chromosomes.



WO 2007/030510 A3

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

PLANTS MODIFIED WITH MINI-CHROMOSOMES

This application claims priority to U.S. Provisional Patent Application No. 60/715,976, filed September 8, 2005.

5

BACKGROUND OF THE INVENTION

Two general approaches are used for introduction of new genetic information ("transformation") into cells. One approach is to introduce the new genetic information as part of another DNA molecule, referred to as an "episomal vector," or "mini-chromosome", which can be maintained as an independent unit (an episome) apart from the host chromosomal DNA molecule(s). Episomal vectors contain all the necessary DNA sequence elements required for DNA replication and maintenance of the vector within the cell. Many episomal vectors are available for use in bacterial cells (for example, see Maniatis et al., "Molecular Cloning: a Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 1982). However, only a few episomal vectors that function in higher eukaryotic cells have been developed. Higher eukaryotic episomal vectors were primarily based on naturally occurring viruses. In higher plant systems gemini viruses are double-stranded DNA viruses that replicate through a double-stranded intermediate upon which an episomal vector could be based, although the gemini virus is limited to an approximately 800 bp insert. Although an episomal plant vector based on the Cauliflower Mosaic Virus has been developed, its capacity to carry new genetic information also is limited (Brisson et al., Nature, 310:511,1984.).

The other general method of genetic transformation involves integration of introduced DNA sequences into the recipient cell's chromosomes, permitting the new information to be replicated and partitioned to the cell's progeny as a part of the natural chromosomes. The introduced DNA usually is broken and joined together in various combinations before it is integrated at random sites into the cell's chromosome (see, for example Wigler *et al.*, Cell, 11:223, 1977). Common problems with this procedure are the rearrangement of introduced DNA sequences and unpredictable levels of expression due to the location of the transgene in the genome or so called "position effect variegation" (Shingo *et al.*, Mol. Cell. Biol., 6:1787, 1986). Further, unlike episomal DNA, integrated DNA cannot normally be precisely

removed. A more refined form of integrative transformation can be achieved by exploiting naturally occurring viruses that integrate into the host's chromosomes as part of their life cycle, such as retroviruses (see Chepko *et al.*, Cell, 37:1053, 1984).

One common genetic transformation method used in higher plants is based on the transfer of bacterial DNA into plant chromosomes that occurs during infection by the phytopathogenic soil bacterium *Agrobacterium* (see Nester *et al.*, Ann. Rev. Plant Phys., 35:387-413, 1984). By substituting genes of interest for the naturally transferred bacterial sequences (called T-DNA), investigators have been able to introduce new DNA into plant cells. However, even this more "refined" integrative transformation system is limited in three major ways. First, DNA sequences introduced into plant cells using the *Agrobacterium* T-DNA system are frequently rearranged (see Jones *et al.*, Mol Gen. Genet., 207:478, 1987). Second, the expression of the introduced DNA sequences varies between individual transformants (see Jones *et al.*, Embo J., 4:2411-2418, 1985). This variability is presumably caused by rearranged sequences and the influence of surrounding sequences in the plant chromosome (*i.e.*, position effects), as well as methylation of the transgene. Finally, insertion of extra elements into the genome can disrupt the genes, promoters or other genetic elements necessary for normal plant growth and function.

Another widely used technique to genetically transform plants involves the use of microprojectile bombardment. In this process, a nucleic acid containing the desired genetic elements to be introduced into the plant is deposited on or in small metallic particles, e.g., tungsten, platinum, or preferably gold, which are then delivered at a high velocity into the plant tissue or plant cells. However, similar problems arise as with *Agrobacterium*-mediated gene transfer, and as noted above expression of the inserted DNA can be unpredictable and insertion of extra elements into the genome can disrupt and adversely impact plant processes.

One attractive alternative to commonly used methods of transformation is the use of an artificial chromosome. Artificial chromosomes are man-made linear or circular DNA molecules constructed in part from cis-acting DNA sequence elements that provide replication and partitioning of the constructed chromosomes (see Murray *et al.*, Nature, 305:189-193, 1983). Desired elements include: (1) origin

of replication, which are the sites for initiation of DNA replication, (2) Centromeres (site of kinetochore assembly and responsible for proper distribution of replicated chromosomes into daughter cells at mitosis or meiosis), and (3) if the chromosome is linear, telomeres (specialized DNA structures at the ends of linear chromosomes that function to stabilize the ends and facilitate the complete replication of the extreme termini of the DNA molecule). An additional desired element is a chromatin organizing sequence. It is well documented that centromere function is crucial for stable chromosomal inheritance in almost all eukaryotic organisms (reviewed in Nicklas 1988). The centromere accomplishes this by attaching, via centromere binding proteins, to the spindle fibers during mitosis and meiosis, thus ensuring proper gene segregation during cell divisions.

The essential chromosomal elements for construction of artificial chromosomes have been precisely characterized in lower eukaryotic species, and more recently in mouse and human. Autonomous replication sequences (ARSs) have been isolated from unicellular fungi, including *Saccharomyces cerevisiae* (brewer's yeast) and *Schizosaccharomyces pombe* (see Stinchcomb *et al.*, 1979 and Hsiao *et al.*, 1979). An ARS behaves like an origin of replication allowing DNA molecules that contain the ARS to be replicated in concert with the rest of the genome after introduction into the cell nuclei of these fungi. DNA molecules containing these sequences replicate, but in the absence of a centromere they are not partitioned into daughter cells in a controlled fashion that ensures efficient chromosome inheritance.

Artificial chromosomes have been constructed in yeast using the three cloned essential chromosomal elements (see Murray *et al.*, Nature, 305:189-193, 1983). None of the essential components identified in unicellular organisms, however, function in higher eukaryotic systems. For example, a yeast CEN sequence will not confer stable inheritance upon vectors transformed into higher eukaryotes.

In contrast to the detailed studies done in yeast, less is known about the molecular structure of functional centromeric DNA of higher eukaryotes. Ultrastructural studies indicate that higher eukaryotic kinetochores, which are specialized complexes of proteins that form on the centromere during late prophase, are large structures (mammalian kinetochore plates are approximately 0.3 μm in diameter) which possess multiple microtubule attachment sites (reviewed in Rieder,

1982). It is therefore possible that the centromeric DNA regions of these organisms will be correspondingly large, although the minimal amount of DNA necessary for centromere function may be much smaller.

While the above studies have been useful in elucidating the structure and function of centromeres, it was not known whether information derived from lower eukaryotic or mammalian higher eukaryotic organisms would be applicable to plants. There exists a need for cloned centromeres from higher eukaryotic organisms, particularly plant organisms, which would represent a first step in production of artificial chromosomes. There further exists a need for plant cells, plants, seeds and progeny containing functional, stable, and autonomous artificial chromosomes capable of carrying a large number of different genes and genetic elements.

SUMMARY OF THE INVENTION

In one aspect, the present invention addresses mini-chromosomes comprising a centromere having one or more selected repeated nucleotide sequences and adchromosomal *Zea mays* (corn) plants comprising a mini-chromosome of the invention. The invention provides for the mini-chromosomes, described in further detail herein, having a centromere comprising a selected repeated nucleotide sequence derived from *Zea mays*.

In another aspect, the invention is based on the production of modified plants, containing functional, stable, autonomous mini-chromosomes. Such mini-chromosomes have been shown herein to be meiotically transmitted to progeny. The present invention particularly addresses adchromosomal *Zea mays* (corn) plants. The invention provides for adchromosomal plants, described in further detail herein, comprising a mini-chromosome, wherein said mini-chromosome preferably has a transmission efficiency during mitotic division of at least 90%, for example, at least 95%. Additionally, these adchromosomal plants may comprise a mini-chromosome having a transmission efficiency during meiotic division of, e.g., at least 80%, at least 85%, at least 90% or at least 95%.

In one embodiment, the mini-chromosomes of the invention comprise a centromere comprising any one of (a) a repeated nucleotide sequence derived from (i.e. is a fragment or variant of) the sequence denoted as CentC, an exemplary sequence of which is provided as GenBank Accession No. AY1290008 (SEQ ID NO:

77), (b) a fragment derived from the sequence denoted as CRM, an exemplary sequence of which is provided as GenBank Accession No. AY129008, or (c) a fragment derived from the sequence denoted as CentA, an exemplary sequence of which is provided as GenBank Accession No. AF078917 (SEQ ID NO: 78), or combinations thereof. Such a sequence or fragment derived from CentC, CRM or CentA preferably hybridizes under highly selective conditions to a representative CentC, CRM or CentA sequence, respectively, or retains at least 70%, 75%, 80%, 85%, 90% or 95% overall identity over the length of the sequence or fragment to a representative CentC, CRM or CentA sequence.

Particularly, the invention provides for mini-chromosomes comprising centromeres having the CentC repeated nucleotide sequence of SEQ ID NO: 70, SEQ ID NO: 71 or variants thereof, e.g. the variants provided in Tables 17 and 22. The invention further provides for mini-chromosomes comprising centromeres having a repeated nucleotide sequence that hybridizes to SEQ ID NO: 70 or SEQ ID NO: 71 under highly selective conditions comprising 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C, or alternatively comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour. The invention also provides for mini-chromosomes comprising a repeated nucleotide sequence that is at least 70%, 75%, 80%, 85%, 90% or 95% identical to SEQ ID NO: 70 or SEQ ID NO: 71. For example, a CentC variant may utilize any nucleotide displayed at a particular base position in Table 17 or 22 together with any nucleotide displayed at any other base position in Table 17 or 22 in any combination, provided that the sequence of the CentC variant retains overall identity over its length of at least 70% to SEQ ID NO: 70 or 71, or would hybridize under highly selective conditions to SEQ ID NO: 70 or 71.

In another embodiment, the invention provides for mini-chromosomes comprising centromeres having a CRM repeated nucleotide sequence that is a fragment of SEQ ID NO: 77 or variant thereof. Such fragments of CRM preferably include at least 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450, or 500 bp of CRM, most preferably at least 50 bp of CRM. The invention further provides for mini-chromosomes comprising centromeres having a variant CRM repeated nucleotide

sequence that hybridizes to SEQ ID NO: 77 under highly selective conditions comprising 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C, or alternatively comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour. Exemplary fragments of CRM include nucleotides 1-515 (515 bp),
 5 nucleotides 1-930 (930 bp), nucleotides 1- 1434 (1434 bp), nucleotides 1508-3791 (2284 bp), nucleotides 1508-5417 (3910 bp), nucleotides 2796-2890 (95 bp), nucleotides 2796-2893 (98 bp), nucleotides 4251- 4744 (494 bp), nucleotides 4626-4772 (147 bp), nucleotides 4945-6236 (1295 bp), nucleotides 4983-5342 (360 bp), nucleotides 5487-5569 (83 bp), nucleotides 5757- 6212 (456 bp), nucleotides 5765-7571 (1807 bp),
 10 nucleotides 6529-6653 (125 bp), nucleotides 6608-6658 (51 bp), nucleotides 6638-7571 (934 bp) and/or nucleotides 6640-7156 (517 bp) of SEQ ID NO: 77. The invention also provides for mini-chromosomes comprising a repeated nucleotide sequence that retains overall identity over its length of at least 70%, 75%, 80%, 85%, 90% or 95% to SEQ ID NO: 77. The invention contemplates fragments of CRM ranging in size up to 51 bp, 83
 15 bp, 95 bp, 98 bp, 125 bp, 147 bp, 360 bp, 456 bp, 494 bp, 515 bp, 517 bp, 930 bp, 934 bp, 1295 bp, 1434 bp, 1807 bp, 2284 bp or 3910 bp in length.

The invention also provides for mini-chromosomes comprising centromeres having a CentA repeated nucleotide sequence that is a fragment of SEQ ID NO: 78 or variant thereof. Exemplary fragments of CentA are up to 512 bp or 513 bp in length (see
 20 Table 15 below) or range in size from 50 to 512 bp or 50 to 513 bp. The invention further provides for mini-chromosomes comprising centromeres having a variant CentA repeated nucleotide sequence that hybridizes to SEQ ID NO: 79 under highly selective conditions comprising 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C, or alternatively comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a
 25 wash at 65°C for a half hour. The invention also provides for mini-chromosomes comprising a repeated nucleotide sequences that retains overall identity over its length of at least 70%, 75%, 80%, 85%, 90% or 95% to SEQ ID NO: 78.

In another embodiment, the centromeres of any of the preceding mini-chromosomes comprise a combination of two or more of the repeated nucleotides
 30 sequences described herein, including those derived from CentC, CRM or CentA

sequences. The invention provides for mini-chromosomes having a centromere comprising (a) a first repeated nucleotide sequence that hybridizes under highly selective conditions comprising 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C, or alternatively comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour, to the nucleotide sequence of either SEQ ID NO: 70 or SEQ ID NO: 71, and (b) a second repeated nucleotide sequence that hybridizes under highly selective conditions comprising 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C, or alternatively comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour, to the nucleotide sequence of SEQ ID NO: 78. Preferably the second repeated nucleotide sequence comprises at least 50 base pairs of SEQ ID NO: 78. Alternatively, the second nucleotide sequence can hybridize under highly selective conditions to the nucleotide sequence of SEQ ID NO: 77. In particular, the invention contemplates mini-chromosomes having a centromere comprising the repeated nucleotide sequence of SEQ ID NO: 70 or a variant thereof and a 50 bp fragment of SEQ ID NO: 77. The invention also contemplates mini-chromosomes having a centromere comprising the repeated nucleotide sequence of SEQ ID NO: 71 or a variant thereof and a 50 bp fragment of SEQ ID NO: 77.

The invention contemplates mini-chromosomes having centromeres comprising at least 50 bp of the contig segments identified in Tables 14 and 18 as homologous to any of the following sequences: Mo 17 locus bz (GenBank Accession No. AY664416), rust resistance gene rp3-1 (GenBank Accession No. AY5704035), coliphage phi-X174 (GenBank Accession No. J02482), 40S ribosomal protein S8 (GenBank Accession No. AY530951), gag-pol (GenBank Accession No. AF464738), retrotransposon (GenBank Accession No. AY574035), Mol7 locus 9008 (GenBank Accession No. AY664418), alpha zein gene cluster (GenBank Accession No. AF090447), Mol7 locus 9009 (GenBank Accession No. AY664419), B73 locus 9002 (GenBank Accession No. AY664413), Magnaporthe grisea (GenBank Accession No. XM_367004), yeast 26S ribosomal RNA (GenBank Accession No. AY046113), Tnl (GenBank Accession No. AF162223), and polynucleotides having the sequence of any of SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID

- 7a -

NO: 85, SEQ ID NO: 86, SEQ ID NO: 87 and SEQ ID NO: 88. The invention also contemplates mini-chromosomes having a centromere comprising a fragment or a variant of any of these nucleotide sequences.

The invention further contemplates that, for any of the contig fragments identified
5 in any of the tables herein by their beginning and ending

nucleotide numbers, isolated nucleic acids may be prepared (including single stranded or double stranded) that retain exact identity to the identified fragment or complement thereof, or that are further fragments or variants thereof that preferably retain ability to hybridize to the original identified fragment. Such isolated nucleic acids are used,
5 *e.g.*, as components of mini-chromosomes of the invention, as probes to isolate centromere sequences for use in mini-chromosomes of the invention, or for transcription of desired complementary strands.

The invention also contemplates mini-chromosomes having a centomere comprising one or more of the following simple repeat sequences: AT-rich repeat,
10 (GCA)_n repeat, GA-rich repeat, CT-rich repeat, T-rich or (TTTTC)_n repeat.

In another embodiment, any of the preceding mini-chromosomes comprise centromeres having *n* copies of a repeated nucleotide sequence, wherein *n* is less than 1000, or less than 500, 250 or 100. In exemplary embodiments, the centromeres of the mini-chromosomes of the invention comprise *n* copies of a
15 repeated nucleotide sequence, wherein *n* is at least 5, wherein *n* is at least 15, wherein *n* is at least 25, wherein *n* is at least 50 and wherein *n* is at least 100.

In additional exemplary embodiments, the centromeres of the mini-chromosomes of the invention comprise *n* copies of a repeated nucleotide sequence where *n* ranges from 5 to 15, 5 to 25, 5 to 50, 5 to 100, 5 to 250, 5 to 500, 5 to
20 1000, 15 to 25, 15 to 50, 15 to 100, 15 to 250, 15 to 500, 15 to 1000, 25 to 50, 25 to 100, 25 to 250, 25 to 500, 25 to 1000, 50 to 100, 50 to 250, 50 to 500, 50 to 1000, 100 to 250, 100 to 500, 100 to 1000, 250 to 500, 250 to 1000, or 500 to 1000.

According to the rough sequence assembly described in Example 6, BAC clones ZB19 has long stretches of CentC repeat and BAC clone ZP113 has long
25 stretches of CentC repeats and/or CRM repeats. For example, the BAC clone ZB19 has stretches of 50 copies of CentC repeats in about 7.5 kb of the nucleotide sequence of contig 30 (SEQ ID NO: 50) and 70 copies of CentC repeats in about 10.5 kb of the nucleotide sequence of contig 31 (SEQ ID NO: 51). The BAC clone ZB113 has stretches of 7 copies of CentC repeats in about 1 kb of the nucleotide sequence of
30 contig 4 (SEQ ID NO: 55), 13 copies of CentC repeats in 1.5 kb of the nucleotide sequence of contig 8 (SEQ ID NO: 59), 24 copies of CentC repeats in about 3.5 kb of the nucleotide sequence of contig 11 (SEQ ID NO: 62), 70 copies of CentC repeats in

about 10.7 kb of the nucleotide sequence of contig 15 (SEQ ID NO: 66), 85 copies of CentC repeats in about 13.5 kb of the nucleotide sequence of contig 17 (SEQ ID NO: 68), and 68 copies of CentC repeats in about 20 kb of the nucleotide sequence of contig 18 (SEQ ID NO: 69). In addition, BAC clone ZB113 has 10 copies of CRM repeats and 20 copies of CentC repeats in about 8.5 kb of the nucleotide sequence of contig 14 (SEQ ID NO: 65). BAC clone ZB113 has 11 copies of CRM repeat and 1 copies of CentA repeat in 15.5 kb of the nucleotide sequence of contig 16 (SEQ ID NO: 67). These are examples of stretches of repeated nucleotide sequence in two functional mini-chromosomes.

10 The invention contemplates mini-chromosomes having a centromere comprising any of the following: at least 5, 6, 7, 8, 9 or 10 repeated nucleotide sequences in about 1.3 kb of nucleotide sequence, at least 15, 20, 25, 30, 35 or 37 repeated nucleotide sequences in about 5.5 kb of nucleotide sequence; or at least 40, 45, 50, 55, 60, 65, 70, 75 or 76 repeated nucleotide sequences in about 13.5 kb of
15 nucleotide sequence.

 In an embodiment of the invention, any of the preceding mini-chromosomes comprising a centromere having at least 5 consecutive repeated nucleotide sequences in head to tail orientation.. The invention also provides for any of the preceding mini-chromosomes comprising a centromere having at least 5
20 repeated nucleotide sequences that are consecutive. Consecutive repeated nucleotide sequences may be in any orientation, e.g. head to tail, tail to tail, or head to head, and need not be directly adjacent to each other (e.g., may be 1-50 bp apart).

 The invention further provides for any of the preceding mini-chromosomes comprising a centromere having at least 5 of the consecutive repeated
25 nucleotide sequences separated by less than n number of nucleotides, wherein n ranges from 1 to 10, or 1 to 20, or 1 to 30, or 1 to 40, or 1 to 50 or wherein n is less than 10 bp or n is less than 20 bp or n is less than 30 bp or n is less than 40 bp or n is less than 50 bp.

 In one embodiment, the mini-chromosomes of the invention are 1000
30 kilobases or less in length. In exemplary embodiments, the mini-chromosome is 600 kilobases or less in length or 500 kilobases or less in length.

In another embodiment, the mini-chromosomes of the invention comprises a site for site-specific recombination.

In another embodiment, the invention provides for the mini-chromosome, further comprising a centromeric nucleic acid insert that comprises
5 artificially synthesized repeated nucleotide sequences. These artificially synthesized repeated nucleotide sequences may be derived from natural centromere sequences, combinations or fragments of natural centromere sequences including a combination of repeats of different lengths, a combination of different sequences, a combination of both different repeat lengths and different sequences, a combination of repeats from
10 two or more plant species, a combination of different artificially synthesized sequences or a combination of natural centromere sequence(s) and artificially synthesized sequence(s).

The invention also provides for a mini-chromosome, wherein the mini-chromosome is derived from a donor clone or a centromere clone and has
15 substitutions, deletions, insertions, duplications or arrangements of one or more nucleotides in the mini-chromosome compared to the nucleotide sequence of the donor clone or centromere clone. In one embodiment, the mini-chromosome is obtained by passage of the mini-chromosome through one or more hosts. In another embodiment, the mini-chromosome is obtained by passage of the mini-chromosome
20 through two or more different hosts. The host may be selected from the group consisting of viruses, bacteria, yeasts, plants, prokaryotic organisms, or eukaryotic organisms. In another embodiment, the mini-chromosome is obtained from a donor clone by *in vitro* methods that introduce sequence variation during template-based replication of the donor clone, or its complementary sequence. In one embodiment
25 this variation may be introduced by a DNA-dependent DNA polymerase. In a further embodiment a minichromosome derived by an *in vitro* method may be further modified by passage of the mini-chromosome through one or more hosts.

The invention also provides for mini-chromosomes that preferably have a transmission efficiency during mitotic division of at least 90%, for example, at
30 least 95%. Additionally, these adchromosomal mini-chromosomes have a transmission efficiency during meiotic division of, e.g., at least 80%, at least 85%, at least 90% or at least 95%.

The invention also provides for a mini-chromosome, wherein the mini-chromosome comprises one or more exogenous nucleic acids. In further exemplary embodiments, the mini-chromosome comprises at least two or more, at least three or more, at least four or more, at least five or more or at least ten or more exogenous
5 nucleic acids.

In one embodiment, at least one exogenous nucleic acid of any of the preceding mini-chromosomes of a plant is operably linked to a heterologous regulatory sequence functional in plant cells. The invention provides for exogenous nucleic acids linked to a plant regulatory sequence. The invention also provides for
10 exogenous nucleic acids linked to a non-plant regulatory sequence, such as an arthropod, viral, bacterial, vertebrate or yeast regulatory sequence. Exemplary regulatory sequences comprise any one of SEQ ID NOS: 1 to 20 or a functional fragment or variant thereof.

In another embodiment, the mini-chromosome comprises an
15 exogenous nucleic acid that confers herbicide resistance, insect resistance, disease resistance, or stress resistance on the plant. The invention provides for mini-chromosomes comprising an exogenous nucleic acid that confers resistance to phosphinothricin or glyphosate herbicide. The invention also provides for mini-chromosomes comprising an exogenous nucleic acid that encodes a phosphinothricin
20 acetyltransferase, glyphosate acetyltransferase, acetohydroxydic synthase or a mutant enolpyruvylshikimate phosphate (EPSP) synthase.

The invention also provides for mini-chromosomes comprising an exogenous nucleic acid that encodes a *Bacillus thuringiensis* toxin gene or *Bacillus cereus* toxin gene. The invention further provides for mini-chromosomes comprising
25 an exogenous nucleic acid that confers resistance to drought, heat, chilling, freezing, excessive moisture, ultraviolet light, ionizing radiation, toxins, pollution, mechanical stress or salt stress. The invention also provides for a mini-chromosome that comprises an exogenous nucleic acid that confers resistance to a virus, bacteria, fungi or nematode.

30 In another embodiment, the mini-chromosome comprises an exogenous nucleic acid conferring herbicide resistance, an exogenous nucleic acid conferring insect resistance, and at least one additional exogenous nucleic acid.

The invention provides for mini-chromosomes comprising an exogenous nucleic acid selected from the group consisting of a nitrogen fixation gene, a plant stress-induced gene, a nutrient utilization gene, a gene that affects plant pigmentation, a gene that encodes an antisense or ribozyme molecule, a gene encoding a secretable antigen, a toxin gene, a receptor gene, a ligand gene, a seed storage gene, a hormone gene, an enzyme gene, an interleukin gene, a clotting factor gene, a cytokine gene, an antibody gene, a growth factor gene, a transcription factor gene, a transcriptional repressor gene, a DNA-binding protein gene, a recombination gene, a DNA replication gene, a programmed cell death gene, a kinase gene, a phosphatase gene, a G protein gene, a cyclin gene, a cell cycle control gene, a gene involved in transcription, a gene involved in translation, a gene involved in RNA processing, a gene involved in RNAi, an organellar gene, a intracellular trafficking gene, an integral membrane protein gene, a transporter gene, a membrane channel protein gene, a cell wall gene, a gene involved in protein processing, a gene involved in protein modification, a gene involved in protein degradation, a gene involved in metabolism, a gene involved in biosynthesis, a gene involved in assimilation of nitrogen or other elements or nutrients, a gene involved in controlling carbon flux, gene involved in respiration, a gene involved in photosynthesis, a gene involved in light sensing, a gene involved in organogenesis, a gene involved in embryogenesis, a gene involved in differentiation, a gene involved in meiotic drive, a gene involved in self incompatibility, a gene involved in development, a gene involved in nutrient, metabolite or mineral transport, a gene involved in nutrient, metabolite or mineral storage, a calcium-binding protein gene, or a lipid-binding protein gene.

The invention also provides for a mini-chromosome comprising an exogenous enzyme gene selected from the group consisting of a gene that encodes an enzyme involved in metabolizing biochemical wastes for use in bioremediation, a gene that encodes an enzyme for modifying pathways that produce secondary plant metabolites, a gene that encodes an enzyme that produces a pharmaceutical, a gene that encodes an enzyme that improves changes the nutritional content of a plant, a gene that encodes an enzyme involved in vitamin synthesis, a gene that encodes an enzyme involved in carbohydrate, polysaccharide or starch synthesis, a gene that encodes an enzyme involved in mineral accumulation or availability, a gene that encodes a phytase, a gene that encodes an enzyme involved in fatty acid, fat or oil

synthesis, a gene that encodes an enzyme involved in synthesis of chemicals or plastics, a gene that encodes an enzyme involved in synthesis of a fuel and a gene that encodes an enzyme involved in synthesis of a fragrance, a gene that encodes an enzyme involved in synthesis of a flavor, a gene that encodes an enzyme involved in synthesis of a pigment or dye, a gene that encodes an enzyme involved in synthesis of a hydrocarbon, a gene that encodes an enzyme involved in synthesis of a structural or fibrous compound, a gene that encodes an enzyme involved in synthesis of a food additive, a gene that encodes an enzyme involved in synthesis of a chemical insecticide, a gene that encodes an enzyme involved in synthesis of an insect repellent, or a gene controlling carbon flux in a plant.

In another embodiment of the invention, any of the preceding mini-chromosomes comprise a telomere.

The invention also provides embodiments wherein any of the preceding mini-chromosomes are linear or circular.

In one embodiment, the invention provides for corn plant cells comprising any of the preceding mini-chromosomes. The invention also provides for corn plant tissue and corn plants comprising these cells. The invention further provides for corn seed obtained from the corn plants of the invention.

In another embodiment, the invention provides for adchromosomal Zea mays (corn) plants comprising any of the preceding mini-chromosomes. In addition, the invention provides for corn plant cells, tissues and seeds obtained from the adchromosomal plants.

In one embodiment of the invention, any of the preceding adchromosomal plants are a monocotyledon. In another embodiment of the invention, any of the preceding adchromosomal plants are a dicotyledone. The invention also provides that the adchromosomal plants of the invention are, e.g., crop plants, cereal plants, vegetable crops, field crops, fruit and vine crops, wood or fiber crops or ornamental plants. The invention also provides exemplary adchromosomal plants that are Zea species.

Another embodiment of the invention is a part of any of the preceding adchromosomal plants. Exemplary plant parts of the invention include a pod, root, cutting, tuber, stem, stalk, fruit, berry, nut, flower, leaf, bark, wood, epidermis,

vascular tissue, organ, protoplast, crown, callus culture, petiole, petal, sepal, stamen, stigma, style, bud, meristem, cambium, cortex, pith, sheath, silk or embryo. Other exemplary plant parts are a meiocyte or gamete or ovule or pollen or endosperm of any of the preceding adchromosomal plants. Other exemplary plant parts are a seed,
5 embryo or propagule of any of the preceding adchromosomal plants.

An embodiment of the invention is a progeny of any of the preceding adchromosomal plants of the invention. These progeny of the invention may be the result of self-breeding, cross-breeding, apomyxis or clonal propagation. In exemplary embodiments, the invention also provides for progeny that comprise a mini-
10 chromosome that is descended from a parental mini-chromosome that contained a centromere less than 150 kilobases in length, less than 100 kilobases in length or less than 50 kilobases in length.

In another aspect, the invention provides for methods of making a mini-chromosome for use in any of the preceding adchromosomal plants of the
15 invention. These methods comprise identifying a centromere nucleotide sequence in a genomic DNA library using a multiplicity of diverse probes, and constructing a mini-chromosome comprising the centromere nucleotide sequence. These methods may further comprise determining hybridization scores for hybridization of the multiplicity of diverse probes to genomic clones within the genomic nucleic acid library,
20 determining a classification for genomic clones within the genomic nucleic acid library according to the hybridization scores for at least two of the diverse probes, and selecting one or more genomic clones within one or more classifications for constructing the mini-chromosome.

In exemplary embodiments, the step of determining a classification for
25 genomic clones within the genomic nucleic acid library may utilize the hybridization scores for at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 or more different probes. A classification may comprise a pattern of high, medium or low hybridization scores to various probes.

Exemplary embodiments of probes useful in this method include a
30 probe that hybridizes to the centromere region of a chromosome, a probe that hybridizes to satellite repeat DNA, a probe that hybridizes to retroelement DNA, a probe that hybridizes to portions of genomic DNA that are heavily methylated, a

probe that hybridizes to arrays of tandem repeats in genomic DNA, a probe that hybridizes to telomere DNA or a probe that hybridizes to a pseudogene. Other exemplary probes include, a probe that hybridizes to ribosomal DNA, a probe that hybridizes to mitochondrial DNA, or a probe that hybridizes to chloroplast DNA, for which preferably a classification comprises a low hybridization score for hybridization to said probe.

Another aspect of the invention provides for methods of making any one of the preceding adchromosomal plants comprising delivering a mini-chromosome to a plant cell using a biolistic method, wherein a particle suitable for use in biolistic method is delivered in a liquid with the mini-chromosome, and regenerating a plant from the plant cell. The liquid may further comprise a divalent ion and a di- or poly-amine. In exemplary embodiments, the liquid comprises water, CaCl₂, and spermidine, and the particles are gold particles. Suitable alternatives to spermidine are, e.g., spermine or other aliphatic or conjugated di- or poly-amines such as 1, 5-diaminopentane, 1, 6-diaminohexane, 1,7-diaminoheptane, 1,8-diaminooctane, histamine or related molecules.

A further aspect of the invention provides for methods of making any of the preceding adchromosomal plant comprising co-delivering to a plant cell a mini-chromosome and a nucleic acid encoding a growth inducing gene, wherein said nucleic acid is not part of the mini-chromosome, and regenerating a plant from the plant cell. The invention further provides for methods comprising co-delivering a nucleic acid encoding a growth inducing gene is not expressed or alternatively is not present in the regenerated plant. The invention also provides for methods wherein the co-delivered nucleic acid encodes a growth inducing gene expressed during regeneration. The growth inducing gene may be a plant growth regulator gene, an organogenesis-promoting gene, an embryogenesis-promoting gene or regeneration-promoting gene, such as *Agrobacterium tumefaciens* isopentenyl transferase gene, *Agrobacterium rhizogenes* isopentenyl transferase gene, *Agrobacterium tumefaciens* indole-3-acetamide hydrolase (IAAH) gene or *Agrobacterium tumefaciens* tryptophan-2-monooxygenase (IAAM) gene.

Another aspect of the invention provides for methods of using any of the preceding adchromosomal plants for a food product, a pharmaceutical product or

chemical product, according to which a suitable exogenous nucleic acid is expressed in adchromosomal plants or plant cells and the plant or plant cells are grown. The plant may secrete the product into its growth environment or the product may be contained within the plant, in which case the plant is harvested and desirable products are extracted.

Thus, the invention contemplates methods of using any of the preceding adchromosomal plants to produce a modified food product, for example, by growing a plant that expresses a exogenous nucleic acid that alters the nutritional content of the plant, and harvesting or processing the corn plant.

The invention also contemplates methods of using any of the preceding adchromosomal plants to produce a recombinant protein, by growing a plant comprising a mini-chromosome that comprises an exogenous nucleic acid encoding the recombinant protein. Optionally the plant is harvested and the desired recombinant protein is isolated from the plant. Exemplary recombinant proteins include pharmaceutical proteins or industrial enzymes.

The invention also contemplates methods of using any of the preceding adchromosomal plants to produce a recombinant protein, by growing a plant comprising a mini-chromosome that comprises an exogenous nucleic acid encoding an enzyme involved in synthesis of the chemical product. Optionally the plant is harvested and the desired chemical product is isolated from the plant. Exemplary chemical products include pharmaceutical products.

Thus, the invention provides in one aspect an isolated mini-chromosome comprising a centromere, wherein the centromere comprises

(a) at least seven first repeated nucleotide sequences that hybridize under conditions comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour to any one of: the nucleotide sequence of SEQ ID NO: 70 or the nucleotide sequence of SEQ ID NO: 71, wherein the repeated nucleotide sequences are within about a 1.3 kb stretch of nucleotide sequence,

- 16a -

(b) at least a second nucleotide sequence comprising a nucleotide sequence that hybridizes under conditions comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour to any one of: the nucleotide sequence of SEQ ID NO: 77 or the nucleotide sequence of SEQ ID NO: 78, and

(c) at least three exogenous nucleic acids,

wherein the centromere confers the ability to segregate to daughter cells.

The invention further provides in one aspect a corn plant cell comprising a minichromosome of the invention.

The invention further provides a corn plant cell comprising a minichromosome of the invention that (i) is not integrated into the plant cell genome and (ii) confers an altered phenotype on the plant cell associated with at least one structural gene within the minichromosome.

The invention further provides a method of using a plant comprising a corn plant cell of the invention, the method comprising the step of growing the plant to produce a recombinant protein encoded by a structural gene of the minichromosome.

SEQUENCES OF THE INVENTION

The following table indicates the identity of the SEQ ID NOs in the sequence listing:

SEQ ID NOS: 1- 6 - *Drosophila melanogaster* promoter sequences

SEQ ID NOS: 7-20 *Saccharomyces cerevisia* promoter sequences

SEQ ID NOS: 21-51 - contigs 1-31 of ZB19

SEQ ID NOS: 52-69 - contigs 1-18 of ZB113

SEQ ID NO 70 - Consensus repeat sequence of CentC from ZB19

SEQ ID NO 71 - Consensus repeat sequence of CentC from ZB113

SEQ ID NO 72 – Consensus repeat sequence of repeat SmOTOT00200215.1 from ZB113

SEQ ID NO 73 – Consensus repeat sequence of repeat SmOTOT00200215.2 from ZB113

- 5 SEQ ID NO 74 – Consensus repeat sequence of repeat SmOTOT00200480 from ZB113
SEQ ID NO 75 – Consensus repeat sequence of repeat SmOTOT00200588 from ZB113
SEQ ID NO: 76 – Full length sequence of CentC (GenBank Accession no. AY321491)
SEQ ID NO: 77 – Full length sequence of CRM (GenBank Accession no. AY129008)
SEQ ID NO: 78 – Full length sequence of CentA (GenBank Accession no. AF078917)
10 SEQ ID NOS: 79-88 - Additional sequences from ZB19 and ZB113

DETAILED DESCRIPTION OF THE INVENTION

While this invention is susceptible of embodiment in many different forms, and will be described herein in detail, specific embodiments thereof with the
15 understanding that the present disclosure is to be considered as an exemplification of the principles of the invention and is not intended to limit the invention to the specific embodiments illustrated.

In another aspect, the invention is based on the production of modified plants, containing functional, stable, autonomous mini-chromosomes. Such mini-
20 chromosomes have been shown herein to be meiotically transmitted to progeny. The present invention particularly addresses adchromosmoal *Zea mays* (corn) plants. The invention provides for adchromosomal plants, described in further detail herein, comprising a mini-chromosome, wherein said mini-chromosome preferably has a transmission efficiency during mitotic division of at least 90%, for example, at least
25 95%. Additionally, these adchromosomal plants may comprise a mini-chromosome having a transmission efficiency during meiotic division of, e.g., at least 80%, at least 85%, at least 90% or at least 95%.

One aspect of the invention is related to plants containing functional, stable, autonomous mini-chromosomes, preferably carrying one or more nucleic acids
30 exogenous to the cell. Such plants carrying mini-chromosomes are contrasted to transgenic plants whose genome has been altered by chromosomal integration of an exogenous nucleic acid. Preferably, expression of the exogenous nucleic acid, either constitutively or in response to a signal which may be a challenge or a stimulus, e.g.

tissue specific expression or time specific expression, results in an altered phenotype of the plant.

The invention provides for mini-chromosomes comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 250, 500, 1000 or more exogenous nucleic acids.

The invention contemplates that any plants, including but not limited to monocots, dicots, gymnosperm, field crops, vegetable crops, fruit and vine crops, or any specific plants named herein, may be modified by carrying autonomous mini-chromosomes as described herein. A related aspect of the invention is plant parts or plant tissues, including pollen, silk, endosperm, ovule, seed, embryo, pods, roots, cuttings, tubers, stems, stalks, fruit, berries, nuts, flowers, leaves, bark, whole plant, plant cell, plant organ, protoplast, cell culture, or any group of plant cells organized into a structural and functional unit, any cells of which carry mini-chromosomes.

A related aspect of the invention is adchromosomal plant parts or plant tissues, including pollen, silk, endosperm, ovule, seed, embryo, pods, roots, cuttings, tubers, stems, stalks, crown, callus culture, petiole, petal, sepal, stamen, stigma, style, bud, fruit, berries, nuts, flowers, leaves, bark, wood, whole plant, plant cell, plant organ, protoplast, cell culture, or any group of plant cells organized into a structural and functional unit. In one preferred embodiment, the exogenous nucleic acid is primarily expressed in a specific location or tissue of a plant, for example, epidermis, vascular tissue, meristem, cambium, cortex, pith, leaf, sheath, flower, root or seed. Tissue-specific expression can be accomplished with, for example, localized presence of the mini-chromosome, selective maintenance of the mini-chromosome, or with promoters that drive tissue-specific expression.

Another related aspect of the invention is meiocytes, pollen, ovules, endosperm, seed, somatic embryos, apomictic embryos, embryos derived from fertilization, vegetative propagules and progeny of the originally adchromosomal plant and of its filial generations that retain the functional, stable, autonomous mini-chromosome. Such progeny include clonally propagated plants, embryos and plant parts as well as filial progeny from self- and cross-breeding, and from apomyxis.

Preferably the mini-chromosome is transmitted to subsequent generations of viable daughter cells during mitotic cell division with a transmission efficiency of at least 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%. More preferably, the mini-chromosome is transmitted to viable gametes during meiotic cell division with a transmission efficiency of at least 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% when more than one copy of the mini-chromosome is present in the gamete mother cells of the plant. Preferably, the mini-chromosome is transmitted to viable gametes during meiotic cell division with a transmission frequency of at least 20%, 30%, 40%, 45%, 46%, 47%, 48%, or 49% when one copy of the mini-chromosome is present in the gamete mother cells of the plant. For production of seeds via sexual reproduction or by apomyxis the mini-chromosome is preferably transferred into at least 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of viable embryos when cells of the plant contain more than one copy of the mini-chromosome. For production of seeds via sexual reproduction or by apomyxis from plants with one mini-chromosome per cell, the mini-chromosome is preferably transferred into at least 20%, 30%, 40%, 45%, 46%, 47%, 48%, or 49% of viable embryos.

Preferably, a mini-chromosome that comprises an exogenous selectable trait or exogenous selectable marker can be employed to increase the frequency in subsequent generations of adchromosomal cells, tissues, gametes, embryos, endosperm, seeds, plants or progeny. More preferably, the frequency of transmission of mini-chromosomes into viable cells, tissues, gametes, embryos, endosperm, seeds, plants or progeny can be at least 95%, 96%, 97%, 98%, 99% or 99.5% after mitosis or meiosis by applying a selection that favors the survival of adchromosomal cells, tissues, gametes, embryos, endosperm, seeds, plants or progeny over such cells, tissues, gametes, embryos, endosperm, seeds, plants or progeny lacking the mini-chromosome.

Transmission efficiency may be measured as the percentage of progeny cells or plants that carry the mini-chromosome as measured by one of several assays taught herein including detection of reporter gene fluorescence, PCR detection of a sequence that is carried by the mini-chromosome, RT-PCR detection of a gene transcript for a gene carried on the mini-chromosome, Western analysis of a protein produced by a gene carried on the mini-chromosome, Southern analysis of the DNA

(either in total or a portion thereof) carried by the mini-chromosome, fluorescence in situ hybridization (FISH) or in situ localization by repressor binding, to name a few. Any assay used to detect the presence of the mini-chromosome (or a portion of the mini-chromosome) may be used to measure the efficiency of a parental cell or plant
5 transmits the mini-chromosome to its progeny. Efficient transmission as measured by some benchmark percentage should indicate the degree to which the mini-chromosome is stable through the mitotic and meiotic cycles.

Plants of the invention may also contain chromosomally integrated exogenous nucleic acid in addition to the autonomous mini-chromosomes. The
10 adchromosomal plants or plant parts, including plant tissues of the invention may include plants that have chromosomal integration of some portion of the mini-chromosome (*e.g.* exogenous nucleic acid or centromere sequences) in some or all cells the plant. The plant, including plant tissue or plant cell is still characterized as adchromosomal despite the occurrence of some chromosomal integration. In one
15 aspect of the invention, the autonomous mini-chromosome can be isolated from integrated exogenous nucleic acid by crossing the adchromosomal plant containing the integrated exogenous nucleic acid with plants producing some gametes lacking the integrated exogenous nucleic acid and subsequently isolating offspring of the cross, or subsequent crosses, that are adchromosomal but lack the integrated exogenous nucleic
20 acid. This independent segregation of the mini-chromosome is one measure of the autonomous nature of the mini-chromosome.

Another aspect of the invention relates to methods for producing and isolating such adchromosomal plants containing functional, stable, autonomous mini-chromosomes.

25 In one embodiment, the invention contemplates improved methods for isolating native centromere sequences. In another embodiment, the invention contemplates methods for generating variants of native or artificial centromere sequences by passage through bacterial or plant or other host cells.

In a further embodiment, the invention contemplates methods for
30 delivering the mini-chromosome into plant cells or tissues to transform the cells or tissues.

In yet another embodiment, the invention contemplates improved methods for regenerating plants, including methods for co-delivery of growth inducing genes with mini-chromosomes. The growth delivery genes include *Agrobacterium tumefaciens* or *Arhizogenes* isopentenyl transferase (IPT) genes involved in cytokinin biosynthesis, plant isopentenyl transferase (IPT) genes involved in cytokinin biosynthesis (from any plant), *Agrobacterium tumefaciens* IAAH, IAAM genes involved in auxin biosynthesis (indole-3-acetamide hydrolase and tryptophan-2-monooxygenase, respectively), *Agrobacterium rhizogenes* rolA, rolB and rolC genes involved in root formation, *Agrobacterium tumefaciens* Aux1, Aux2 genes involved in auxin biosynthesis (indole-3-acetamide hydrolase or tryptophan-2-monooxygenase genes), *Arabidopsis thaliana* leafy cotyledon genes (e.g. Lec1, Lec2) promoting embryogenesis and shoot formation (see Stone et al., Proc. Natl Acad. Sci USA 98: 11806-11811), *Arabidopsis thaliana* ESR1 gene involved in shoot formation (see Banno et al., Plant Cell 13: 2609-2618), *Arabidopsis thaliana* PGA6/WUSCHEL gene involved in embryogenesis (see Zuo et al., Plant J. 30: 349-359).

In yet a further embodiment, the invention contemplates methods for selecting modified plant cells or plant parts containing mini-chromosomes for regeneration. Such methods include assays for identifying adchromosomal plants or cells by determining that mini-chromosomes within the modified plant cell or plant are functional, stable, and autonomous. Exemplary assays for assessing mini-chromosome performance include lineage-based inheritance assays, use of chromosome loss agents to demonstrate autonomy, exonucleas digestion, global mitotic mini-chromosome inheritance assays (sectoring assays) with or without the use of agents inducing chromosomal loss, assays measuring expression levels of marker genes in the mini-chromosome over time and space in a plant, physical assays for separation of autonomous mini-chromosomes from endogenous nuclear chromosomes of plants, molecular assays demonstrating conserved mini-chromosome structure, such as PCR, Southern blots, mini-chromosome rescue, cloning and characterization of mini-chromosome sequences present in the plant, cytological assays detecting mini-chromosome presence in the cell's genome (e.g. FISH) and meiotic mini-chromosome inheritance assays, which measure the levels of mini-chromosome inheritance into a subsequent generation of plants via meiosis and gametes, embryos, endosperm or seeds.

The invention also contemplates novel methods of screening for adchromosomal plant cells that involve use of relatively low, sub-killing concentrations of selection agent (*e.g.* sub-killing antibiotic concentrations), and also involve use of a screenable marker (*e.g.*, a visible marker gene) to identify clusters of modified cells carrying the screenable marker, after which these screenable cells are manipulated to homogeneity. Another aspect of the present invention is related to methods of making and compositions of non-plant promoters for expressing genes in plants.

The invention further provides isolated promoter nucleic acid sequences comprising any one of SEQ ID NOS: 1 to 20, or fragments or variants thereof that retain expression-promoting activity. Mini-chromosomes comprising non-plant promoter sequences such as these that are operably linked to plant-expressed genes (*e.g.*, genes that confer a different phenotype on plants), are contemplated as are plants comprising such mini-chromosomes.

Another aspect is related to methods for using exonuclease to enrich for circular mini-chromosome DNA in genomic DNA preparations.

Another aspect of the invention relates to methods for using such adchromosomal plants containing a mini-chromosome for producing food products, pharmaceutical products and chemical products by appropriate expression of exogenous nucleic acid(s) contained within the mini-chromosome(s).

Mini-chromosomes containing centromeres from one plant species, when inserted into plant cells of a different species or even a different genus or family, can be stable, functional and autonomous. Thus, another aspect of the invention is an adchromosomal plant comprising a functional, stable, autonomous mini-chromosome that contains centromere sequence derived from a different taxonomic plant species, or derived from a different taxonomic plant species, genus, family, order or class.

Yet another aspect of the invention provides novel autonomous mini-chromosomes with novel compositions and structures which are used to transform plant cells which are in turn used to generate a plant (or multiple plants). Exemplary mini-chromosomes of the invention are contemplated to be of a size 2000 kb or less in length. Other exemplary sizes of mini-chromosomes include less than or equal to,

e.g., 1500 kb, 1000 kb, 900 kb, 800 kb, 700 kb, 600 kb, 500 kb, 450 kb, 400 kb, 350 kb, 300 kb, 250 kb, 200 kb, 150 kb, 100 kb, 80 kb, 60 kb, or 40 kb in length.

In a related aspect, novel centromere compositions as characterized by sequence content, size or other parameters are provided. Preferably, the minimal size
5 of centromeric sequence is utilized in mini-chromosome construction. Exemplary sizes include a centromeric nucleic acid insert derived from a portion of plant genomic DNA, that is less than or equal to 1000 kb, 900 kb, 800 kb, 700 kb, 600 kb, 500 kb, 400 kb, 300 kb, 200 kb, 150 kb, 100 kb, 95 kb, 90 kb, 85 kb, 80 kb, 75 kb, 70 kb, 65 kb, 60 kb, 55 kb, 50 kb, 45 kb, 40 kb, 35 kb, 30 kb, 25 kb, 20 kb, 15 kb, 10 kb,
10 5 kb, or 2 kb in length. For example, rescued functional variant soybean centromeric sequences have been shown to be less than 30 kb in size. Another related aspect is the novel structure of the mini-chromosome, particularly structures lacking bacterial sequences, *e.g.*, required for bacterial propagation.

In exemplary embodiments, the invention also contemplates mini-
15 chromosomes or other vectors comprising fragments or variants of the genomic DNA inserts of the BAC clones [identified as ZB19, or ZB113] deposited on February 23, 2005 with the American Type Culture Collection (ATCC), P.O. Box 1549 Manassas, VA 20108, USA, under Accession Nos. PTA-6604 and, PTA-6605, respectively], or naturally occurring descendants thereof, that retain the ability to segregate during
20 mitotic or meiotic division as described herein, as well as adchromosomal plants or parts containing these mini-chromosomes. Other exemplary embodiments include fragments or variants of the genomic DNA inserts of any of the BAC clones identified herein, or descendants thereof, and fragments or variants of the centromeric nucleic acid inserts of any of the vectors or mini-chromosomes identified herein.

25 In other exemplary embodiments, the invention contemplates mini-chromosomes or other vectors comprising centromeric nucleotide sequence that when hybridized to 1, 2, 3, 4, 5, 6, 7, 8 or more of the probes described in the examples herein, under hybridization conditions described herein, *e.g.* low, medium or high stringency, provides relative hybridization scores as described in the examples herein.
30 Exemplary stringent hybridization conditions comprise 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C or 0.5 x SSC 0.25% SDS at 65° for 15 minutes, followed by a wash at 65 degrees for a half hour. Preferably the probes for which

relative hybridization scores are described herein as 5/10 or greater are used, and a hybridization signal greater than background for one or more of these probes is used to select clones. Adchromosomal plants or parts containing such mini-chromosomes are contemplated.

5 The advantages of the present invention include: provision of an autonomous, independent genetic linkage group for accelerating breeding; lack of disruption of host genome; multiple gene “stacking” of large an potentially unlimited numbers of genes; uniform genetic composition exogenous DNA sequences in plant cells and plants containing autonomous mini-chromosomes; defined genetic context
10 for predictable gene expression; higher frequency occurrence and recovery of plant cells and plants containing stably maintained exogenous DNA due to elimination of inefficient integration step; and the ability to eliminate mini-chromosomes in any tissues.

I. Composition of mini-chromosomes and mini-chromosome construction

15 The mini-chromosome vector of the present invention may contain a variety of elements, including (1) sequences that function as plant centromeres, (2) one or more exogenous nucleic acids, including, for example, plant-expressed genes, (3) sequences that function as an origin of replication, which may be included in the region that functions as plant centromere, (4) optionally, a bacterial plasmid backbone
20 for propagation of the plasmid in bacteria, (5) optionally, sequences that function as plant telomeres, (6) optionally, additional “stuffer DNA” sequences that serve to separate the various components on the mini-chromosome from each other, (7) optionally “buffer” sequences such as MARs or SARs, (8) optionally marker sequences of any origin, including but not limited to plant and bacterial origin, (9)
25 optionally, sequences that serve as recombination sites, and (10) “chromatin packaging sequences” such as cohesion and condensing binding sites.

 The mini-chromosomes of the present invention may be constructed to include various components which are novel, which include, but are not limited to, the centromere comprising novel repeating centromeric sequences, and the promoters,
30 particularly promoters derived from non-plant species, as described in further detail below.

The mini-chromosomes of the present invention may be constructed to include various components which are novel, which include, but are not limited to, the centromere comprising novel repeating centromeric sequences, and the promoters, particularly promoters derived from non-plant species, as described in further detail
5 below.

Novel centromere compositions

The centromere in the mini-chromosome of the present invention may comprise novel repeating centromeric sequences.

Exemplary embodiments of centromere nucleic acid sequences
10 according to the present invention include fragments or variants of the genomic DNA inserts of the BAC clones [identified as ZB19, or ZB113 deposited on February 23, 2005 with the American Type Culture Collection (ATCC), P.O. Box 1549 Manassas, VA 20108, USA, under Accession Nos. PTA-6604 and PTA-6605, respectively] that retain the ability to segregate during mitotic or meiotic division as described herein.
15 Variants of such sequences include artificially produced modifications as described herein and modifications produced via passaging through one or more bacterial, plant or other host cells as described herein.

Vectors comprising one, two, three, four, five, six, seven, eight, nine, ten, 15 or 20 or more of the elements contained in any of the exemplary vectors
20 described in the examples below are also contemplated.

The invention specifically contemplates the alternative use of fragments or variants (mutants) of any of the nucleic acids described herein that retain the desired activity, including nucleic acids that function as centromeres, nucleic acids that function as promoters or other regulatory control sequences, or exogenous nucleic
25 acids. Variants may have one or more additions, substitutions or deletions of nucleotides within the original nucleotide sequence or consensus sequence. Variants include nucleic acid sequences that are at least 50%, 55%, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% identical to the original nucleic acid sequence. Variants also include nucleic acid sequences that
30 hybridize under low, medium, high or very high stringency conditions to the original nucleic acid sequence. Similarly, the specification also contemplates the alternative use of fragments or variants of any of the polypeptides described herein.

The comparison of sequences and determination of percent identity between two nucleotide sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) J. Mol. Biol. 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix. Preferably parameters are set so as to maximize the percent identity.

As used herein, the term "hybridizes under low stringency, medium stringency, and high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in Current Protocols in Molecular Biology (1989) John Wiley & Sons, N.Y., 6.3.1-6.3.6. Aqueous and non-aqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.5 x SSC, 0.1% SDS, at least at 50°C; 2) medium stringency hybridization conditions in 6 x SSC at about 45°C, followed by one or more washes in 0.2 x SSC, 0.1% SDS at 55°C; 3) high stringency hybridization conditions in 6 x SSC at about 45°C, followed by one or more washes in 0.2 x SSC, 0.1% SDS at 65°C. Other exemplary highly selective or stringent hybridization conditions comprise 0.02 M to 0.15 M NaCl at temperatures of about 50°C to 70°C or 0.5 x SSC 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65 degrees for a half hour.

Mini-chromosome sequence content and structure

Plant-expressed genes from non-plant sources may be modified to accommodate plant codon usage, to insert preferred motifs near the translation initiation ATG codon, to remove sequences recognized in plants as 5' or 3' splice sites, or to better reflect plant GC/ AT content. Plant genes typically have a GC content of more than 35%, and coding sequences which are rich in A and T nucleotides can be problematic. For example, ATTTA motifs may destabilize mRNA; plant polyadenylation signals such as AATAAA at inappropriate positions within the message may cause premature truncation of transcription; and monocotyledons may recognize AT-rich sequences as splice sites.

Each exogenous nucleic acid or plant-expressed gene may include a promoter, a coding region and a terminator sequence, which may be separated from each other by restriction endonuclease sites or recombination sites or both. Genes may also include introns, which may be present in any number and at any position within the transcribed portion of the gene, including the 5' untranslated sequence, the coding region and the 3' untranslated sequence. Introns may be natural plant introns derived from any plant, or artificial introns based on the splice site consensus that has been defined for plant species. Some intron sequences have been shown to enhance expression in plants. Optionally the exogenous nucleic acid may include a plant transcriptional terminator, non-translated leader sequences derived from viruses that enhance expression, a minimal promoter, or a signal sequence controlling the targeting of gene products to plant compartments or organelles.

The coding regions of the genes can encode any protein, including but not limited to visible marker genes (for example, fluorescent protein genes, other genes conferring a visible phenotype to the plant) or other screenable or selectable marker genes (for example, conferring resistance to antibiotics, herbicides or other toxic compounds or encoding a protein that confers a growth advantage to the cell expressing the protein) or genes which confer some commercial or agronomic value to the adchromosomal plant. Multiple genes can be placed on the same mini-chromosome vector, limited only by the number of restriction endonuclease sites or site-specific recombination sites present in the vector. The genes may be separated from each other by restriction endonuclease sites, homing endonuclease sites, recombination sites or any combinations thereof. Any number of genes can be present.

The mini-chromosome vector may also contain a bacterial plasmid backbone for propagation of the plasmid in bacteria such as *E. coli*, *A. tumefaciens*, or *A. rhizogenes*. The plasmid backbone may be that of a low-copy vector or in other embodiments it may be desirable to use a mid to high level copy backbone. In one embodiment of the invention, this backbone contains the replicon of the F' plasmid of *E. coli*. However, other plasmid replicons, such as the bacteriophage P1 replicon, or other low-copy plasmid systems such as the RK2 replication origin, may also be used. The backbone may include one or several antibiotic-resistance genes conferring resistance to a specific antibiotic to the bacterial cell in which the plasmid is present.

Bacterial antibiotic-resistance genes include but are not limited to kanamycin-, ampicillin-, chloramphenicol-, streptomycin-, spectinomycin-, tetracycline- and gentamycin-resistance genes.

5 The mini-chromosome vector may also contain plant telomeres. An exemplary telomere sequence is TTTAGGG or its complement. Telomeres are specialized DNA structures at the ends of linear chromosomes that function to stabilize the ends and facilitate the complete replication of the extreme termini of the DNA molecule (Richards et. al., Cell.1988 Apr 8;53(1):127-36; Ausubel et al., Current Protocols in Molecular Biology, Wiley & Sons, 1997).

10 Additionally, the mini-chromosome vector may contain "stuffer DNA" sequences that serve to separate the various components on the mini-chromosome (centromere, genes, telomeres) from each other. The stuffer DNA may be of any origin, prokaryotic or eukaryotic, and from any genome or species, plant, animal, microbe or organelle, or may be of synthetic origin. The stuffer DNA can range from
15 100 bp to 10 Mb in length and can be repetitive in sequence, with unit repeats from 10 to 1,000,000 bp. Examples of repetitive sequences that can be used as stuffer DNAs include but are not limited to: rDNA, satellite repeats, retroelements, transposons, pseudogenes, transcribed genes, microsatellites, tDNA genes, short sequence repeats and combinations thereof. Alternatively, the stuffer DNA can consist of unique, non-
20 repetitive DNA of any origin or sequence. The stuffer sequences may also include DNA with the ability to form boundary domains, such as but not limited to scaffold attachment regions (SARs) or matrix attachment regions (MARs). The stuffer DNA may be entirely synthetic, composed of random sequence. In this case, the stuffer DNA may have any base composition, or any A/T or G/C content. For example, the
25 G/C content of the stuffer DNA could resemble that of the plant (~30-40%), or could be much lower (0-30%) or much higher (40-100%). Alternatively, the stuffer sequences could be synthesized to contain an excess of any given nucleotide such as A, C, G or T. Different synthetic stuffers of different compositions may also be combined with each other. For example a fragment with low G/C content may be
30 flanked or abutted by a fragment of medium or high G/C content, or vice versa.

In one embodiment of the invention, the mini-chromosome has a circular structure without telomeres. In another embodiment, the mini-chromosome

has a circular structure with telomeres. In a third embodiment, the mini-chromosome has a linear structure with telomeres, as would result if a "linear" structure were to be cut with a unique endonuclease, exposing the telomeres at the ends of a DNA molecule that contains all of the sequence contained in the original, closed construct with the exception of the an antibiotic-resistance gene. In a fourth embodiment of the invention, the telomeres could be placed in such a manner that the bacterial replicon, backbone sequences, antibiotic-resistance genes and any other sequences of bacterial origin and present for the purposes of propagation of the mini-chromosome in bacteria, can be removed from the plant-expressed genes, the centromere, telomeres, and other sequences by cutting the structure with an unique endonuclease. This results in a mini-chromosome from which much of, or preferably all, bacterial sequences have been removed. In this embodiment, bacterial sequence present between or among the plant-expressed genes or other mini-chromosome sequences would be excised prior to removal of the remaining bacterial sequences by cutting the mini-chromosome with a homing endonuclease and re-ligating the structure such that the antibiotic-resistance gene has been lost. The unique endonuclease site may be the recognition sequence of a homing endonuclease. Alternatively, the endonucleases and their sites can be replaced with any specific DNA cutting mechanism and its specific recognition site such as rare-cutting endonuclease or recombinase and its specific recognition site, as long as that site is present in the mini-chromosomes only at the indicated positions.

Various structural configurations are possible by which mini-chromosome elements can be oriented with respect to each other. A centromere can be placed on a mini-chromosome either between genes or outside a cluster of genes next to one telomere or next to the other telomere. Stuffer DNAs can be combined with these configurations to place the stuffer sequences inside the telomeres, around the centromere between genes or any combination thereof. Thus, a large number of alternative mini-chromosome structures are possible, depending on the relative placement of centromere DNA, genes, stuffer DNAs, bacterial sequences, telomeres, and other sequences. The sequence content of each of these variants is the same, but their structure may be different depending on how the sequences are placed. These variations in architecture are possible both for linear and for circular mini-chromosomes.

Exemplary centromere components

Centromere components may be isolated or derived from native plant genome, for example, modified through recombinant techniques or through the cell-based techniques described below. Alternatively, wholly artificial centromere components may be constructed using as a general guide the sequence of native centromeres. Combinations of centromere components derived from natural sources and/or combinations of naturally derived and artificial components are also contemplated. As noted above, centromere sequence from one taxonomic plant species has been shown to be functional in another taxonomic plant species, genus and family.

In one embodiment, the centromere contains n copies of a repeated nucleotide sequence obtained by the methods disclosed herein; wherein n is at least 2. In another embodiment, the centromere contains n copies of interdigitated repeats. An interdigitated repeat is a DNA sequence that consists of two distinct repetitive elements that combine to create a unique permutation. Potentially any number of repeat copies capable of physically being placed on the recombinant construct could be included on the construct, including about 5, 10, 15, 20, 30, 50, 75, 100, 150, 200, 300, 400, 500, 750, 1,000, 1,500, 2,000, 3,000, 5,000, 7,500, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000 and about 100,000, including all ranges in-between such copy numbers. Moreover, the copies, while largely identical, can vary from each other. Such repeat variation is commonly observed in naturally occurring centromeres. The length of the repeat may vary, but will preferably range from about 20 bp to about 360 bp, from about 20 bp to about 250 bp, from about 50 bp to about 225 bp, from about 75 bp to about 210 bp, such as a 92 bp repeat and a 97 bp repeat, from about 100 bp to about 205 bp, from about 125 bp to about 200 bp, from about 150 bp to about 195 bp, from about 160 bp to about 190 and from about 170 bp to about 185 bp including about 180 bp.

The invention contemplates that two or more of these repeated nucleotide sequences, or similar repeated nucleotide sequences, may be oriented head to tail within the centromere. The term "head to tail" refers to multiple consecutive copies of the same or similar repeated nucleotide sequence (e.g., at least 70% identical) that are in the same 5'-3' orientation. The invention also contemplates that two or more of these repeated nucleotide sequences may be consecutive within the

centromere. The term "consecutive" refers to the same or similar repeated nucleotide sequences (e.g., at least 70% identical) that follow one after another without being interrupted by other significant sequence elements. Such consecutive repeated nucleotide sequences may be in any orientation, *e.g.* head to tail, tail to tail, or head to head, and
 5 may be separated by *n* number of nucleotides, wherein *n* ranges from 1 to 10, or 1 to 20, or 1 to 30, or 1 to 40, or 1 to 50. Exemplary repeated nucleotide sequences derived from corn, and identified by the methods described herein, are CentC, CRM and CentA. An exemplary sequence of CentC is provided as GenBank Accession No. AY1290008 (SEQ ID NO: 77). The consensus sequence of CentC derived from BAC clone ZB19 is set out
 10 as SEQ ID NO: 70, and the consensus sequence of CentC derived from BAC clone ZB113 is set out as SEQ ID NO: 71. Variants of these CentC consensus sequences within the BAC clones were identified and are set out in Tables 17 and 22.

An exemplary sequence of CRM is provided as GenBank Accession No. AY129008 (SEQ ID NO: 77). The fragments of SEQ ID NO: 77 that are observed within
 15 the BAC clone ZB113 are as follows: nucleotides 1-515, nucleotides 1-930, nucleotides 1-1434, nucleotides 1508-3791, nucleotides 1508-5417, nucleotides 2796-2890, nucleotides 2796-2893, nucleotides 4251-4744, nucleotides 4626-4772, nucleotides 4945-6236, nucleotides 4983-5342, nucleotides 5487-5569, nucleotides 5757-6212, nucleotides 5765-7571, nucleotides 6529-6653, nucleotides 6608-6658, nucleotides 6638-7571
 20 and/or nucleotides 6640-7156 of SEQ ID NO: 78.

An exemplary sequence of CentA is provided as GenBank Accession No. AF078917 (SEQ ID NO: 78). The fragment of SEQ ID NO: 78 that are observed in the BAC clone ZB113 are as follows comprise nucleotides 9589- 10101 of SEQ ID NO: 37. (contig 16).

25

Modification of centromeres isolated from native plant genome

Modification and changes may be made in the centromeric DNA segments of the current invention and still obtain a functional molecule with desirable characteristics. The

- 31a -

following is a discussion based upon changing the nucleic acids of a centromere to create an equivalent, or even an improved, second generation molecule.

In particular embodiments of the invention, mutated centromeric sequences are contemplated to be useful for increasing the utility of the centromere. It is specifically contemplated that the function of the centromeres of the current invention may be based in part or in whole upon the secondary structure of the DNA sequences of the centromere, modification of the DNA with methyl groups or other adducts, and / or the proteins which interact with the centromere. By changing the DNA sequence of the centromere, one may alter the affinity of one or more centromere-associated protein(s) for the centromere and / or the secondary structure or modification of the centromeric sequences, thereby changing the activity of the centromere. Alternatively, changes may be made in the centromeres of the invention which do not affect the activity of the centromere. Changes in the centromeric sequences which reduce the size of the DNA segment needed to confer centromere activity are contemplated to be particularly useful in the current invention, as would changes which increased the fidelity with which the centromere was transmitted during mitosis and meiosis.

Modification of centromeres by passage through bacteria, plant or other hosts or processes

In the methods of the present invention, the resulting mini-chromosome DNA sequence may also be a derivative of the parental clone or centromere clone having substitutions, deletions, insertions, duplications and/or rearrangements of one or more nucleotides in the nucleic acid sequence. Such nucleotide mutations may occur individually or consecutively in stretches of 1, 2, 3, 4, 5, 10, 20, 40, 80, 100, 200, 400, 800, 1000, 2000, 4000, 8000, 10000, 50000, 100000, and about 200000, including all ranges in-between.

Variations of mini-chromosomes may arise through passage of mini-chromosomes through various hosts including virus, bacteria, yeast, plant or other prokaryotic or eukaryotic organism and may occur through passage of multiple hosts or individual host. Variations may also occur by replicating the mini-chromosome in vitro.

Derivatives may be identified through sequence analysis, or variations in mini-chromosome molecular weight through electrophoresis such as, but not limited to, CHEF gel analysis, column or gradient separation, or any other methods used in the field to determine and/or analyze DNA molecular weight or sequence

content. Alternately, derivatives may be identified by the altered activity of a derivative in conferring centromere function to a mini-chromosome.

Exemplary exogenous nucleic acids including plant-expressed genes

Of particular interest in the present invention are exogenous nucleic acids which when introduced into plants will alter the phenotype of the plant, a plant organ, plant tissue, or portion of the plant. Exemplary exogenous nucleic acids encode polypeptides involved in one or more important biological properties in plants. Other exemplary exogenous nucleic acids alter expression of exogenous or endogenous genes, either increasing or decreasing expression, optionally in response to a specific signal or stimulus.

As used herein, the term "trait" can refer either to the altered phenotype of interest or the nucleic acid which causes the altered phenotype of interest.

One of the major purposes of transformation of crop plants is to add some commercially desirable, agronomically important traits to the plant. Such traits include, but are not limited to, herbicide resistance or tolerance; insect (pest) resistance or tolerance; disease resistance or tolerance (viral, bacterial, fungal, nematode or other pathogens); stress tolerance and/or resistance, as exemplified by resistance or tolerance to drought, heat, chilling, freezing, excessive moisture, salt stress, mechanical stress, extreme acidity, alkalinity, toxins, UV light, ionizing radiation or oxidative stress; increased yields, whether in quantity or quality; enhanced or altered nutrient acquisition and enhanced or altered metabolic efficiency; enhanced or altered nutritional content and makeup of plant tissues used for food, feed, fiber or processing; physical appearance; male sterility; drydown; standability; prolificacy; starch quantity and quality; oil quantity and quality; protein quality and quantity; amino acid composition; modified chemical production; altered pharmaceutical or nutraceutical properties; altered bioremediation properties; increased biomass; altered growth rate; altered fitness; altered biodegradability; altered CO₂ fixation; presence of bioindicator activity; altered digestibility by humans or animals; altered allergenicity; altered mating characteristics; altered pollen dispersal; improved environmental impact; altered nitrogen fixation capability; the production of a pharmaceutically active protein; the production of a small molecule with medicinal properties; the production of a chemical including those with

industrial utility; the production of nutraceuticals, food additives, carbohydrates, RNAs, lipids, fuels, dyes, pigments, vitamins, scents, flavors, vaccines, antibodies, hormones, and the like; and alterations in plant architecture or development, including changes in developmental timing, photosynthesis, signal transduction, cell growth, reproduction, or differentiation. Additionally one could create a library of an entire genome from any organism or organelle including mammals, plants, microbes, fungi, or bacteria, represented on mini-chromosomes.

In one embodiment, the modified plant may exhibit increased or decreased expression or accumulation of a product of the plant, which may be a natural product of the plant or a new or altered product of the plant. Exemplary products include an enzyme, an RNA molecule, a nutritional protein, a structural protein, an amino acid, a lipid, a fatty acid, a polysaccharide, a sugar, an alcohol, an alkaloid, a carotenoid, a propanoid, a phenylpropanoid, or terpenoid, a steroid, a flavonoid, a phenolic compound, an anthocyanin, a pigment, a vitamin or a plant hormone. In another embodiment, the modified plant has enhanced or diminished requirements for light, water, nitrogen, or trace elements. In another embodiment the modified plant has an enhance ability to capture or fix nitrogen from its environment. In yet another embodiment, the modified plant is enriched for an essential amino acid as a proportion of a protein fraction of the plant. The protein fraction may be, for example, total seed protein, soluble protein, insoluble protein, water-extractable protein, and lipid-associated protein. The modification may include overexpression, underexpression, antisense modulation, sense suppression, inducible expression, inducible repression, or inducible modulation of a gene.

A brief summary of exemplary improved properties and polypeptides of interest for either increased or decreased expression is provided below.

(i) Herbicide Resistance

A herbicide resistance (or tolerance) trait is a characteristic of a modified plant that is resistant to dosages of an herbicide that is typically lethal to a non-modified plant. Exemplary herbicides for which resistance is useful in a plant include glyphosate herbicides, phosphinothricin herbicides, oxynil herbicides, imidazolinone herbicides, dinitroaniline herbicides, pyridine herbicides, sulfonylurea herbicides, bialaphos herbicides, sulfonamide herbicides and glufosinate herbicides.

Other herbicides would be useful as would combinations of herbicide genes on the same mini-chromosome.

The genes encoding phosphinothricin acetyltransferase (bar), glyphosate tolerant EPSP synthase genes, glyphosate acetyltransferase, the glyphosate
5 degradative enzyme gene gox encoding glyphosate oxidoreductase, deh (encoding a dehalogenase enzyme that inactivates dalapon), herbicide resistant (e.g., sulfonylurea and imidazolinone) acetolactate synthase, and bxn genes (encoding a nitrilase enzyme that degrades bromoxynil) are good examples of herbicide resistant genes for use in transformation. The bar gene codes for an enzyme, phosphinothricin acetyltransferase
10 (PAT), which inactivates the herbicide phosphinothricin and prevents this compound from inhibiting glutamine synthetase enzymes. The enzyme 5-enolpyruvylshikimate 3-phosphate synthase (EPSP Synthase), is normally inhibited by the herbicide N-(phosphonomethyl)glycine (glyphosate). However, genes are known that encode glyphosate resistant EPSP synthase enzymes. These genes are particularly
15 contemplated for use in plant transformation. The deh gene encodes the enzyme dalapon dehalogenase and confers resistance to the herbicide dalapon. The bxn gene codes for a specific nitrilase enzyme that converts bromoxynil to a non herbicidal degradation product. The glyphosate acetyl transferase gene inactivates the herbicide glyphosate and prevents this compound from inhibiting EPSP synthase.

20 Polypeptides that may produce plants having tolerance to plant herbicides include polypeptides involved in the shikimate pathway, which are of interest for providing glyphosate tolerant plants. Such polypeptides include polypeptides involved in biosynthesis of chorismate, phenylalanine, tyrosine and tryptophan.

25 (ii) *Insect Resistance*

Potential insect resistance (or tolerance) genes that can be introduced include *Bacillus thuringiensis* toxin genes or Bt genes (Watrud et al., In: Engineered Organisms and the Environment, 1985). Bt genes may provide resistance to lepidopteran or coleopteran pests such as European Corn Borer (ECB). Preferred Bt
30 toxin genes for use in such embodiments include the CryIA(b) and CryIA(c) genes. Endotoxin genes from other species of *B. thuringiensis* which affect insect growth or development also may be employed in this regard.

It is contemplated that preferred Bt genes for use in the mini- chromosomes disclosed herein will be those in which the coding sequence has been modified to effect increased expression in plants, and for example, in monocot plants. Means for preparing synthetic genes are well known in the art and are disclosed in, for example, U.S. Patent No. 5,500,365 and U.S. Patent Number No. 5,689,052. Examples of such modified Bt toxin genes include a synthetic Bt CryIA(b) gene (Perlak et al., Proc. Natl. Acad. Sci. USA, 88:3324-3328, 1991), and the synthetic CryIA(c) gene termed 1800b (PCT Application WO 95/06128). Some examples of other Bt toxin genes known to those of skill in the art are given in Table 1 below.

Table 1 : *Bacillus thuringiensis* Endotoxin Genes^a

New Nomenclature	Old Nomenclature	GenBank Accession
CryIAa	CryIA(a)	M11250
CryIAb	CryIA(b)	M13898
CryIAc	CryIA(c)	M11068
CryIAd	CryIA(d)	M73250
CryIAe	CryIA(e)	M65252
CryIBa	CryIB	X06711
CryIBb	ET5	L32020
CryIBc	PEG5	Z46442
CryIBd	CryEI	U70726
CryICa	CryIC	X07518
CryICb	CryIC(b)	M97880
CryIDa	CryID	X54160
CryIDb	PrtB	Z22511
CryIEa	CryIE	X53985
CryIEb	CryIE(b)	M73253
CryIFa	CryIF	M63897
CryIFb	PrtD	Z22512
CryIGa	PrtA	Z22510
CryIGb	CryH2	U70725
CryIHa	PrtC	Z22513

Cry1Hb		U35780
Cry1Ia	CryV	X62821
Cry1Ib	CryV	U07642
Cry1Ja	ET4	L32019
Cry1Jb	ET1	U31527
Cry1K		U28801
Cry2Aa	CryIIA	M31738
Cry2Ab	CryIIB	M23724
Cry2Ac	CryIIC	X57252
Cry3A	CryIIIA	M22472
Cry3Ba	CryIIIB	X17123
Cry3Bb	CryIIIB2	M89794
Cry3C	CryIIID	X59797
Cry4A	CryIVA	Y00423
Cry4B	CryIVB	X07423
Cry5Aa	CryVA(a)	L07025
Cry5Ab	CryVA(b)	L07026
Cry6A	CryVIA	L07022
Cry6B	CryVIB	L07024
Cry7Aa	CryIIIC	M64478
Cry7Ab	CryIIICb	U04367
Cry8A	CryIIIE	U04364
Cry8B	CryIIIG	U04365
Cry8C	CryIIIF	U04366
Cry9A	CryIG	X58120
Cry9B	CryIX	X75019
Cry9C	CryIIC	Z37527
Cry10A	CryIVC	M12662
Cry11A	CryIVD	M31737
Cry11B	Jeg80	X86902
Cry12A	CryVB	L07027
Cry13A	CryVC	L07023
Cry14A	CryVD	U13955
Cry15A	34kDa	M76442

Cry16A	cbm71	X94146
Cry17A	cbm71	X99478
Cry18A	CryBP1	X99049
Cry19A	Jeg65	Y08920
Cyt1Aa	CytA	X03182
Cyt1Ab	CytM	X98793
Cyt2A	CytB	Z14147
Cyt2B	CytB	U52043

^a**Adapted from:**

http://epunix.biols.susx.ac.uk/Home/Neil_Crickmore/Bt/index.html

Protease inhibitors also may provide insect resistance (Johnson et al., Proc Natl Acad Sci U S A. 1989 December; 86(24): 9871–9875.), and will thus have utility in plant transformation. The use of a protease inhibitor II gene, pinII, from tomato or potato is envisioned to be particularly useful. Even more advantageous is the use of a pinII gene in combination with a Bt toxin gene, the combined effect of which has been discovered to produce synergistic insecticidal activity. Other genes which encode inhibitors of the insect's digestive system, or those that encode enzymes or co factors that facilitate the production of inhibitors, also may be useful. This group may be exemplified by oryzacystatin and amylase inhibitors such as those from wheat and barley.

Amylase inhibitors are found in various plant species and are used to ward off insect predation via inhibition of the digestive amylases of attacking insects. Several amylase inhibitor genes have been isolated from plants and some have been introduced as exogenous nucleic acids, conferring an insect resistant phenotype that is potentially useful ("Plants, Genes, and Crop Biotechnology" by Maarten J. Chrispeels and David E. Sadava (2003) Jones and Bartlett Press).

Genes encoding lectins may confer additional or alternative insecticide properties. Lectins are multivalent carbohydrate binding proteins which have the ability to agglutinate red blood cells from a range of species. Lectins have been identified recently as insecticidal agents with activity against weevils, ECB and rootworm (Murdock et al., Phytochemistry, 29:85-89, 1990, Czapla & Lang, J. Econ. Entomol., 83:2480-2485, 1990). Lectin genes contemplated to be useful include, for example, barley and wheat germ agglutinin (WGA) and rice lectins (Gatehouse et al., J. Sci. Food. Agric., 35:373-380, 1984), with WGA being preferred.

Genes controlling the production of large or small polypeptides active against insects when introduced into the insect pests, such as, e.g., lytic peptides, peptide hormones and toxins and venoms, form another aspect of the invention. For example, it is contemplated that the expression of juvenile hormone esterase, directed
5 towards specific insect pests, also may result in insecticidal activity, or perhaps cause cessation of metamorphosis (Hammock et al., Nature, 344:458-461, 1990).

Genes which encode enzymes that affect the integrity of the insect cuticle form yet another aspect of the invention. Such genes include those encoding, e.g., chitinase, proteases, lipases and also genes for the production of nikkomycin, a
10 compound that inhibits chitin synthesis, the introduction of any of which is contemplated to produce insect resistant plants. Genes that code for activities that affect insect molting, such as those affecting the production of ecdysteroid UDP glucosyl transferase, also fall within the scope of the useful exogenous nucleic acids of the present invention.

15 Genes that code for enzymes that facilitate the production of compounds that reduce the nutritional quality of the host plant to insect pests also are encompassed by the present invention. It may be possible, for instance, to confer insecticidal activity on a plant by altering its sterol composition. Sterols are obtained by insects from their diet and are used for hormone synthesis and membrane stability.
20 Therefore alterations in plant sterol composition by expression of novel genes, e.g., those that directly promote the production of undesirable sterols or those that convert desirable sterols into undesirable forms, could have a negative effect on insect growth and/or development and hence endow the plant with insecticidal activity. Lipoxygenases are naturally occurring plant enzymes that have been shown to exhibit
25 anti nutritional effects on insects and to reduce the nutritional quality of their diet. Therefore, further embodiments of the invention concern modified plants with enhanced lipoxygenase activity which may be resistant to insect feeding.

Tripsacum dactyloides is a species of grass that is resistant to certain insects, including corn root worm. It is anticipated that genes encoding proteins that
30 are toxic to insects or are involved in the biosynthesis of compounds toxic to insects will be isolated from *Tripsacum* and that these novel genes will be useful in conferring resistance to insects. It is known that the basis of insect resistance in

Tripsacum is genetic, because said resistance has been transferred to *Zea mays* via sexual crosses (Branson and Guss, Proceedings North Central Branch Entomological Society of America, 27:91-95, 1972). It is further anticipated that other cereal, monocot or dicot plant species may have genes encoding proteins that are toxic to insects which would be useful for producing insect resistant plants.

Further genes encoding proteins characterized as having potential insecticidal activity also may be used as exogenous nucleic acids in accordance herewith. Such genes include, for example, the cowpea trypsin inhibitor (CpTI; Hilder et al., Nature, 330:160-163, 1987) which may be used as a rootworm deterrent; genes encoding avermectin (Avermectin and Abamectin., Campbell, W.C., Ed., 1989; Ikeda et al., J. Bacteriol., 169:5615-5621, 1987) which may prove particularly useful as a corn rootworm deterrent; ribosome inactivating protein genes; and even genes that regulate plant structures. Modified plants including anti insect antibody genes and genes that code for enzymes that can convert a non toxic insecticide (pro insecticide) applied to the outside of the plant into an insecticide inside the plant also are contemplated.

Polypeptides that may improve plant tolerance to the effects of plant pests or pathogens include proteases, polypeptides involved in anthocyanin biosynthesis, polypeptides involved in cell wall metabolism, including cellulases, glucosidases, pectin methylesterase, pectinase, polygalacturonase, chitinase, chitosanase, and cellulose synthase, and polypeptides involved in biosynthesis of terpenoids or indole for production of bioactive metabolites to provide defense against herbivorous insects. It is also anticipated that combinations of different insect resistance genes on the same mini-chromosome will be particularly useful.

Vegetative Insecticidal Proteins (VIP) are a relatively new class of proteins originally found to be produced in the vegetative growth phase of the bacterium, *Bacillus cereus*, but do have a spectrum of insect lethality similar to the insecticidal genes found in strains of *Bacillus thuriengensis*. Both the vip1a and vip3A genes have been isolated and have demonstrated insect toxicity. It is anticipated that such genes may be used in modified plants to confer insect resistance ("Plants, Genes, and Crop Biotechnology" by Maarten J. Chrispeels and David E. Sadava (2003) Jones and Bartlett Press).

(iii) *Environment or Stress Resistance*

Improvement of a plant's ability to tolerate various environmental stresses such as, but not limited to, drought, excess moisture, chilling, freezing, high temperature, salt, and oxidative stress, also can be effected through expression of novel genes. It is proposed that benefits may be realized in terms of increased resistance to freezing temperatures through the introduction of an "antifreeze" protein such as that of the Winter Flounder (Cutler et al., J. Plant Physiol., 135:351-354, 1989) or synthetic gene derivatives thereof. Improved chilling tolerance also may be conferred through increased expression of glycerol 3 phosphate acetyltransferase in chloroplasts (Wolter et al., The EMBO J., 4685-4692, 1992). Resistance to oxidative stress (often exacerbated by conditions such as chilling temperatures in combination with high light intensities) can be conferred by expression of superoxide dismutase (Gupta et al., 1993), and may be improved by glutathione reductase (Bowler et al., Ann Rev. Plant Physiol., 43:83-116, 1992). Such strategies may allow for tolerance to freezing in newly emerged fields as well as extending later maturity higher yielding varieties to earlier relative maturity zones.

It is contemplated that the expression of novel genes that favorably affect plant water content, total water potential, osmotic potential, or turgor will enhance the ability of the plant to tolerate drought. As used herein, the terms "drought resistance" and "drought tolerance" are used to refer to a plant's increased resistance or tolerance to stress induced by a reduction in water availability, as compared to normal circumstances, and the ability of the plant to function and survive in lower water environments. In this aspect of the invention it is proposed, for example, that the expression of genes encoding for the biosynthesis of osmotically active solutes, such as polyol compounds, may impart protection against drought. Within this class are genes encoding for mannitol L phosphate dehydrogenase (Lee and Saier, 1982) and trehalose 6 phosphate synthase (Kaasen et al., J. Bacteriology, 174:889-898, 1992). Through the subsequent action of native phosphatases in the cell or by the introduction and coexpression of a specific phosphatase, these introduced genes will result in the accumulation of either mannitol or trehalose, respectively, both of which have been well documented as protective compounds able to mitigate the effects of stress. Mannitol accumulation in transgenic tobacco has been verified and preliminary results indicate that plants expressing high levels of this metabolite

are able to tolerate an applied osmotic stress (Tarczynski et al., Science, 259:508-510, 1993, Tarczynski et al Proc. Natl. Acad. Sci. USA, 89:1-5, 1993).

Similarly, the efficacy of other metabolites in protecting either enzyme function (e.g., alanopine or propionic acid) or membrane integrity (e.g., alanopine) has been documented (Loomis et al., J. Expt. Zoology, 252:9-15, 1989), and therefore expression of genes encoding for the biosynthesis of these compounds might confer drought resistance in a manner similar to or complimentary to mannitol. Other examples of naturally occurring metabolites that are osmotically active and/or provide some direct protective effect during drought and/or desiccation include fructose, erythritol (Coxson et al., Biotropica, 24:121-133, 1992), sorbitol, dulcitol (Karsten et al., Botanica Marina, 35:11-19, 1992), glucosylglycerol (Reed et al., J. Gen. Microbiology, 130:1-4, 1984; Erdmann et al., J. Gen. Microbiology, 138:363-368, 1992), sucrose, stachyose (Koster and Leopold, Plant Physiol., 88:829-832, 1988; Blackman et al., Plant Physiol., 100:225-230, 1992), raffinose (Bernal Lugo and Leopold, Plant Physiol., 98:1207-1210, 1992), proline (Rensburg et al., J. Plant Physiol., 141:188-194, 1993), glycine betaine, ononitol and pinitol (Vernon and Bohnert, The EMBO J., 11:2077-2085, 1992). Continued canopy growth and increased reproductive fitness during times of stress will be augmented by introduction and expression of genes such as those controlling the osmotically active compounds discussed above and other such compounds. Currently preferred genes which promote the synthesis of an osmotically active polyol compound are genes which encode the enzymes mannitol 1 phosphate dehydrogenase, trehalose 6 phosphate synthase and myoinositol 0 methyltransferase.

It is contemplated that the expression of specific proteins also may increase drought tolerance. Three classes of Late Embryogenic Abundant (LEA) Proteins have been assigned based on structural similarities (see Dure et al., Plant Molecular Biology, 12:475-486, 1989). All three classes of LEAs have been demonstrated in maturing (e.g. desiccating) seeds. Within these 3 types of LEA proteins, the Type II (dehydrin type) have generally been implicated in drought and/or desiccation tolerance in vegetative plant parts (e.g. Mundy and Chua, The EMBO J., 7:2279-2286, 1988; Piatkowski et al., Plant Physiol., 94:1682-1688, 1990; Yamaguchi Shinozaki et al., Plant Cell Physiol., 33:217-224, 1992). Expression of a Type III LEA (HVA 1) in tobacco was found to influence plant height, maturity and drought

tolerance (Fitzpatrick, *Gen. Engineering News*, 22:7, 1993). In rice, expression of the HVA 1 gene influenced tolerance to water deficit and salinity (Xu et al., *Plant Physiol.*, 110:249-257, 1996). Expression of structural genes from any of the three LEA groups may therefore confer drought tolerance. Other types of proteins induced during water stress include thiol proteases, aldolases or transmembrane transporters (Guerrero et al., *Plant Molecular Biology*, 15:11-26, 1990), which may confer various protective and/or repair type functions during drought stress. It also is contemplated that genes that effect lipid biosynthesis and hence membrane composition might also be useful in conferring drought resistance on the plant.

Many of these genes for improving drought resistance have complementary modes of action. Thus, it is envisaged that combinations of these genes might have additive and/or synergistic effects in improving drought resistance in plants. Many of these genes also improve freezing tolerance (or resistance); the physical stresses incurred during freezing and drought are similar in nature and may be mitigated in similar fashion. Benefit may be conferred via constitutive expression of these genes, but the preferred means of expressing these novel genes may be through the use of a turgor induced promoter (such as the promoters for the turgor induced genes described in Guerrero et al., *Plant Molecular Biology*, 15:11-26, 1990 and Shagan et al., *Plant Physiol.*, 101:1397-1398, 1993). Spatial and temporal expression patterns of these genes may enable plants to better withstand stress.

It is proposed that expression of genes that are involved with specific morphological traits that allow for increased water extractions from drying soil would be of benefit. For example, introduction and expression of genes that alter root characteristics may enhance water uptake. It also is contemplated that expression of genes that enhance reproductive fitness during times of stress would be of significant value. For example, expression of genes that improve the synchrony of pollen shed and receptiveness of the female flower parts, e.g., silks, would be of benefit. In addition it is proposed that expression of genes that minimize kernel abortion during times of stress would increase the amount of grain to be harvested and hence be of value.

Given the overall role of water in determining yield, it is contemplated that enabling plants to utilize water more efficiently, through the introduction and expression of novel genes, will improve overall performance even when soil water availability is not limiting. By introducing genes that improve the ability of plants to maximize water usage across a full range of stresses relating to water availability, yield stability or consistency of yield performance may be realized.

Polypeptides that may improve stress tolerance under a variety of stress conditions include polypeptides involved in gene regulation, such as serine/threonine-protein kinases, MAP kinases, MAP kinase kinases, and MAP kinase kinase kinases; polypeptides that act as receptors for signal transduction and regulation, such as receptor protein kinases; intracellular signaling proteins, such as protein phosphatases, GTP binding proteins, and phospholipid signaling proteins; polypeptides involved in arginine biosynthesis; polypeptides involved in ATP metabolism, including for example ATPase, adenylate transporters, and polypeptides involved in ATP synthesis and transport; polypeptides involved in glycine betaine, jasmonic acid, flavonoid or steroid biosynthesis; and hemoglobin. Enhanced or reduced activity of such polypeptides in modified plants will provide changes in the ability of a plant to respond to a variety of environmental stresses, such as chemical stress, drought stress and pest stress.

Other polypeptides that may improve plant tolerance to cold or freezing temperatures include polypeptides involved in biosynthesis of trehalose or raffinose, polypeptides encoded by cold induced genes, fatty acyl desaturases and other polypeptides involved in glycerolipid or membrane lipid biosynthesis, which find use in modification of membrane fatty acid composition, alternative oxidase, calcium-dependent protein kinases, LEA proteins or uncoupling protein.

Other polypeptides that may improve plant tolerance to heat include polypeptides involved in biosynthesis of trehalose, polypeptides involved in glycerolipid biosynthesis or membrane lipid metabolism (for altering membrane fatty acid composition), heat shock proteins or mitochondrial NDK.

Other polypeptides that may improve tolerance to extreme osmotic conditions include polypeptides involved in proline biosynthesis.

Other polypeptides that may improve plant tolerance to drought conditions include aquaporins, polypeptides involved in biosynthesis of trehalose or wax, LEA proteins or invertase.

(iv) *Disease Resistance*

5 It is proposed that increased resistance (or tolerance) to diseases may be realized through introduction of genes into plants, for example, into monocotyledonous plants such as maize. It is possible to produce resistance to diseases caused by viruses, viroids, bacteria, fungi and nematodes. It also is contemplated that control of mycotoxin producing organisms may be realized through
10 expression of introduced genes. Resistance can be affected through suppression of endogenous factors that encourage disease-causing interactions, expression of exogenous factors that are toxic to or otherwise provide protection from pathogens, or expression of factors that enhance the plant's own defense responses.

Resistance to viruses may be produced through expression of novel
15 genes. For example, it has been demonstrated that expression of a viral coat protein in a modified plant can impart resistance to infection of the plant by that virus and perhaps other closely related viruses (Cuozzo et al., Bio/Technology, 6:549-553, 1988, Hemenway et al., The EMBO J., 7:1273-1280, 1988, Abel et al., Science, 232:738-743, 1986). It is contemplated that expression of antisense genes targeted at
20 essential viral functions may also impart resistance to viruses. For example, an antisense gene targeted at the gene responsible for replication of viral nucleic acid may inhibit replication and lead to resistance to the virus. It is believed that interference with other viral functions through the use of antisense genes also may increase resistance to viruses. Further, it is proposed that it may be possible to
25 achieve resistance to viruses through other approaches, including, but not limited to the use of satellite viruses.

It is proposed that increased resistance to diseases caused by bacteria and fungi may be realized through introduction of novel genes. It is contemplated that genes encoding so called "peptide antibiotics," pathogenesis related (PR)
30 proteins, toxin resistance, or proteins affecting host pathogen interactions such as morphological characteristics will be useful. Peptide antibiotics are polypeptide sequences which are inhibitory to growth of bacteria and other microorganisms. For

example, the classes of peptides referred to as cecropins and magainins inhibit growth of many species of bacteria and fungi. It is proposed that expression of PR proteins in plants, for example, monocots such as maize, may be useful in conferring resistance to bacterial disease. These genes are induced following pathogen attack on a host
5 plant and have been divided into at least five classes of proteins (Bol, Linthorst, and Cornelissen, 1990). Included amongst the PR proteins are beta 1, 3 glucanases, chitinases, and osmotin and other proteins that are believed to function in plant resistance to disease organisms. Other genes have been identified that have antifungal properties, *e.g.*, UDA (stinging nettle lectin), or hevein (Broekaert et al., 1989; Barkai
10 Golan et al., 1978). It is known that certain plant diseases are caused by the production of phytotoxins. It is proposed that resistance to these diseases would be achieved through expression of a novel gene that encodes an enzyme capable of degrading or otherwise inactivating the phytotoxin. It also is contemplated that expression of novel genes that alter the interactions between the host plant and
15 pathogen may be useful in reducing the ability of the disease organism to invade the tissues of the host plant; *e.g.*, an increase in the waxiness of the leaf cuticle or other morphological characteristics.

Polypeptides useful for imparting improved disease responses to plants include polypeptides encoded by cercosporin induced genes, antifungal proteins and
20 proteins encoded by R-genes or SAR genes.

Agronomically important diseases caused by fungal phytopathogens include: glume or leaf blotch, late blight, stalk/head rot, rice blast, leaf blight and spot, corn smut, wilt, sheath blight, stem canker, root rot, blackleg or kernel rot.

Exemplary plant viruses include tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Specific fungal, bacterial
25 and viral pathogens of major crops include, but are not limited to:

RICE: rice brown spot fungus (*Cochliobolus miyabeanus*), rice blast fungus--*Magnaporthe grisea* (*Pyricularia grisea*), *Magnaporthe salvinii* (*Sclerotium oryzae*), *Xanthomomas oryzae* pv. *oryzae*, *Xanthomomas oryzae* pv. *oryzicola*,
30 *Rhizoctonia* spp. (including but not limited to *Rhizoctonia solani*, *Rhizoctonia oryzae* and *Rhizoctonia oryzae-sativae*), *Pseudomonas* spp. (including but not limited to *Pseudomonas plantarii*, *Pseudomonas avenae*, *Pseudomonas glumae*, *Pseudomonas*

- fuscovaginae, *Pseudomonas alboprecipitans*, *Pseudomonas syringae* pv. *panici*,
Pseudomonas syringae pv. *syringae*, *Pseudomonas syringae* pv. *oryzae* and
Pseudomonas syringae pv. *aptata*), *Erwinia* spp. (including but not limited to *Erwinia*
herbicola, *Erwinia amylovaora*, *Erwinia chrysanthemi* and *Erwinia carotovora*),
 5 *Achyla* spp. (including but not limited to *Achyla conspicua* and *Achyia klebsiana*),
Pythium spp. (including but not limited to *Pythium dissotocum*, *Pythium irregulare*,
Pythium arrhenomanes, *Pythium myriotylum*, *Pythium catenulatum*, *Pythium*
graminicola and *Pythium spinosum*), *Saprolegnia* spp., *Dictyuchus* spp., *Pythiogeton*
 spp., *Phytophthora* spp., *Alternaria padwickii*, *Cochliobolus miyabeanus*, *Curvularia*
 10 spp. (including but not limited to *Curvularia lunata*, *Curvularia affinis*, *Curvularia*
clavata, *Curvularia eragrostidis*, *Curvularia fallax*, *Curvularia geniculata*, *Curvularia*
inaequalis, *Curvularia intermedia*, *Curvularia oryzae*, *Curvularia oryzae-sativae*,
Curvularia pallescens, *Curvularia senegalensis*, *Curvularia tuberculata*, *Curvularia*
uncinata and *Curvularia verruculosa*), *Sarocladium oryzae*, *Gerlachia oryzae*,
 15 *Fusarium* spp. (including but not limited *Fusarium graminearum*, *Fusarium nivale* and
 to different pathovars of *Fusarium moniliforme*, including pvs. *fujikuroi* and *zeae*),
Sclerotium rolfsii, *Phoma exigua*, *Mucor fragilis*, *Trichoderma viride*, *Rhizopus* spp.,
Cercospora oryzae, *Entyloma oryzae*, *Dreschlera gigantea*, *Scierophthora*
macrospora, *Mycovellosiella oryzae*, *Phomopsis oryzae-sativae*, *Puccinia graminis*,
 20 *Uromyces coronatus*, *Cylindrocladium scoparium*, *Sarocladium oryzae*,
Gaeumannomyces graminis pv. *graminis*, *Myrothecium verrucaria*, *Pyrenochaeta*
oryzae, *Ustilaginoidea virens*, *Neovossia* spp. (including but not limited to *Neovossia*
horrida), *Tilletia* spp., *Balansia oryzae-sativae*, *Phoma* spp. (including but not limited
 to *Phoma sorghina*, *Phoma insidiosa*, *Phoma glumarum*, *Phoma glumicola* and *Phoma*
 25 *oryzina*), *Nigrospora* spp. (including but not limited to *Nigrospora oryzae*, *Nigrospora*
sphaerica, *Nigrospora panici* and *Nigrospora padwickii*), *Epiococcum nigrum*,
Phyllostica spp., *Wolkia decolorans*, *Monascus purpureus*, *Aspergillus* spp.,
Penicillium spp., *Absidia* spp., *Mucor* spp., *Chaetomium* spp., *Dematium* spp.,
Monilia spp., *Streptomyces* spp., *Syncephalastrum* spp., *Verticillium* spp.,
 30 *Nematospora coryli*, *Nakataea sigmoidea*, *Cladosporium* spp., *Bipolaris* spp.,
Coniothyrium spp., *Diplodia oryzae*, *Exserophilum rostratum*, *Helococera oryzae*,
Melanomma glumarum, *Metashaeria* spp., *Mycosphaerella* spp., *Oidium* spp.,
Pestalotia spp., *Phaeoseptoria* spp., *Sphaeropsis* spp., *Trematosphaerella* spp., rice

- black-streaked dwarf virus, rice dwarf virus, rice gall dwarf virus, barley yellow dwarf virus, rice grassy stunt virus, rice hoja blanca virus, rice necrosis mosaic virus, rice ragged stunt virus, rice stripe virus, rice stripe necrosis virus, rice transitory yellowing virus, rice tungro bacilliform virus, rice tungro spherical virus, rice yellow mottle virus, rice tarsonemid mite virus, Echinochloa hoja blanca virus, Echinochloa ragged stunt virus, orange leaf mycoplasma-like organism, yellow dwarf mycoplasma-like organism, Aphelenchoides besseyi, Ditylenchus angustus, Hirschmanniella spp., Criconemella spp., Meloidogyne spp., Heterodera spp., Pratylenchus spp., Hoplolaimus indicus.
- 10 SOYBEANS: *Phytophthora sojae*, *Fusarium solani* f. sp. *Glycines*, *Macrophomina phaseolina*, *Fusarium*, *Pythium*, *Rhizoctonia*, *Phialophora gregata*, *Sclerotinia sclerotiorum*, *Diaporthe phaseolorum* var. *sojae*, *Colletotrichum truncatum*, *Phomopsis longicolla*, *Cercospora kikuchii*, *Diaporthe phaseolonum* var. *meridionalis* (and var. *caulivora*), *Phakopsora pachyrhizi*, *Fusarium solani*,
- 15 *Microsphaera diffusa*, *Septoria glycines*, *Cercospora kikuchii*, *Macrophomina phaseolina*, *Sclerotinia sclerotiorum*, *Corynespora cassiicola*, *Rhizoctonia solani*, *Cercospora sojae*, *Phytophthora megasperma* fsp. *glycinea*, *Macrophomina phaseolina*, *Fusarium oxysporum*, *Diaporthe phaseolorum* var. *sojae* (*Phomopsis sojae*), *Diaporthe phaseolorum* var. *caulivora*, *Sclerotium rolfsii*, *Cercospora kikuchii*,
- 20 *Cercospora sojae*, *Peronospora manshurica*, *Colletotrichum dematium* (*Colletotrichum truncatum*), *Corynespora cassiicola*, *Phyllosticta sojicola*, *Alternaria alternata*, *Pseudomonas syringae* p.v. *glycinea*, *Xanthomonas campestris* p.v. *phaseoli*, *Microsphaera diffusa*, *Fusarium semitectum*, *Phialophora gregata*, Soybean mosaic virus, *Glomerella glycines*, Tobacco Ring spot virus, Tobacco Streak virus,
- 25 *Phakopsora pachyrhizi*, *Pythium aphanidermatum*, *Pythium ultimum*, *Pythium dearyanum*, Tomato spotted wilted virus, *Heterodera glycines*, *Fusarium solani*, Soybean cyst and root knot nematodes.

- CORN: *Fusarium moniliforme* var. *subglutinans*, *Erwinia stewartii*, *Fusarium moniliforme*, *Gibberella zeae* (*Fusarium Graminearum*), *Stenocarpella maydi* (*Diplodia maydis*), *Pythium irregulare*, *Pythium debaryanum*, *Pythium graminicola*, *Pythium splendens*, *Pythium ultimum*, *Pythium aphanidermatum*, *Aspergillus flavus*, *Bipolaris maydis* O, T (*cochliobolus heterostrophus*), *Helminthosporium carbonum* I, II, and III (*Cochliobolus carbonum*), *Exserohilum*

turcicum I, II and III, Helminthosporium pedicellatum, Physoderma maydis, Phyllosticta maydis, Kabatie-maydis, Cercospora sorghi, Ustilago maydis, Puccinia sorghi, Puccinia polysora, Macrophomina phaseolina, Penicillium oxalicum, Nigrospora oryzae, Cladosporium herbarum, Curvularia lunata, Curvularia inaequalis, 5 Curvularia pallescens, Clavibacter michiganese subsp. Nebraskense, Trichoderma viride, Maize dwarf Mosaic Virus A and B, Wheat Streak Mosaic Virus, Maize Chlorotic Dwarf Virus, Claviceps sorghi, Pseudomonas avenae, Erwinia chrysantemi p.v. Zea, Erwinia corotovora, Cornstun spiroplasma, Diplodia macrospora, Sclerophthora macrospora, Peronosclerospora sorghi, Peronosclerospora 10 philippinesis, Peronosclerospora maydis, Peronosclerospora sacchari, Spacelotheca reiliana, Physopella zea, Cephalosporium maydis, Cephalosporium acremonium, Maize Chlorotic Mottle Virus, High Plains Virus, Maize Mosaic Virus, Maize Rayado Fino Virus, Maize Streak Virus, Maize Stripe Virus, Maize Rought Dwarf Virus:

WHEAT: Pseudomonas syringae p.v. atrofaciens, Urocystis agropyri, 15 Xanthomonas campestris p.v. translucens, Pseudomonas syringae p.v. syringae, Alternaria alternata, Cladosporium herbarum, Fusarium graminearum, Fusarium avenaceum, Fusarium culmorum, Ustilago tritici, Ascochyta tritici, Cephalosporium gramineum, Collotetrichum graminicola, Erysiphe graminis f. sp. Tritici, Puccinia graminis f. sp. Tritici, Puccinia recondite f. sp. tritici, puccinia striiformis, 20 Pyrenophora triticirepentis, Septoria nodorum, Septoria tritici, Spetoria avenae, Pseudocercosporella herpotrichoides, Rhizoctonia solani, Rhizoctonia cerealis, Gaeumannomyces graminis var. tritici, Pythium aphanidermatum, Pythium arrhenomanes, Pythium ultimum, Bipolaris sorokiniana, Barley Yellow Dwarf Virus, Brome Mosaic Virus, Soil Borne Wheat Mosaic Virus, Wheat Streak Virus, Wheat 25 Spindle Streak Virus, American Wheat Striate Virus, Claviceps purpurea, Tilletia tritici, Tilletia laevis, Pstilago tritici, Tilletia indica, Rhizoctonia solani, Pythium arrhenomannes, Pythium gramicola, Pythium aphanidermatum, High Plains Virus, European Wheat Striate Virus:

CANOLA: Albugo candida, Alternaria brassicae, Leptosharia 30 maculans, Rhizoctonia solani, Sclerotinia sclerotiorum, Mycosphaerella brassiccola, Pythium ultimum, Peronospora parasitica, Fusarium roseum, Fusarium oxysporum, Tilletia foetida, Tilletia caries, Alternaria alternata:

SUNFLOWER: *Plasmophora halstedii*, *Sclerotinia sclerotiorum*, Aster
 Yellows, *Septoria helianthi*, *Phomopsis helianthi*, *Alternaria helianthi*, *Alternaria*
zinniae, *Botrytis cinerea*, *Phoma macdonaldii*, *Macrophomina phaseolina*, *Erysiphe*
cichoracearum, *Phizopus oryzae*, *Rhizopus arrhizus*, *Rhizopus stolonifer*, *Puccinia*
 5 *helianthi*, *Verticillium Dahliae*, *Erwinia carotovorum* p.v. *carotovora*,
Cephalosporium acremonium, *Phytophthora cryptogea*, *Albugo tragopogonis*.

SORGHUM: *Exserohilum turcicum*, *Colletotrichum graminicola*
 (*Glomerella graminicola*), *Cercospora sorghi*, *Gloeocercospora sorghi*, *Ascochyta*
sorghi, *Pseudomonas syringae* p.v. *syringae*, *Xanthomonas campestris* p.v. *holcicola*,
 10 *Pseudomonas andropogonis*, *Puccinia purpurea*, *Macrophomina phaseolina*, *Periconia*
circinata, *Fusarium moniliforme*, *Alternaria alternate*, *Bipolaris sorghicola*,
Helminthosporium sorghicola, *Curvularia lunata*, *Phoma insidiosa*, *Pseudomonas*
avenae (*Pseudomonas alboprecipitans*), *Ramulispora sorghi*, *Ramulispora sorghicola*,
Phyllachara sacchari *Sporisorium relianum* (*Sphacelotheca reliana*), *Sphacelotheca*
 15 *cruenta*, *Sporisorium sorghi*, Sugarcane mosaic H, Maize Dwarf Mosaic Virus A & B,
Claviceps sorghi, *Rhizoctonia solani*, *Acremonium strictum*, *Sclerophthora*
macrospora, *Peronosclerospora sorghi*, *Peronosclerospora philippinensis*, *Sclerospora*
graminicola, *Fusarium graminearum*, *Fusarium Oxysporum*, *Pythium arrhenomanes*,
Pythium graminicola.

20 ALFALFA: *Clavibacter michiganensis* subsp. *insidiosum*, *Pythium*
ultimum, *Pythium irregulare*, *Pythium splendens*, *Pythium debaryanum*, *Pythium*
aphanidermatum, *Phytophthora megasperma*, *Peronospora trifoliorum*, *Phoma*
medicaginis var. *medicaginis*, *Cercospora medicaginis*, *Pseudopeziza medicaginis*,
Leptotrochila medicaginis, *Fusarium oxysporum*, *Rhizoctonia solani*, *Uromyces*
 25 *striatus*, *Colletotrichum trifolii* race 1 and race 2, *Leptosphaerulina briosiana*,
Stemphylium botryosum, *Stagonospora meliloti*, *Sclerotinia trifoliorum*, Alfalfa
 Mosaic Virus, *Verticillium albo-atrum*, *Xanthomonas campestris* p.v. *alfalfae*,
Aphanomyces euteiches, *Stemphylium herbarum*, *Stemphylium alfalfae*.

(v) *Plant Agronomic Characteristics*

30 Two of the factors determining where crop plants can be grown are the
 average daily temperature during the growing season and the length of time between
 frosts. Within the areas where it is possible to grow a particular crop, there are

varying limitations on the maximal time it is allowed to grow to maturity and be harvested. For example, a variety to be grown in a particular area is selected for its ability to mature and dry down to harvestable moisture content within the required period of time with maximum possible yield. Therefore, crops of varying maturities
5 are developed for different growing locations. Apart from the need to dry down sufficiently to permit harvest, it is desirable to have maximal drying take place in the field to minimize the amount of energy required for additional drying post harvest. Also, the more readily a product such as grain can dry down, the more time there is available for growth and kernel fill. It is considered that genes that influence maturity
10 and/or dry down can be identified and introduced into plant lines using transformation techniques to create new varieties adapted to different growing locations or the same growing location, but having improved yield to moisture ratio at harvest. Expression of genes that are involved in regulation of plant development may be especially useful.

15 It is contemplated that genes may be introduced into plants that would improve standability and other plant growth characteristics. Expression of novel genes in plants which confer stronger stalks, improved root systems, or prevent or reduce ear droppage or shattering would be of great value to the farmer. It is proposed that introduction and expression of genes that increase the total amount of
20 photoassimilate available by, for example, increasing light distribution and/or interception would be advantageous. In addition, the expression of genes that increase the efficiency of photosynthesis and/or the leaf canopy would further increase gains in productivity. It is contemplated that expression of a phytochrome gene in crop plants may be advantageous. Expression of such a gene may reduce
25 apical dominance, confer semidwarfism on a plant, or increase shade tolerance (U.S. Patent No. 5,268,526). Such approaches would allow for increased plant populations in the field.

(vi) *Nutrient Utilization*

30 The ability to utilize available nutrients may be a limiting factor in growth of crop plants. It is proposed that it would be possible to alter nutrient uptake, tolerate pH extremes, mobilization through the plant, storage pools, and availability for metabolic activities by the introduction of novel genes. These modifications

would allow a plant, for example, maize to more efficiently utilize available nutrients. It is contemplated that an increase in the activity of, for example, an enzyme that is normally present in the plant and involved in nutrient utilization would increase the availability of a nutrient or decrease the availability of an antinutritive factor. An
5 example of such an enzyme would be phytase. It is further contemplated that enhanced nitrogen utilization by a plant is desirable. Expression of a glutamate dehydrogenase gene in plants, e.g., *E. coli* *gdhA* genes, may lead to increased fixation of nitrogen in organic compounds. Furthermore, expression of *gdhA* in plants may lead to enhanced resistance to the herbicide glufosinate by incorporation of excess
10 ammonia into glutamate, thereby detoxifying the ammonia. It also is contemplated that expression of a novel gene may make a nutrient source available that was previously not accessible, e.g., an enzyme that releases a component of nutrient value from a more complex molecule, perhaps a macromolecule.

Polypeptides useful for improving nitrogen flow, sensing, uptake,
15 storage and/or transport include those involved in aspartate, glutamine or glutamate biosynthesis, polypeptides involved in aspartate, glutamine or glutamate transport, polypeptides associated with the TOR (Target of Rapamycin) pathway, nitrate transporters, nitrate reductases, amino transferases, ammonium transporters, chlorate transporters or polypeptides involved in tetrapyrrole biosynthesis.

20 Polypeptides useful for increasing the rate of photosynthesis include phytochrome, ribulose biphosphate carboxylase-oxygenase, Rubisco activase, photosystem I and II proteins, electron carriers, ATP synthase, NADH dehydrogenase or cytochrome oxidase.

Polypeptides useful for increasing phosphorus uptake, transport or
25 utilization include phosphatases or phosphate transporters.

(vii) *Male Sterility*

Male sterility is useful in the production of hybrid seed. It is proposed that male sterility may be produced through expression of novel genes. For example, it has been shown that expression of genes that encode proteins, RNAs, or peptides
30 that interfere with development of the male inflorescence and/or gametophyte result in male sterility. Chimeric ribonuclease genes that express in the anthers of

transgenic tobacco and oilseed rape have been demonstrated to lead to male sterility (Mariani et al., Nature, 347:737-741, 1990).

A number of mutations were discovered in maize that confer cytoplasmic male sterility. One mutation in particular, referred to as T cytoplasm, also correlates with sensitivity to Southern corn leaf blight. A DNA sequence, designated TURF 13 (Levings, Science, 250:942-947, 1990), was identified that correlates with T cytoplasm. It is proposed that it would be possible through the introduction of TURF 13 via transformation, to separate male sterility from disease sensitivity. As it is necessary to be able to restore male fertility for breeding purposes and for grain production, it is proposed that genes encoding restoration of male fertility also may be introduced.

(viii) Altered Nutritional Content

Genes may be introduced into plants to improve or alter the nutrient quality or content of a particular crop. Introduction of genes that alter the composition of a crop may greatly enhance the feed or food value. For example, the protein of many grains is suboptimal for feed and food purposes, especially when fed to pigs, poultry, and humans. The protein is deficient in several amino acids that are essential in the diet of these species, requiring the addition of supplements to the grain. Limiting essential amino acids may include lysine, methionine, tryptophan, threonine, valine, arginine, and histidine. Some amino acids become limiting only after corn is supplemented with other inputs for feed formulations. The levels of these essential amino acids in seeds and grain may be elevated by mechanisms which include, but are not limited to, the introduction of genes to increase the biosynthesis of the amino acids, decrease the degradation of the amino acids, increase the storage of the amino acids in proteins, or increase transport of the amino acids to the seeds or grain.

Polypeptides useful for providing increased seed protein quantity and/or quality include polypeptides involved in the metabolism of amino acids in plants, particularly polypeptides involved in biosynthesis of methionine/cysteine and lysine, amino acid transporters, amino acid efflux carriers, seed storage proteins, proteases, or polypeptides involved in phytic acid metabolism.

The protein composition of a crop may be altered to improve the balance of amino acids in a variety of ways including elevating expression of native proteins, decreasing expression of those with poor composition, changing the composition of native proteins, or introducing genes encoding entirely new proteins

5. possessing superior composition.

The introduction of genes that alter the oil content of a crop plant may also be of value. Increases in oil content may result in increases in metabolizable-energy-content and density of the seeds for use in feed and food. The introduced genes may encode enzymes that remove or reduce rate-limitations or regulated steps

10 in fatty acid or lipid biosynthesis. Such genes may include, but are not limited to, those that encode acetyl-CoA carboxylase, ACP-acyltransferase, alpha-ketoacyl-ACP synthase, or other well known fatty acid biosynthetic activities. Other possibilities are genes that encode proteins that do not possess enzymatic activity such as acyl carrier protein. Genes may be introduced that alter the balance of fatty acids present in the

15 oil providing a more healthful or nutritive feedstuff. The introduced DNA also may encode sequences that block expression of enzymes involved in fatty acid biosynthesis, altering the proportions of fatty acids present in crops.

Genes may be introduced that enhance the nutritive value of crops, or of foods derived from crops by increasing the level of naturally occurring

20 phytosterols, or by encoding for proteins to enable the synthesis of phytosterols in crops. The phytosterols from these crops can be processed directly into foods, or extracted and used to manufacture food products.

Genes may be introduced that enhance the nutritive value of the starch component of crops, for example by increasing the degree of branching, resulting in

25 improved utilization of the starch in livestock by delaying its metabolism. Additionally, other major constituents of a crop may be altered, including genes that affect a variety of other nutritive, processing, or other quality aspects. For example, pigmentation may be increased or decreased.

Carbohydrate metabolism may be altered, for example by increased

30 sucrose production and/or transport. Polypeptides useful for affecting on carbohydrate metabolism include polypeptides involved in sucrose or starch metabolism, carbon assimilation or carbohydrate transport, including, for example sucrose transporters or

glucose/hexose transporters, enzymes involved in glycolysis/gluconeogenesis, the pentose phosphate cycle, or raffinose biosynthesis, or polypeptides involved in glucose signaling, such as SNF1 complex proteins.

Feed or food crops may also possess sub-optimal quantities of
5 vitamins, antioxidants or other nutraceuticals, requiring supplementation to provide adequate nutritive value and ideal health value. Introduction of genes that enhance vitamin biosynthesis may be envisioned including, for example, vitamins A, E, B12, choline, or the like. Mineral content may also be sub-optimal. Thus genes that affect the accumulation or availability of compounds containing phosphorus, sulfur,
10 calcium, manganese, zinc, or iron among others would be valuable.

Numerous other examples of improvements of crops may be used with the invention. The improvements may not necessarily involve grain, but may, for example, improve the value of a crop for silage. Introduction of DNA to accomplish this might include sequences that alter lignin production such as those that result in
15 the "brown midrib" phenotype associated with superior feed value for cattle. Other genes may encode for enzymes that alter the structure of extracellular carbohydrates in the stover, or that facilitate the degradation of the carbohydrates in the non-grain portion of the crop so that it can be efficiently fermented into ethanol or other useful carbohydrates.

20 It may be desirable to modify the nutritional content of plants by reducing undesirable components such as fats, starches, etc. This may be done, for example, by the use of exogenous nucleic acids that encode enzymes which increase plant use or metabolism of such components so that they are present at lower quantities. Alternatively, it may be done by use of exogenous nucleic acids that
25 reduce expression levels or activity of native plant enzymes that synthesize such components.

Likewise the elimination of certain undesirable traits may improve the food or feed value of the crop. Many undesirable traits must currently be eliminated by special post-harvest processing steps and the degree to which these can be
30 engineered into the plant prior to harvest and processing would provide significant value. Examples of such traits are the elimination of anti-nutritionals such as phytates and phenolic compounds which are commonly found in many crop species. Also, the

reduction of fats, carbohydrates and certain phytohormones may be valuable for the food and feed industries as they may allow a more efficient mechanism to meet specific dietary requirements.

5 In addition to direct improvements in feed or food value, genes also may be introduced which improve the processing of crops and improve the value of the products resulting from the processing. One use of crops is via wetmilling. Thus novel genes that increase the efficiency and reduce the cost of such processing, for example by decreasing steeping time, may also find use. Improving the value of wetmilling products may include altering the quantity or quality of starch, oil, corn
10 gluten meal, or the components of gluten feed. Elevation of starch may be achieved through the identification and elimination of rate limiting steps in starch biosynthesis by expressing increased amounts of enzymes involved in biosynthesis or by decreasing levels of the other components of crops resulting in proportional increases in starch.

15 Oil is another product of wetmilling, the value of which may be improved by introduction and expression of genes. Oil properties may be altered to improve its performance in the production and use of cooking oil, shortenings, lubricants or other oil-derived products or improvement of its health attributes when used in the food-related applications. Novel fatty acids also may be synthesized
20 which upon extraction can serve as starting materials for chemical syntheses. The changes in oil properties may be achieved by altering the type, level, or lipid arrangement of the fatty acids present in the oil. This in turn may be accomplished by the addition of genes that encode enzymes that catalyze the synthesis of novel fatty acids (*e.g.* fatty acid elongases, desaturases) and the lipids possessing them or by
25 increasing levels of native fatty acids while possibly reducing levels of precursors or breakdown products. Alternatively, DNA sequences may be introduced which slow or block steps in fatty acid biosynthesis resulting in the increase in precursor fatty acid intermediates. Genes that might be added include desaturases, epoxidases, hydratases, dehydratases, or other enzymes that catalyze reactions involving fatty acid
30 intermediates. Representative examples of catalytic steps that might be blocked include the desaturations from stearic to oleic acid or oleic to linolenic acid resulting in the respective accumulations of stearic and oleic acids. Another example is the

blockage of elongation steps resulting in the accumulation of C8 to C12 saturated fatty acids.

Polypeptides useful for providing increased seed oil quantity and/or quality include polypeptides involved in fatty acid and glycerolipid biosynthesis, beta-oxidation enzymes, enzymes involved in biosynthesis of nutritional compounds, such as carotenoids and tocopherols, or polypeptides that increase embryo size or number or thickness of aleurone.

Polypeptides involved in production of galactomannans or arabinogalactans are of interest for providing plants having increased and/or modified reserve polysaccharides for use in food, pharmaceutical, cosmetic, paper and paint industries.

Polypeptides involved in modification of flavonoid/isoflavonoid metabolism in plants include cinnamate-4-hydroxylase, chalcone synthase or flavones synthase. Enhanced or reduced activity of such polypeptides in modified plants will provide changes in the quantity and/or speed of flavonoid metabolism in plants and may improve disease resistance by enhancing synthesis of protective secondary metabolites or improving signaling pathways governing disease resistance.

Polypeptides involved in lignin biosynthesis are of interest for increasing plants' resistance to lodging and for increasing the usefulness of plant materials as biofuels.

(ix) Production or Assimilation of Chemicals or Biological

It may further be considered that a modified plant prepared in accordance with the invention may be used for the production or manufacturing of useful biological compounds that were either not produced at all, or not produced at the same level, in the corn plant previously. Alternatively, plants produced in accordance with the invention may be made to metabolize or absorb and concentrate certain compounds, such as hazardous wastes, thereby allowing bioremediation of these compounds.

The novel plants producing these compounds are made possible by the introduction and expression of one or potentially many genes with the constructs provided by the invention. The vast array of possibilities include but are not limited

to any biological compound which is presently produced by any organism such as proteins, nucleic acids, primary and intermediary metabolites, carbohydrate polymers, enzymes for uses in bioremediation, enzymes for modifying pathways that produce secondary plant metabolites such as falconoid or vitamins, enzymes that could
5 produce pharmaceuticals, and for introducing enzymes that could produce compounds of interest to the manufacturing industry such as specialty chemicals and plastics. The compounds may be produced by the plant, extracted upon harvest and/or processing, and used for any presently recognized useful purpose such as pharmaceuticals, fragrances, and industrial enzymes to name a few.

10 *(x) Other characteristics*

Cell cycle modification: Polypeptides encoding cell cycle enzymes and regulators of the cell cycle pathway are useful for manipulating growth rate in plants to provide early vigor and accelerated maturation. Improvements in quality traits, such as seed oil content, may also be obtained by expression of cell cycle enzymes
15 and cell cycle regulators. Polypeptides of interest for modification of cell cycle pathway include cycling and EIF5 α pathway proteins, polypeptides involved in polyamine metabolism, polypeptides which act as regulators of the cell cycle pathway, including cyclin-dependent kinases (CDKs), CDK-activating kinases, cell cycle-dependent phosphatases, CDK-inhibitors, Rb and Rb-binding proteins, or
20 transcription factors that activate genes involved in cell proliferation and division, such as the E2F family of transcription factors, proteins involved in degradation of cyclins, such as cullins, and plant homologs of tumor suppressor polypeptides.

Plant growth regulators: Polypeptides involved in production of substances that regulate the growth of various plant tissues are of interest in the
25 present invention and may be used to provide modified plants having altered morphologies and improved plant growth and development profiles leading to improvements in yield and stress response. Of particular interest are polypeptides involved in the biosynthesis, or degradation of plant growth hormones, such as gibberellins, brassinosteroids, cytokinins, auxins, ethylene or abscisic acid, and other
30 proteins involved in the activity, uptake and/or transport of such polypeptides, including for example, cytokinin oxidase, cytokinin/purine permeases, F-box proteins, G-proteins or phytosulfokines.

Transcription factors in plants: Transcription factors play a key role in plant growth and development by controlling the expression of one or more genes in temporal, spatial and physiological specific patterns. Enhanced or reduced activity of such polypeptides in modified plants will provide significant changes in gene transcription patterns and provide a variety of beneficial effects in plant growth, development and response to environmental conditions. Transcription factors of interest include, but are not limited to myb transcription factors, including helix-turn-helix proteins, homeodomain transcription factors, leucine zipper transcription factors, MADS transcription factors, transcription factors having AP2 domains, zinc finger transcription factors, CCAAT binding transcription factors, ethylene responsive transcription factors, transcription initiation factors or UV damaged DNA binding proteins.

Homologous recombination: Increasing the rate of homologous recombination in plants is useful for accelerating the introgression of transgenes into breeding varieties by backcrossing, and to enhance the conventional breeding process by allowing rare recombinants between closely linked genes in phase repulsion to be identified more easily. Polypeptides useful for expression in plants to provide increased homologous recombination include polypeptides involved in mitosis and/or meiosis, DNA replication, nucleic acid metabolism, DNA repair pathways or homologous recombination pathways including for example, recombinases, nucleases, proteins binding to DNA double-strand breaks, single-strand DNA binding proteins, strand-exchange proteins, resolvases, ligases, helicases and polypeptide members of the RAD52 epistasis group.

Non-Protein-Expressing Exogenous Nucleic Acids

Plants with decreased expression of a gene of interest can also be achieved, for example, by expression of antisense nucleic acids, dsRNA or RNAi, catalytic RNA such as ribozymes, sense expression constructs that exhibit cosuppression effects, aptamers or zinc finger proteins.

Antisense RNA reduces production of the polypeptide product of the target messenger RNA, for example by blocking translation through formation of RNA:RNA duplexes or by inducing degradation of the target mRNA. Antisense approaches are a way of preventing or reducing gene function by targeting the genetic material as disclosed in U.S. Pat. Nos. 4,801,540; 5,107,065; 5,759,829; 5,910,444;

6,184,439; and 6,198,026. In one approach, an antisense gene sequence is introduced that is transcribed into antisense RNA that is complementary to the target mRNA. For example, part or all of the normal gene sequences are placed under a promoter in inverted orientation so that the 'wrong' or complementary strand is transcribed into a non-protein expressing antisense RNA. The promoter used for the antisense gene may influence the level, timing, tissue, specificity, or inducibility of the antisense inhibition.

Autonomous mini-chromosomes may contain exogenous DNA bounded by recombination sites, for example lox-P sites, that can be recognized by a recombinase, e.g. Cre, and removed from the mini-chromosome. In cases where there is a homologous recombination site or sites in the host genomic DNA, the exogenous DNA excised the mini-chromosome may be integrated into the genome at one of the specific recombination sites and the DNA bounded by the recombination sites will become integrated into the host DNA. The use of a mini-chromosome as a platform for DNA excision or for launching such DNA integration into the host genome may include in vivo induction of the expression of a recombinase encoded in the genomic DNA of a transgenic host, or in a mini-chromosome or other episome.

RNAi gene suppression in plants by transcription of a dsRNA is described in U.S. Pat. No. 6,506,559, U.S. patent application Publication No. 2002/0168707, WO 98/53083, WO 99/53050 and WO 99/61631. The double-stranded RNA or RNAi constructs can trigger the sequence-specific degradation of the target messenger RNA. Suppression of a gene by RNAi can be achieved using a recombinant DNA construct having a promoter operably linked to a DNA element comprising a sense and anti-sense element of a segment of genomic DNA of the gene, e.g., a segment of at least about 23 nucleotides, more preferably about 50 to 200 nucleotides where the sense and anti-sense DNA components can be directly linked or joined by an intron or artificial DNA segment that can form a loop when the transcribed RNA hybridizes to form a hairpin structure.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of the target gene or genes or facilitate molecular reactions. Ribozymes are targeted to a given sequence by hybridization of sequences within the ribozyme to the target mRNA.

Two stretches of homology are required for this targeting, and these stretches of homologous sequences flank the catalytic ribozyme structure. It is possible to design ribozymes that specifically pair with virtually any target mRNA and cleave the target mRNA at a specific location, thereby inactivating it. A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs that are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include Tobacco Ringspot Virus (Prody *et al*, *Science*, 231:1577-1580, 1986), Avocado Sunblotch Viroid (Palukaitis *et al*, *Virology*, 99:145-151, 1979; Symons, *Nucl Acids Res.*, 9:6527-6537, 1981), and Lucerne Transient Streak Virus (Forster and Symons, *Cell*, 49:211-220, 1987), and the satellite RNAs from velvet tobacco mottle virus, *Solanum nodiflorum* mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff, et al., *Nature* 334:585-591 (1988). Several different ribozyme motifs have been described with RNA cleavage activity (Symons, *Annu. Rev. Biochem.*, 61:641-671, 1992). Other suitable ribozymes include sequences from RNase P with RNA cleavage activity (Yuan *et al*, *Proc. Natl. Acad. Sci. USA*, 89:8006-8010, 1992; Yuan and Altman, *Science*, 263:1269-1273, 1994; U. S. Patents 5,168,053 and 5,624,824), hairpin ribozyme structures (Berzal-Herranz *et al*, *Genes and Devel*, 6:129-134, 1992; Chowrira *et al*, *J. Biol. Chem.*, 269:25856-25864, 1994) and Hepatitis Delta virus based ribozymes (U. S. Patent 5,625,047). The general design and optimization of ribozyme directed RNA cleavage activity has been discussed in detail (Haseloff and Gerlach, 1988, *Nature*. 1988 Aug 18;334(6183):585-91, Chowrira *et al.*, *J. Biol. Chem.*, 269:25856-25864, 1994).

Another method of reducing protein expression utilizes the phenomenon of cosuppression or gene silencing (for example, U.S. Pat. Nos. 6,063,947; 5,686,649; or 5,283,184). Cosuppression of an endogenous gene using a full-length cDNA sequence as well as a partial cDNA sequence are known (for example, Napoli *et al.*, *Plant Cell* 2:279-289 [1990]; van der Krol *et al.*, *Plant Cell* 2:291-299 [1990]; Smith *et al.*, *Mol. Gen. Genetics* 224:477-481 [1990]). The phenomenon of cosuppression has also been used to inhibit plant target genes in a tissue-specific manner.

endogenous sequence. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Generally, where inhibition of expression is desired, some transcription of the introduced sequence occurs. The effect may occur where the introduced sequence contains no coding sequence per se, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence.

Yet another method of reducing protein activity is by expressing nucleic acid ligands, so-called aptamers, which specifically bind to the protein. Aptamers may be obtained by the SELEX (Systematic Evolution of Ligands by EXponential Enrichment) method. See U.S. Pat. No. 5,270,163. In the SELEX method, a candidate mixture of single stranded nucleic acids having regions of randomized sequence is contacted with the protein and those nucleic acids having an increased affinity to the target are selected and amplified. After several iterations a nucleic acid with optimal affinity to the polypeptide is obtained and is used for expression in modified plants.

A zinc finger protein that binds a polypeptide-encoding sequence or its regulatory region is also used to alter expression of the nucleotide sequence. Transcription of the nucleotide sequence may be reduced or increased. Zinc finger proteins are, for example, described in Beerli et al. (1998) PNAS 95:14628-14633., or in WO 95/19431, WO 98/54311, or WO 96/06166.

Other examples of non-protein expressing sequences specifically envisioned for use with the invention include tRNA sequences, for example, to alter codon usage, and rRNA variants, for example, which may confer resistance to various agents such as antibiotics.

It is contemplated that unexpressed DNA sequences, including novel synthetic sequences, could be introduced into cells as proprietary "labels" of those cells and plants and seeds thereof. It would not be necessary for a label DNA element to disrupt the function of a gene endogenous to the host organism, as the sole function of this DNA would be to identify the origin of the organism. For example, one could introduce a unique DNA sequence into a plant and this DNA element would identify all cells, plants, and progeny of these cells as having arisen from that labeled source. It is proposed that inclusion of label DNAs would enable one to distinguish proprietary germplasm or germplasm derived from such, from unlabelled germplasm.

10 **Exemplary plant promoters, regulatory sequences and targeting sequences**

Exemplary classes of plant promoters are described below.

Constitutive Expression promoters: Exemplary constitutive expression promoters include the ubiquitin promoter (e.g., sunflower--Binet et al. Plant Science 79: 87-94 (1991); maize--Christensen et al. Plant Molec. Biol. 12: 619-632 (1989); and Arabidopsis--Callis et al., J. Biol. Chem. 265: 12486-12493 (1990) and Norris et al., Plant Mol. Biol. 21: 895-906 (1993)); the CaMV 35S promoter (U.S. Patent Nos. 5,858,742 and 5,322,938); or the actin promoter (e.g., rice-- U.S. Pat. No. 5,641,876; McElroy et al. Plant Cell 2: 163-171 (1990), McElroy et al. Mol. Gen. Genet. 231: 150-160 (1991), and Chibbar et al. Plant Cell Rep. 12: 506-509 (1993)).

20 Inducible Expression promoters: Exemplary inducible expression promoters include the chemically regulatable tobacco PR-1 promoter (e.g., tobacco-- U.S. Pat. No. 5,614,395; Arabidopsis--Lebel et al., Plant J. 16: 223-233 (1998); maize- U.S. Pat. No. 6,429,362). Various chemical regulators may be employed to induce expression, including the benzothiadiazole, isonicotinic acid, and salicylic acid compounds disclosed in U.S. Pat. Nos. 5,523,311 and 5,614,395. Other promoters inducible by certain alcohols or ketones, such as ethanol, include, for example, the alcA gene promoter from *Aspergillus nidulans* (Caddick et al. (1998) Nat. Biotechnol 16:177-180). A glucocorticoid-mediated induction system is described in Aoyama and Chua (1997) The Plant Journal 11: 605-612 wherein gene expression is induced by application of a glucocorticoid, for example a dexamethasone. Another class of useful promoters are water-deficit-inducible promoters, e.g. promoters which are derived from the 5' regulatory region of genes identified as a heat shock protein 17.5 gene (HSP 17.5), an HVA22 gene (HVA22), and a cinnamic acid 4-hydroxylase

(CA4H) gene of *Zea mays*. Another water-deficit-inducible promoter is derived from the rab-17 promoter as disclosed by Vilardell et al., Plant Molecular Biology, 17(5):985-993, 1990. See also U.S. Pat. No. 6,084,089 which discloses cold inducible promoters, U.S. Pat. No. 6,294,714 which discloses light inducible promoters, U.S. Pat. No. 6,140,078 which discloses salt inducible promoters, U.S. Pat. No. 6,252,138 which discloses pathogen inducible promoters, and U.S. Pat. No. 6,175,060 which discloses phosphorus deficiency inducible promoters.

As another example, numerous wound-inducible promoters have been described (e.g. Xu et al. Plant Molec. Biol. 22: 573-588 (1993), Logemann et al. Plant Cell 1: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22: 783-792 (1993), Firek *et al.* Plant Molec. Biol. 22: 129-142 (1993), Warner et al. Plant J. 3: 191-201 (1993)). Logemann describe 5' upstream sequences of the potato wun1 gene. Xu et al. show that a wound-inducible promoter from the dicotyledon potato (pin2) is active in the monocotyledon rice. Rohrmeier & Lehle describe maize Wipl cDNA which is wound induced and which can be used to isolate the cognate promoter. Firek et al. and Warner et al. have described a wound-induced gene from the monocotyledon *Asparagus officinalis*, which is expressed at local wound and pathogen invasion sites.

Tissue-Specific Promoters: Exemplary promoters that express genes only in certain tissues are useful according to the present invention. For example root specific expression may be attained using the promoter of the maize metallothionein-like (MTL) gene described by de Framond (FEBS 290: 103-106 (1991)) and also in U.S. Pat. No. 5,466,785. U.S. Pat. No. 5,837,848 discloses a root specific promoter. Another exemplary promoter confers pith-preferred expression (see Int'l. Pub. No. WO 93/07278, which describes the maize trpA gene and promoter that is preferentially expressed in pith cells). Leaf-specific expression may be attained, for example, by using the promoter for a maize gene encoding phosphoenol carboxylase (PEPC) (see Hudspeth & Grula, Plant Molec Biol 12: 579-589 (1989)). Pollen-specific expression may be conferred by the promoter for the maize calcium-dependent protein kinase (CDPK) gene which is expressed in pollen cells (WO 93/07278). U.S. Pat. Appl. Pub. No. 20040016025 describes tissue-specific promoters. Pollen-specific expression may be conferred by the tomato LAT52 pollen-specific promoter (Bate et. al., Plan mol Biol. 1998 Jul;37(5):859-69).

See also U.S. Pat. No. 6,437,217 which discloses a root-specific maize RS81 promoter, U.S. Pat. No. 6,426,446 which discloses a root specific maize RS324 promoter, U.S. Pat. No. 6,232,526 which discloses a constitutive maize A3 promoter, U.S. Pat. No. 6,177,611 which discloses constitutive maize promoters, U.S. Pat. No. 6,433,252 which discloses a maize L3 oleosin promoter that are aleurone and seed coat-specific promoters, U.S. Pat. No. 6,429,357 which discloses a constitutive rice actin 2 promoter and intron, U.S. patent application Pub. No. 20040216189 which discloses an inducible constitutive leaf specific maize chloroplast aldolase promoter.

Optionally a plant transcriptional terminator can be used in place of the plant-expressed gene native transcriptional terminator. Exemplary transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the tml terminator, the nopaline synthase terminator and the pea rbcS E9 terminator. These can be used in both monocotyledons and dicotyledons.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize Adhl gene have been found to significantly enhance expression. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, Genes Develop. 1: 1183-1200 (1987)). The intron from the maize bronze1 gene also enhances expression. Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader. U.S. Patent Application Publication 2002/0192813 discloses 5', 3' and intron elements useful in the design of effective plant expression vectors.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV, the "omega-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g. Gallie *et al.* Nucl. Acids Res. 15: 8693-8711 (1987); Skuzeski *et al.* Plant Molec. Biol. 15: 65-79 (1990)). Other leader sequences known in the art include but are not limited to: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein, O., Fuerst, T. R., and Moss, B. PNAS USA 86:6126-

6130 (1989)); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison et al., 1986); MDMV leader (Maize Dwarf Mosaic Virus); Virology 154:9-20); human immunoglobulin heavy-chain binding protein (BiP) leader, (Macejak, D. G., and Sarnow, P., Nature 353: 90-94 (1991); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling, S. A., and Gehrke, L., Nature 325:622-625 (1987); tobacco mosaic virus leader (TMV), (Gallie et al., Molecular Biology of RNA, pages 237-256 (1989); or Maize Chlorotic Mottle Virus leader (MCMV) (Lommel et al., Virology 81:382-385 (1991). See also, Della-Cioppa et al., Plant Physiology 84:965-968 (1987).

10 A minimal promoter may also be incorporated. Such a promoter has low background activity in plants when there is no transactivator present or when enhancer or response element binding sites are absent. One exemplary minimal promoter is the Bz1 minimal promoter, which is obtained from the bronze1 gene of maize. Roth et al., Plant Cell 3: 317 (1991). A minimal promoter may also be created
15 by use of a synthetic TATA element. The TATA element allows recognition of the promoter by RNA polymerase factors and confers a basal level of gene expression in the absence of activation (see generally, Mukumoto (1993) Plant Mol Biol 23: 995-1003; Green (2000) Trends Biochem Sci 25: 59-63).

Sequences controlling the targeting of gene products also may be
20 included. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins which is cleaved during chloroplast import to yield the mature protein (e.g. Comai et al. J. Biol. Chem. 263: 15104-15109 (1988)). These signal sequences can be fused to heterologous gene products to effect the import of heterologous products into the
25 chloroplast (van den Broeck, et al. Nature 313: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the EPSP synthase enzyme, the GS2 protein or many other proteins which are known to be chloroplast localized. Other gene products are localized to other organelles such as the mitochondrion and the
30 peroxisome (e.g. Unger et al. Plant Molec. Biol. 13: 411-418 (1989)). Examples of sequences that target to such organelles are the nuclear-encoded ATPases or specific aspartate amino transferase isoforms for mitochondria. Targeting cellular protein bodies has been described by Rogers et al. (Proc. Natl. Acad. Sci. USA 82: 6512-6516

(1985)). In addition, amino terminal and carboxy-terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, Plant Cell 2: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for
5 vacuolar targeting of gene products (Shinshi et al. Plant Molec. Biol. 14: 357-368 (1990)).

Another possible element which may be introduced is a matrix attachment region element (MAR), such as the chicken lysozyme A element (Stief, 1989), which can be positioned around an expressible gene of interest to effect an
10 increase in overall expression of the gene and diminish position dependent effects upon incorporation into the plant genome (Stief et al., Nature, 341:343, 1989; Phi-Van et al., Mol. Cell. Biol., 10:2302-2307.1990).

Use of non-plant promoter regions isolated from *Drosophila melanogaster* and *Saccharomyces cerevisiae* to express genes in plants

15 The promoter in the mini-chromosome of the present invention can be derived from plant or non-plant species. In a preferred embodiment, the nucleotide sequence of the promoter is derived from non-plant species for the expression of genes in plant cells, including but not limited to dicotyledon plant cells such as tobacco, tomato, potato, soybean, canola, sunflower, alfalfa, cotton and *Arabidopsis*,
20 or monocotyledonous plant cell, such as wheat, maize, rye, rice, turf grass, oat, barley, sorghum, millet, and sugarcane. In one embodiment, the non-plant promoters are constitutive or inducible promoters derived from insect, e.g., *Drosophila melanogaster* or yeast, e.g., *Saccharomyces cerevisiae*. Table 2 lists the promoters from *Drosophila melanogaster* and *Saccharomyces cerevisiae* that are used to derive
25 the examples of non-plant promoters in the present invention. Promoters derived from any animal, protist, or fungi are also contemplated. SEQ ID NOS: 1-20 are examples of promoter sequences derived from *Drosophila melanogaster* or *Saccharomyces cerevisiae*. These non-plant promoters can be operably linked to nucleic acid sequences encoding polypeptides or non-protein-expressing sequences
30 including, but not limited to, antisense RNA and ribozymes, to form nucleic acid constructs, vectors, and host cells (prokaryotic or eukaryotic), comprising the promoters.

Table 2- *Drosophila melanogaster* Promoters (Information obtained from the Flybase Web Site at <http://flybase.bio.indiana.edu/> which is a database of the *Drosophila* Genome)

SEQ ID NO:	SSymbol	Flybase ID	Standard promoter gene name	Gene Product	Chromosome
1	Pgd	FBgn0004654	Phosphogluconate dehydrogenase	6-phosphogluconate dehydrogenase	X
2	Grim	FBgn0015946	grim	grim-P138	3
3	Uro	FBgn0003961	Urate oxidase	Uro-P1	2
4	Sna	FBgn0003448	snail	sna-P1	2
5	Rh3	FBgn0003249	Rhodopsin 3	Rh3	3
6	Lsp-1 γ	FBgn0002564	Larval serum protein 1 γ	Lsp1 γ -P1	3

***Saccharomyces cerevisiae* Promoters**

(Information obtained from the *Saccharomyces* Genome Database Web site at <http://www.yeastgenome.org/SearchContents.shtml>)

Seq No.	SSymbol	Systematic Name	Standard promoter gene name	Gene Product	Chromosome
7	Tef-2	YBR118W	TEF2 (Translation elongation factor promoter)	Translation elongation factor EF-1 α	2
8	Leu-1	YGL009C	LEU1 (LEUcine biosynthesis)	isopropylmalate isomerase	7
9	Met16	YPR167C	METHionine requiring	3'phosphoadenylylsulfate reductase	16
10	Leu-2	YCL018W	LEU2 (leucine biosynthesis)	beta-IPM (isopropylmalate) dehydrogenase	3
11	His-4	YCL030C	HIS4 (HISTidine requiring)	histidinol dehydrogenase	3
12	Met-2	YNL277W	MET2 (methionine requiring)	L-homoserine-O-acetyltransferase	14
13	Ste-3	YKL178C	STE3 (alias DAF2 Sterile)	a-factor receptor	11
14	Arg-1	YOL058W	ARG1 (alias ARG10 ARGinine requiring)	arginosuccinate synthetase	15
15	Pgk-1	YCR012W	PGK1 (phosphoglycerate kinase)	phosphoglycerate kinase	3
16	GPD-1	YDL022W	GPD1 (alias DAR1/HOR1/OSG OSR5: glycerol-3-phosphate dehydrogenase activity)	glycerol-3-phosphate dehydrogenase	4
17	ADH1	YOL086C	ADH1 (alias ADC1)	alcohol dehydrogenase	15

SEQ ID NO:	SSymbol	Flybase ID	Standard promoter gene name	Gene Product	Chromosome
18	GPD-2	YOL059W	GPD2 (alias GPD3: glycerol-3-phosphate dehydrogenase activity	glycerol-3-phosphate dehydrogenase	15
19	Arg-4	YHR018C	ARGinine requiring	argininosuccinate lyase	8
20	Yat-1	YAR035W	YAT-1(carnitine acetyltransferase)	carnitine acetyltransferase	1

The present invention relates to methods for producing a polypeptide, comprising cultivating plant material for the production of the polypeptide at any level, wherein the plant host cells comprises a first nucleic acid sequence encoding the polypeptide operably linked to a second nucleic acid sequence comprising a heterologous promoter foreign to the nucleic acid sequence, wherein the promoter comprises a sequence selected from the group consisting of SEQ ID NOS:1 to 20 or subsequences thereof; and mutant, hybrid, or tandem promoters thereof that retain promoter activity.

The present invention also relates to methods for producing non-protein expressed sequences, comprising cultivating plant material for the production of the non-protein expressed sequence, wherein the plant host cell comprises a first nucleic acid sequence encoding the non-protein expressed sequences operably linked to a second nucleic acid sequence comprising a heterologous promoter foreign to the nucleic acid sequence, wherein the promoter comprises a sequence selected from the group consisting of SEQ ID NOS: 1 to 20 or subsequences thereof; and mutant, hybrid, or tandem promoters thereof.

The present invention also relates to isolated promoter sequences and to constructs, vectors, or plant host cells comprising one or more of the promoters operably linked to a nucleic acid sequence encoding a polypeptide or non-protein expressing sequence.

In the methods of the present invention, the promoter may also be a mutant of the promoters having a substitution, deletion, and/or insertion of one or more nucleotides in the nucleic acid sequence of SEQ ID NOS: 1 to 20.

The present invention also relates to methods for obtaining derivative promoters of SEQ ID NOS: 1 to 20.

The techniques used to isolate or clone a nucleic acid sequence comprising a promoter of interest are known in the art and include isolation from genomic DNA. The cloning procedures may involve excision or amplification, for example by polymerase chain reaction, and isolation of a desired nucleic acid
5 fragment comprising the nucleic acid sequence encoding the promoter, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into the plant cell.

Definitions

The term “adchromosomal” plant or plant part as used herein means a
10 plant or plant part that contains functional, stable and autonomous mini-chromosomes. Adchromosomal plants or plant parts may be chimeric or not chimeric (chimeric meaning that mini-chromosomes are only in certain portions of the plant, and are not uniformly distributed throughout the plant). An adchromosomal plant cell contains at least one functional, stable and autonomous mini-chromosome.

15 The term “autonomous” as used herein means that when delivered to plant cells, at least some mini-chromosomes are transmitted through mitotic division to daughter cells and are episomal in the daughter plant cells, i.e. are not chromosomally integrated in the daughter plant cells. Daughter plant cells that contain autonomous mini-chromosomes can be selected for further replication using,
20 for example, selectable or screenable markers. During the introduction into a cell of a mini-chromosome, or during subsequent stages of the cell cycle, there may be chromosomal integration of some portion or all of the DNA derived from a mini-chromosome in some cells. The mini-chromosome is still characterized as autonomous despite the occurrence of such events if a plant may be regenerated that
25 contains episomal descendants of the mini-chromosome distributed throughout its parts, or if gametes or progeny can be derived from the plant that contain episomal descendants of the mini-chromosome distributed through its parts.

As used herein, a “centromere” is any DNA sequence that confers an ability to segregate to daughter cells through cell division. In one context, this
30 sequence may produce a transmission efficiency to daughter cells ranging from about 1% to about 100%, including to about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or about 95% of daughter cells. Variations in such a transmission efficiency may find important applications within the scope of the invention; for

example, mini-chromosomes carrying centromeres that confer 100% stability could be maintained in all daughter cells without selection, while those that confer 1% stability could be temporarily introduced into a transgenic organism, but be eliminated when desired. In particular embodiments of the invention, the centromere may confer stable transmission to daughter cells of a nucleic acid sequence, including a recombinant construct comprising the centromere, through mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A plant centromere is not necessarily derived from plants, but has the ability to promote DNA transmission to daughter plant cells.

10 As used herein, the term "circular permutations" refer to variants of a sequence that begin at base n within the sequence, proceed to the end of the sequence, resume with base number one of the sequence, and proceed to base n - 1. For this analysis, n may be any number less than or equal to the length of the sequence. For example, circular permutations of the sequence ABCD are: ABCD, BCDA, CDAB, 15 and DABC.

The term "co-delivery" as used herein refers to the delivery of two nucleic acid segments to a cell. In co-delivery of plant growth inducing genes and mini-chromosomes, the two nucleic acid segments are delivered simultaneously using the same delivery method. Alternatively, the nucleic acid segment containing the growth inducing gene, optionally as part of an episomal vector, such as a viral vector or a plasmid vector, may be delivered to the plant cells before or after delivery of the mini-chromosome, and the mini-chromosome may carry an exogenous nucleic acid that induces expression of the earlier-delivered growth inducing gene. In this embodiment, the two nucleic acid segments may be delivered separately at different times provided the encoded growth inducing factors are functional during the appropriate time period.

The term "coding sequence" is defined herein as a nucleic acid sequence that is transcribed into mRNA which is translated into a polypeptide when placed under the control of promoter sequences. The boundaries of the coding sequence are generally determined by the ATG start codon located at the start of the open reading frame, near the 5' end of the mRNA, and TAG, TGA or TAA stop codons at the end of the coding sequence, near the 3' end of the mRNA, and in some

cases, a transcription terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, genomic DNA, cDNA, semisynthetic, synthetic, or recombinant nucleic acid sequences.

5 As used herein the term "consensus" refers to a nucleic acid sequence derived by comparing two or more related sequences. A consensus sequence defines both the conserved and variable sites between the sequences being compared. Any one of the sequences used to derive the consensus or any permutation defined by the consensus may be useful in construction of mini-chromosomes.

10 The term "exogenous" when used in reference to a nucleic acid, for example, is intended to refer to any nucleic acid that has been introduced into a recipient cell, regardless of whether the same or similar nucleic acid is already present in such a cell. Thus, as an example, "exogenous DNA" can include an additional copy of DNA that is already present in the plant cell, DNA from another plant, DNA from a
15 different organism, or a DNA generated externally, such as a DNA sequence containing an antisense message of a gene, or a DNA sequence encoding a synthetic or modified version of a gene. An "exogenous gene" can be a gene not normally found in the host genome in an identical context, or an extra copy of a host gene. The gene may be isolated from a different species than that of the host genome, or
20 alternatively, isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene.

 The term "functional" as used herein to describe a mini-chromosome means that when an exogenous nucleic acid is present within the mini-chromosome the exogenous nucleic acid can function in a detectable manner when the mini-
25 chromosome is within a plant cell; exemplary functions of the exogenous nucleic acid include transcription of the exogenous nucleic acid, expression of the exogenous nucleic acid, regulatory control of expression of other exogenous nucleic acids, recognition by a restriction enzyme or other endonuclease, ribozyme or recombinase; providing a substrate for DNA methylation, DNA glycolation or other DNA chemical
30 modification; binding to proteins such as histones, helix-loop-helix proteins, zinc binding proteins, leucine zipper proteins, MADS box proteins, topoisomerases, helicases, transposases, TATA box binding proteins, viral protein, reverse

transcriptases; or cohesins; providing an integration site for homologous recombination; providing an integration site for a transposon, T-DNA or retrovirus; providing a substrate for RNAi synthesis; priming of DNA replication; aptamer binding; or kinetochore binding. If multiple exogenous nucleic acids are present
5 within the mini-chromosome, the function of one or preferably more of the exogenous nucleic acids can be detected under suitable conditions permitting function thereof.

As used herein, a "library" is a pool of cloned DNA fragments that represents some or all DNA sequences collected, prepared or purified from a specific source. Each library may contain the DNA of a given organism inserted as discrete
10 restriction enzyme generated fragments or as randomly sheared fragments into many thousands of plasmid vectors. For purposes of the present invention, *E. coli*, yeast, and *Salmonella* plasmids are particularly useful for propagating the genome inserts from other organisms. In principle, any gene or sequence present in the starting DNA preparation can be isolated by screening the library with a specific hybridization
15 probe (see, for example, Young et al., In: Eukaryotic Genetic Systems ICN-UCLA Symposia on Molecular and Cellular Biology, VII, 315-331, 1977).

As used herein, the term "linker" refers to a DNA molecule, generally up to 50 or 60 nucleotides long and composed of two or more complementary oligonucleotides that have been synthesized chemically, or excised or amplified from
20 existing plasmids or vectors. In a preferred embodiment, this fragment contains one, or preferably more than one, restriction enzyme site for a blunt cutting enzyme and/or a staggered cutting enzyme, such as BamHI. One end of the linker is designed to be ligatable to one end of a linear DNA molecule and the other end is designed to be ligatable to the other end of the linear molecule, or both ends may be designed to be
25 ligatable to both ends of the linear DNA molecule.

As used herein, a "mini-chromosome" is a recombinant DNA construct including a centromere and capable of transmission to daughter cells. A mini-chromosome may remain separate from the host genome (as episomes) or may integrate into host chromosomes. The stability of this construct through cell division
30 could range between from about 1% to about 100%, including about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and about 95%. The mini-chromosome construct may be a circular or linear molecule. It may include elements such as one

or more telomeres, origin of replication sequences, stuffer sequences, buffer sequences, chromatin packaging sequences, linkers and genes. The number of such sequences included is only limited by the physical size limitations of the construct itself. It could contain DNA derived from a natural centromere, although it may be preferable to limit the amount of DNA to the minimal amount required to obtain a transmission efficiency in the range of 1-100%. The mini-chromosome could also contain a synthetic centromere composed of tandem arrays of repeats of any sequence, either derived from a natural centromere, or of synthetic DNA. The mini-chromosome could also contain DNA derived from multiple natural centromeres. The mini-chromosome may be inherited through mitosis or meiosis, or through both meiosis and mitosis. As used herein, the term mini-chromosome specifically encompasses and includes the terms "plant artificial chromosome" or "PLAC," or engineered chromosomes or microchromosomes and all teachings relevant to a PLAC or plant artificial chromosome specifically apply to constructs within the meaning of the term mini-chromosome.

The term "non-protein expressing sequence" or "non-protein coding sequence" is defined herein as a nucleic acid sequence that is not eventually translated into protein. The nucleic acid may or may not be transcribed into RNA. Exemplary sequences include ribozymes or antisense RNA.

The term "operably linked" is defined herein as a configuration in which a control sequence, e.g., a promoter sequence, directs transcription or translation of another sequence, for example a coding sequence. For example, a promoter sequence could be appropriately placed at a position relative to a coding sequence such that the control sequence directs the production of a polypeptide encoded by the coding sequence.

"Phenotype" or "phenotypic trait(s)", as used herein, refers to an observable property or set of properties resulting from the expression of a gene. The set of properties may be observed visually or after biological or biochemical testing, and may be constantly present or may only manifest upon challenge with the appropriate stimulus or activation with the appropriate signal.

The term "plant," as used herein, refers to any type of plant. Exemplary types of plants are listed below, but other types of plants will be known to

those of skill in the art and could be used with the invention. Modified plants of the invention include, for example, dicots, gymnosperm, monocots, mosses, ferns, horsetails, club mosses, liver worts, hornworts, red algae, brown algae, gametophytes and sporophytes of pteridophytes, and green algae.

5 The term “crop plant” refers to plants grown for agricultural or commercial rather than experimental purposes and specifically excludes *Arabidopsis thaliana*. Some plants grown for experimental purposes may take on commercial importance when used to produce pharmaceutical or chemical products. Centromeres “derived from crop plants” according to the present invention specifically exclude
10 centromeres that are fragments of naturally occurring *Arabidopsis thaliana* centromeres or naturally occurring descendants thereof. Centromeres derived from crop plants include variants (mutants) of *Arabidopsis thaliana* centromeres, or artificial centromeres synthesized based on nucleotide sequences of *Arabidopsis thaliana* centromeres.

15 A common class of plants exploited in agriculture are vegetable crops, including artichokes, kohlrabi, arugula, leeks, asparagus, lettuce (*e.g.*, head, leaf, romaine), bok choy, malanga, broccoli, melons (*e.g.*, muskmelon, watermelon, crenshaw, honeydew, cantaloupe), brussels sprouts, cabbage, cardoni, carrots, napa, cauliflower, okra, onions, celery, parsley, chick peas, parsnips, chicory, chinese
20 cabbage, peppers, collards, potatoes, cucumber plants (marrows, cucumbers), pumpkins, cucurbits, radishes, dry bulb onions, rutabaga, eggplant, salsify, escarole, shallots, endive, garlic, spinach, green onions, squash, greens, beet (sugar beet or fodder beet), sweet potatoes, swiss chard, horseradish, tomatoes, kale, turnips, or spices.

25 Other types of plants frequently finding commercial use include fruit and vine crops such as apples, grapes, apricots, cherries, nectarines, peaches, pears, plums, prunes, quince, almonds, chestnuts, filberts, pecans, pistachios, walnuts, citrus, blueberries, boysenberries, cranberries, currants, loganberries, raspberries, strawberries, blackberries, grapes, avocados, bananas, kiwi, persimmons,
30 pomegranate, pineapple, tropical fruits, pomes, melon, mango, papaya, or lychee.

Modified wood and fiber or pulp plants of particular interest include, but are not limited to maple, oak, cherry, mahogany, poplar, aspen, birch, beech,

spruce, fir, kenaf, pine, walnut, cedar, redwood, chestnut, acacia, bombax, alder, eucalyptus, catalpa, mulberry, persimmon, ash, honeylocust, sweetgum, privet, sycamore, magnolia, sourwood, cottonwood, mesquite, buckthorn, locust, willow, elderberry, teak, linden, bubinga, basswood or elm.

5 Modified flowers and ornamental plants of particular interest, include, but are not limited to, roses, petunias, pansy, peony, olive, begonias, violets, phlox, nasturtiums, irises, lilies, orchids, vinca, philodendron, poinsettias, opuntia, cyclamen, magnolia, dogwood, azalea, redbud, boxwood, Viburnum, maple, elderberry, hosta, agave, asters, sunflower, pansies, hibiscus, morning glory, alstromeria, zinnia,
10 geranium, Prosopis, artemesia, clematis, delphinium, dianthus, gallium, coreopsis, iberis, lamium, poppy, lavender, leucophyllum, sedum, salvia, verbascum, digitalis, penstemon, savory, pyrethrum, or oenothera. Modified nut-bearing trees of particular interest include, but are not limited to pecans, walnuts, macadamia nuts, hazelnuts, almonds, or pistachios, cashews, pignolas or chestnuts.

15 Many of the most widely grown plants are field crop plants such as evening primrose, meadow foam, corn (field, sweet, popcorn), hops, jojoba, peanuts, rice, safflower, small grains (barley, oats, rye, wheat, etc.), sorghum, tobacco, kapok, leguminous plants (beans, lentils, peas, soybeans), oil plants (rape, mustard, poppy, olives, sunflowers, coconut, castor oil plants, cocoa beans, groundnuts, oil palms),
20 fibre plants (cotton, flax, hemp, jute), lauraceae (cinnamon, camphor), or plants such as coffee, sugarcane, cocoa, tea, or natural rubber plants.

 Still other examples of plants include bedding plants such as flowers, cactus, succulents or ornamental plants, as well as trees such as forest (broad-leaved trees or evergreens, such as conifers), fruit, ornamental, or nut-bearing trees, as well
25 as shrubs or other nursery stock.

 Modified crop plants of particular interest in the present invention include, but are not limited to, soybean (*Glycine max*), cotton, canola (also known as rape), wheat, sunflower, sorghum, alfalfa, barley, safflower, millet, rice, tobacco, fruit and vegetable crops or turfgrasses. Exemplary cereals include maize, wheat, barley,
30 oats, rye, millet, sorghum, rice tritcale, secale, einkorn, spelt, emmer, teff, milo, flax, gramma grass, *Tripsacum* sp., or teosinte. Oil-producing plants include plant species that produce and store triacylglycerol in specific organs, primarily in seeds. Such

species include soybean (*Glycine max*), rapeseed or canola (including *Brassica napus*, *Brassica rapa* or *Brassica campestris*), *Brassica juncea*, *Brassica carinata*, sunflower (*Helianthus annuus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), cocoa (*Theobroma cacao*), safflower (*Carthamus tinctorius*), oil palm (*Elaeis guineensis*),
5 coconut palm (*Cocos nucifera*), flax (*Linum usitatissimum*), castor (*Ricinus communis*) or peanut (*Arachis hypogaea*).

The term "plant part" as used herein includes pollen, silk, endosperm, ovule, seed, embryo, pods, roots, cuttings, tubers, stems, stalks, fruit, berries, nuts, flowers, leaves, bark, wood, whole plant, plant cell, plant organ, epidermis, vascular
10 tissue, protoplast, cell culture, crown, callus culture, petiole, petal, sepal, stamen, stigma, style, bud, meristem, cambium, cortex, pith, sheath or any group of plant cells organized into a structural and functional unit. In one preferred embodiment, the exogenous nucleic acid is expressed in a specific location or tissue of a plant, for example, epidermis, vascular tissue, meristem, cambium, cortex, pith, leaf, sheath,
15 flower, root or seed.

The term "promoter" is defined herein as a DNA sequence that allows the binding of RNA polymerase (including but not limited to RNA polymerase I, RNA polymerase II and RNA polymerase III from eukaryotes) and directs the polymerase to a downstream transcriptional start site of a nucleic acid sequence
20 encoding a polypeptide to initiate transcription. RNA polymerase effectively catalyzes the assembly of messenger RNA complementary to the appropriate DNA strand of the coding region.

A "promoter operably linked to a heterologous gene" is a promoter that is operably linked to a gene that is different from the gene to which the promoter is
25 normally operably linked in its native state. Similarly, an "exogenous nucleic acid operably linked to a heterologous regulatory sequence" is a nucleic acid that is operably linked to a regulatory control sequence to which it is not normally linked in its native state.

The term "hybrid promoter" is defined herein as parts of two or more
30 promoters that are fused together to generate a sequence that is a fusion of the two or more promoters, which is operably linked to a coding sequence and mediates the transcription of the coding sequence into mRNA.

The term "tandem promoter" is defined herein as two or more promoter sequences each of which is operably linked to a coding sequence and mediates the transcription of the coding sequence into mRNA.

5 The term "constitutive active promoter" is defined herein as a promoter that allows permanent stable expression of the gene of interest.

The term "Inducible promoter" is defined herein as a promoter induced by the presence or absence of biotic or an abiotic factor.

The term "polypeptide" does not refer to a specific length of the encoded product and, therefore, encompasses peptides, oligopeptides, and proteins.

10 The term "exogenous polypeptide" is defined as a polypeptide which is not native to the plant cell, a native polypeptide in which modifications have been made to alter the native sequence, or a native polypeptide whose expression is quantitatively altered as a result of a manipulation of the plant cell by recombinant DNA techniques.

As used herein, the term "pseudogene" refers to a non-functional copy

15 of a protein-coding gene; pseudogenes found in the genomes of eukaryotic organisms are often inactivated by mutations and are thus presumed to be non-essential to that organism; pseudogenes of reverse transcriptase and other open reading frames found in retroelements are abundant in the centromeric regions of *Arabidopsis* and other organisms and are often present in complex clusters of related sequences.

20 As used herein the term "regulatory sequence" refers to any DNA sequence that influences the efficiency of transcription or translation of any gene. The term includes, but is not limited to, sequences comprising promoters, enhancers and terminators.

As used herein the term "repeated nucleotide sequence" refers to any

25 nucleic acid sequence of at least 25 bp present in a genome or a recombinant molecule, other than a telomere repeat, that occurs at least two or more times and that are preferably at least 80% identical either in head to tail or head to head orientation either with or without intervening sequence between repeat units.

As used herein, the term "retroelement" or "retrotransposon" refers to a

30 genetic element related to retroviruses that disperse through an RNA stage; the abundant retroelements present in plant genomes contain long terminal repeats (LTR

retrotransposons) and encode a polyprotein gene that is processed into several proteins including a reverse transcriptase. Specific retroelements (complete or partial sequences) can be found in and around plant centromeres and can be present as dispersed copies or complex repeat clusters. Individual copies of retroelements may
5 be truncated or contain mutations; intact retroelements are rarely encountered.

As used herein the term "satellite DNA" refers to short DNA sequences (typically < 1000 bp) present in a genome as multiple repeats, mostly arranged in a tandemly repeated fashion, as opposed to a dispersed fashion. Repetitive arrays of specific satellite repeats are abundant in the centromeres of many higher eukaryotic
10 organisms.

As used herein, a "screenable marker" is a gene whose presence results in an identifiable phenotype. This phenotype may be observable under standard conditions, altered conditions such as elevated temperature, or in the presence of certain chemicals used to detect the phenotype. The use of a screenable marker
15 allows for the use of lower, sub-killing antibiotic concentrations and the use of a visible marker gene to identify clusters of transformed cells, and then manipulation of these cells to homogeneity. Preferred screenable markers of the present include genes that encode fluorescent proteins that are detectable by a visual microscope such as the fluorescent reporter genes DsRed, ZsGreen, ZsYellow, AmCyan, Green Fluorescent
20 Protein (GFP). An additional preferred screenable marker gene is *lac*.

The invention also contemplates novel methods of screening for adchromosomal plant cells that involve use of relatively low, sub-killing concentrations of a selection agent (e.g. sub-killing antibiotic concentrations), and also involve use of a screenable marker (e.g., a visible marker gene) to identify
25 clusters of modified cells carrying the screenable marker, after which these screenable cells are manipulated to homogeneity. As used herein, a "selectable marker" is a gene whose presence results in a clear phenotype, and most often a growth advantage for cells that contain the marker. This growth advantage may be present under standard conditions, altered conditions such as elevated temperature, specialized media
30 compositions, or in the presence of certain chemicals such as herbicides or antibiotics. Use of selectable markers is described, for example, in Broach et al. Gene, 8:121-133, 1979. Examples of selectable markers include the thymidine kinase gene, the cellular

adenine phosphoribosyltransferase gene and the dihydrolfolate reductase gene, hygromycin phosphotransferase genes, the bar gene, neomycin phosphotransferase genes and phosphomannose isomerase, among others. Preferred selectable markers in the present invention include genes whose expression confer antibiotic or herbicide resistance to the host cell, or proteins allowing utilization of a carbon source not normally utilized by plant cells. Expression of one of these markers should be sufficient to enable the maintenance of a vector within the host cell, and facilitate the manipulation of the plasmid into new host cells. Of particular interest in the present invention are proteins conferring cellular resistance to kanamycin, G 418, paramomycin, hygromycin, bialaphos, and glyphosate for example, or proteins allowing utilization of a carbon source, such as mannose, not normally utilized by plant cells.

The term "stable" as used herein means that the mini-chromosome can be transmitted to daughter cells over at least 8 mitotic generations. Some embodiments of mini-chromosomes may be transmitted as functional, autonomous units for less than 8 mitotic generations, *e.g.* 1, 2, 3, 4, 5, 6, or 7. Preferred mini-chromosomes can be transmitted over at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 generations, for example, through the regeneration or differentiation of an entire plant, and preferably are transmitted through meiotic division to gametes. Other preferred mini-chromosomes can be further maintained in the zygote derived from such a gamete or in an embryo or endosperm derived from one or more such gametes. A "functional and stable" mini-chromosome is one in which functional mini-chromosomes can be detected after transmission of the mini-chromosomes over at least 8 mitotic generations, or after inheritance through a meiotic division. During mitotic division, as occurs occasionally with native chromosomes, there may be some non-transmission of mini-chromosomes; the mini-chromosome may still be characterized as stable despite the occurrence of such events if an adchromosomal plant that contains descendants of the mini-chromosome distributed throughout its parts may be regenerated from cells, cuttings, propagules, or cell cultures containing the mini-chromosome, or if an adchromosomal plant can be identified in progeny of the plant containing the mini-chromosome.

As used herein, a "structural gene" is a sequence which codes for a polypeptide or RNA and includes 5' and 3' ends. The structural gene may be from the host into which the structural gene is transformed or from another species. A structural gene will preferably, but not necessarily, include one or more regulatory sequences which modulate the expression of the structural gene, such as a promoter, terminator or enhancer. A structural gene will preferably, but not necessarily, confer some useful phenotype upon an organism comprising the structural gene, for example, herbicide resistance. In one embodiment of the invention, a structural gene may encode an RNA sequence which is not translated into a protein, for example a tRNA or rRNA gene.

As used herein, the term "telomere" or "telomere DNA" refers to a sequence capable of capping the ends of a chromosome, thereby preventing degradation of the chromosome end, ensuring replication and preventing fusion to other chromosome sequences. Telomeres can include naturally occurring telomere sequences or synthetic sequences. Telomeres from one species may confer telomere activity in another species. An exemplary telomere DNA is a heptanucleotide telomere repeat TTTAGGG (and its complement) found in the majority of plants.

"Transformed," "transgenic," "modified," and "recombinant" refer to a host organism such as a plant into which an exogenous or heterologous nucleic acid molecule has been introduced, and includes meiocytes, seeds, zygotes, embryos, endosperm, or progeny of such plant that retain the exogenous or heterologous nucleic acid molecule but which have not themselves been subjected to the transformation process.

When the phrase "transmission efficiency" of a certain percent is used, transmission percent efficiency is calculated by measuring mini-chromosome presence through one or more mitotic or meiotic generations. It is directly measured as the ratio (expressed as a percentage) of the daughter cells or plants demonstrating presence of the mini-chromosome to parental cells or plants demonstrating presence of the mini-chromosome. Presence of the mini-chromosome in parental and daughter cells is demonstrated with assays that detect the presence of an exogenous nucleic acid carried on the mini-chromosome. Exemplary assays can be the detection of a screenable marker (e.g. presence of a fluorescent protein or any gene whose

expression results in an observable phenotype), a selectable marker, or PCR amplification of any exogenous nucleic acid carried on the mini-chromosome.

Constructing mini-chromosomes by site-specific recombination

Plant mini-chromosomes may be constructed using site-specific recombination sequences (for example those recognized by the bacteriophage P1 Cre recombinase, or the bacteriophage lambda integrase, or similar recombination enzymes). A compatible recombination site, or a pair of such sites, is present on both the centromere containing DNA clones and the donor DNA clones. Incubation of the donor clone and the centromere clone in the presence of the recombinase enzyme causes strand exchange to occur between the recombination sites in the two plasmids; the resulting mini-chromosomes contain centromere sequences as well as mini-chromosome vector sequences. The DNA molecules formed in such recombination reactions is introduced into *E. coli*, other bacteria, yeast or plant cells by common methods in the field including, but not limited to, heat shock, chemical transformation, electroporation, particle bombardment, whiskers, or other transformation methods followed by selection for marker genes including chemical, enzymatic, color, or other marker present on either parental plasmid, allowing for the selection of transformants harboring mini-chromosomes.

II. Methods of detecting and characterizing mini-chromosomes in plant cells or of scoring mini-chromosome performance in plant cells:

Identification of candidate centromere fragments by probing BAC libraries

Centromere clones are identified from a large genomic insert library such as a Bacterial Artificial Chromosome library. Probes are labeled using nick-translation in the presence of radioactively labeled dCTP, dATP, dGTP or dTTP as in, for example, the commercially available Rediprime kit (Amersham) as per the manufacturer's instructions. Other labeling methods familiar to those skilled in the art could be substituted. The libraries are screened and deconvoluted. Genomic clones are screened by probing with small centromere-specific clones. Other embodiments of this procedure would involve hybridizing a library with other centromere sequences. Of the BAC clones identified using this procedure, a representative set are identified as having high hybridization signals to some probes, and optionally low hybridization signals to other probes. These are selected, the bacterial clones grown up in cultures and DNA prepared by methods familiar to those skilled in the art such

as alkaline lysis. The DNA composition of purified clones is surveyed using for example fingerprinting by digesting with restriction enzymes such as, but not limited to, HinfI or HindIII. In a preferred embodiment the restriction enzyme cuts within the tandem centromere satellite repeat (see below). A variety of clones showing different fingerprints are selected for conversion into mini-chromosomes and inheritance testing. It can also be informative to use multiple restriction enzymes for fingerprinting or other enzymes which can cleave DNA.

Fingerprinting analysis of BACs and mini-chromosomes

Centromere function may be associated with large tandem arrays of satellite repeats. To assess the composition and architecture of the centromere BACs, the candidate BACs are digested with a restriction enzyme, such as HindIII, which cuts with known frequency within the consensus sequence of the unit repeat of the tandemly repeated centromere satellite. Digestion products are then separated by agarose gel electrophoresis. Large insert clones containing a large array of tandem repeats will produce a strong band of the unit repeat size, as well as less intense bands at 2x and 3x the unit repeat size, and further multiples of the repeat size. These methods are well-known and there are many possible variations known to those skilled in the art.

Determining sequence composition of mini-chromosomes by shotgun cloning/sequencing, sequence analysis

To determine the sequence composition of the mini-chromosome, the insert is sequenced. To generate DNA suitable for sequencing mini-chromosomes are fragmented, for example by using a random shearing method (such as sonication, nebulization, etc). Other fragmentation techniques may also be used such as enzymatic digestion. These fragments are then cloned into a plasmid vector and sequenced. The resulting DNA sequence is trimmed of poor-quality sequence and of sequence corresponding to the plasmid vector. The sequence is then compared to the known DNA sequences using an algorithm such as BLAST to search a sequence database such as GenBank.

To determine the consensus of the satellite repeat in the mini-chromosome, the sequences containing satellite repeat are aligned using a DNA sequence alignment program such as ContigExpress from Vector NTI. The sequences may also be aligned to previously determined repeats for that species. The sequences

are trimmed to unit repeat length using the consensus as a template. Sequences trimmed from the ends of the alignment are realigned with the consensus and further trimmed until all sequences are at or below the consensus length. The sequences are then aligned with each other. The consensus is determined by the frequency of a specific nucleotide at each position; if the most frequent base is three times more frequent than the next most frequent base, it was considered the consensus.

Methods for determining consensus sequence are well known in the art, see, e.g., U.S. Pat. App. Pub. No. 20030124561; Hall & Preuss (2002). These methods, including DNA sequencing, assembly, and analysis, are well-known and there are many possible variations known to those skilled in the art. Other alignment parameters may also be useful such as using more or less stringent definitions of consensus.

Non-selective mini-chromosome mitotic inheritance assays

The following list of assays and potential outcomes illustrates how various assays can be used to distinguish autonomous events from integrated events.

Assay #1: transient assay

Mini-chromosomes are tested for their ability to become established as chromosomes and their ability to be inherited in mitotic cell divisions. In this assay, mini-chromosomes are delivered to plant cells, for example suspension cells in liquid culture. The cells used can be at various stages of growth. In this example, a population in which some cells were undergoing division was used. The mini-chromosome is then assessed over the course of several cell divisions, by tracking the presence of a screenable marker, e.g. a visible marker gene such as a fluorescent protein. Mini-chromosomes that are inherited well may show an initial delivery into many single cells; after several cell divisions, these single cells divide to form clusters of mini-chromosome-containing cells. Other exemplary embodiments of this method include delivering mini-chromosomes to other mitotic cell types, including roots and shoot meristems.

Assay #2: Non-lineage based inheritance assays on modified transformed cells and plants

Mini-chromosome inheritance is assessed on modified cell lines and plants by following the presence of the mini-chromosome over the course of multiple cell divisions. An initial population of mini-chromosome containing cells is assayed

for the presence of the mini-chromosome, by the presence of a marker gene, including but not limited to a fluorescent protein, a colored protein, a protein assayable by histochemical assay, and a gene affecting cell morphology. All nuclei are stained with a DNA-specific dye including but not limited to DAPI, Hoechst 33258, 5 OliGreen, Giemsa YOYO, or TOTO, allowing a determination of the number of cells that do not contain the mini-chromosome. After the initial determination of the percent of cells carrying the mini-chromosome, the cells are allowed to divide over the course of several cell divisions. The number of cell divisions, n , is determined by a method including but not limited to monitoring the change in total weight of cells, 10 and monitoring the change in volume of the cells or by directly counting cells in an aliquot of the culture. After a number of cell divisions, the population of cells is again assayed for the presence of the mini-chromosome. The loss rate per generation is calculated by the equation:

$$\text{Loss rate per generation} = 1 - (F/I)^{1/n}$$

15 The population of mini-chromosome-containing cells may include suspension cells, roots, leaves, meristems, flowers, or any other tissue of modified plants, or any other cell type containing a mini-chromosome.

These methods are well-known and there are many possible variations known to those skilled in the art; they have been used before with human cells and 20 yeast cells.

Assay #3: Lineage based inheritance assays on modified cells and plants

Mini-chromosome inheritance is assessed on modified cell lines and plants by following the presence of the mini-chromosome over the course of multiple cell divisions. In cell types that allow for tracking of cell lineage, including but not 25 limited to root cell files, trichomes, and leaf stomata guard cells, mini-chromosome loss per generation does not need to be determined statistically over a population, it can be discerned directly through successive cell divisions. In other manifestations of this method, cell lineage can be discerned from cell position, or methods including but not limited to the use of histological lineage tracing dyes, and the induction of 30 genetic mosaics in dividing cells.

In one simple example, the two guard cells of the stomata are daughters of a single precursor cell. To assay mini-chromosome inheritance in this

cell type, the epidermis of the leaf of a plant containing a mini-chromosome is examined for the presence of the mini-chromosome by the presence of a marker gene, including but not limited to a fluorescent protein, a colored protein, a protein assayable by histochemical assay, and a gene affecting cell morphology. The number of loss events in which one guard cell contains the mini-chromosome (L) and the number of cell divisions in which both guard cells contain the mini-chromosome (B) are counted. The loss rate per cell division is determined as $L/(L+B)$. Other lineage-based cell types are assayed in similar fashion. These methods are well-known and there are many possible variations known to those skilled in the art; they have been used before with yeast cells.

Lineal mini-chromosome inheritance may also be assessed by examining root files or clustered cells in callus over time. Changes in the percent of cells carrying the mini-chromosome will indicate the mitotic inheritance.

Assay #4: Inheritance assays on modified cells and plants in the presence of chromosome loss agents

Any of the above three assays can be done in the presence of chromosome loss agents (including but not limited to colchicine, colcemid, caffeine, etoposide, nocodazole, oryzalin, trifluran). It is likely that an autonomous mini-chromosome will prove more susceptible to loss induced by chromosome loss agents; therefore, autonomous mini-chromosomes should show a lower rate of inheritance in the presence of chromosome loss agents. These methods have been used to study chromosome loss in fruit flies and yeast; there are many possible variations known to those skilled in the art..

III. Transformation of plant cells and plant regeneration

Various methods may be used to deliver DNA into plant cells. These include biological methods, such as *Agrobacterium*, *E. coli*, and viruses, physical methods such as biolistic particle bombardment, nanocopy device, the Stein beam gun, silicon carbide whiskers and microinjection, electrical methods such as electroporation, and chemical methods such as the use of poly-ethylene glycol and other compounds known to stimulate DNA uptake into cells. Examples of these techniques are described by Paszkowski et al., EMBO J 3: 2717-2722 (1984), Potrykus et al., Mol. Gen. Genet. 199: 169-177 (1985), Reich et al., Biotechnology 4:

1001-1004 (1986), and Klein et al., Nature 327: 70-73 (1987). Transformation using silicon carbide whiskers, e.g. in maize, is described in Brisibe, J. Exp. Bot. 51(343):187-196 (2000) and Dunwell, Methods Mol. Biol. 111:375-82 (1999) and U.S. Patent No. 5,464,765.

5

Agrobacterium-mediated delivery

Agrobacterium-mediated transformation is one method for introducing a desired genetic element into a plant. Several *Agrobacterium* species mediate the transfer of a specific DNA known as "T-DNA" that can be genetically engineered to carry a desired piece of DNA into many plant species. Plasmids used for delivery contain the T-DNA flanking the nucleic acid to be inserted into the plant. The major events marking the process of T-DNA mediated pathogenesis are induction of virulence genes, processing and transfer of T-DNA.

There are three common methods to transform plant cells with *Agrobacterium*. The first method is co-cultivation of *Agrobacterium* with cultured isolated protoplasts. This method requires an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts. The second method is transformation of cells or tissues with *Agrobacterium*. This method requires (a) that the plant cells or tissues can be modified by *Agrobacterium* and (b) that the modified cells or tissues can be induced to regenerate into whole plants. The third method is transformation of seeds, apices or meristems with *Agrobacterium*. This method requires exposure of the meristematic cells of these tissues to *Agrobacterium* and micropropagation of the shoots or plan organs arising from these meristematic cells.

Those of skill in the art are familiar with procedures for growth and suitable culture conditions for *Agrobacterium* as well as subsequent inoculation procedures. Liquid or semi-solid culture media can be used. The density of the *Agrobacterium* culture used for inoculation and the ratio of *Agrobacterium* cells to explant can vary from one system to the next, as can media, growth procedures, timing and lighting conditions.

Transformation of dicotyledons using *Agrobacterium* has long been known in the art, and transformation of monocotyledons using *Agrobacterium* has

also been described. See, WO 94/00977 and U.S. Pat. No. 5,591,616. See also, Negrotto et al., Plant Cell Reports 19: 798-803 (2000).

A number of wild-type and disarmed strains of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* harboring Ti or Ri plasmids can be used for gene transfer into plants. Preferably, the *Agrobacterium* hosts contain disarmed Ti and Ri plasmids that do not contain the oncogenes that cause tumorigenesis or rhizogenesis. Exemplary strains include *Agrobacterium tumefaciens* strain C58, a nopaline-type strain that is used to mediate the transfer of DNA into a plant cell, octopine-type strains such as LBA4404 or succinamopine-type strains, e.g., EHA101 or EHA105. The use of these strains for plant transformation has been reported and the methods are familiar to those of skill in the art. U.S. Application No. 20040244075 published December 2, 2004 describes improved methods of *Agrobacterium*-mediated transformation. The efficiency of transformation by *Agrobacterium* may be enhanced by using a number of methods known in the art. For example, the inclusion of a natural wound response molecule such as acetosyringone (AS) to the *Agrobacterium* culture has been shown to enhance transformation efficiency with *Agrobacterium tumefaciens* (Shahla et al., (1987) Plant Molec. Biol. 8:291-298). Alternatively, transformation efficiency may be enhanced by wounding the target tissue to be modified or transformed. Wounding of plant tissue may be achieved, for example, by punching, maceration, bombardment with microprojectiles, etc. (See e.g., Bidney et al., (1992) Plant Molec. Biol. 18:301-313).

In addition, a recent method described by Broothaerts, et. al. (Nature 433: 629-633, 2005) expands the bacterial genera that can be used to transfer genes into plants. This work involved the transfer of a disarmed Ti plasmid without T-DNA and another vector with T-DNA containing the marker enzyme beta-glucuronidase, into three different bacteria. Gene transfer was successful and this method significantly expands the tools available for gene delivery into plants.

Microprojectile bombardment delivery

Another widely used technique to genetically transform plants involves the use of microprojectile bombardment, in this process, a nucleic acid containing the

desired genetic elements to be introduced into the plant is deposited on or in small dense particles, e.g., tungsten, platinum, or preferably 1 micron gold particles, which are then delivered at a high velocity into the plant tissue or plant cells using a specialized biolistics device. Many such devices have been designed and constructed; one in particular, the PDS1000/He sold by BioRad, is the instrument most commonly used for biolistics of plant cells. The advantage of this method is that no specialized sequences need to be present on the nucleic acid molecule to be delivered into plant cells; delivery of any nucleic acid sequence is theoretically possible.

For the bombardment, cells in suspension are concentrated on filters or solid culture medium. Alternatively, immature embryos, seedling explants, or any plant tissue or target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate.

Various biolistics protocols have been described that differ in the type of particle or the manner in which DNA is coated onto the particle. Any technique for coating microprojectiles that allows for delivery of transforming DNA to the target cells may be used. For example, particles may be prepared by functionalizing the surface of a gold oxide particle by providing free amine groups. DNA, having a strong negative charge, will then bind to the functionalized particles.

Parameters such as the concentration of DNA used to coat microprojectiles may influence the recovery of transformants containing a single copy of the transgene. For example, a lower concentration of DNA may not necessarily change the efficiency of the transformation but may instead increase the proportion of single copy insertion events. In this regard, ranges of approximately 1 ng to approximately 10 μ g (10,000 ng), approximately 5 ng to 8 μ g or approximately 20 ng, 50 ng, 100 ng, 200 ng, 500 ng, 1 μ g, 2 μ g, 5 μ g, or 7 μ g of transforming DNA may be used per each 1.0-2.0 mg of starting 1.0 micron gold particles.

Other physical and biological parameters may be varied, such as manipulation of the DNA/microprojectile precipitate, factors that affect the flight and velocity of the projectiles, manipulation of the cells before and immediately after bombardment (including osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells), the orientation of an immature embryo or other target

tissue relative to the particle trajectory, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. One may particularly wish to adjust physical parameters such as DNA concentration, gap distance, flight distance, tissue distance, and helium pressure.

5 The particles delivered via biolistics can be “dry” or “wet.” In the “dry” method, the mini-chromosome DNA-coated particles such as gold are applied onto a macrocarrier (such as a metal plate, or a carrier sheet made of a fragile material such as mylar) and dried. The gas discharge then accelerates the macrocarrier into a stopping screen, which halts the macrocarrier but allows the particles to pass through;
10 the particles then continue their trajectory until they impact the tissue being bombarded. For the “wet” method, the droplet containing the mini-chromosome DNA-coated particles is applied to the bottom part of a filter holder, which is attached to a base which is itself attached to a rupture disk holder used to hold the rupture disk to the helium egress tube for bombardment. The gas discharge directly displaces the
15 DNA/gold droplet from the filter holder and accelerates the particles and their DNA cargo into the tissue being bombarded. The wet biolistics method has been described in detail elsewhere but has not previously been applied in the context of plants (Mialhe et al., Mol Mar Biol Biotechnol. 4(4):275-831995). The concentrations of the various components for coating particles and the physical parameters for delivery can
20 be optimized using procedures known in the art.

 A variety of plant cells/tissues are suitable for transformation, including immature embryos, scutellar tissue, suspension cell cultures, immature inflorescence, shoot meristem, epithelial peels, nodal explants, callus tissue, hypocotyl tissue, cotyledons, roots, and leaves, meristem cells, and gametic cells such
25 as microspores, pollen, sperm and egg cells. It is contemplated that any cell from which a fertile plant may be regenerated is useful as a recipient cell. Callus may be initiated from tissue sources including, but not limited to, immature embryos, seedling apical meristems, microspore-derived embryos, roots, hypocotyls, cotyledons and the like. Those cells which are capable of proliferating as callus also are recipient cells for
30 genetic transformation.

 Any suitable plant culture medium can be used. Examples of suitable media would include but are not limited to MS-based media (Murashige and Skoog,

Physiol. Plant, 15:473-497, 1962) or N6-based media(Chu et al., Scientia Sinica 18:659, 1975) supplemented with additional plant growth regulators including but not limited to auxins such as picloram (4-amino-3,5,6-trichloropicolinic acid), 2,4-D (2,4-dichlorophenoxyacetic acid), naphthalene-acetic acid (NAA) and dicamba (3,6-dichloroanisic acid), cytokinins such as BAP (6-benzylaminopurine) and kinetin, and gibberellins. Other media additives can include but are not limited to amino acids, macroelements, iron, microelements, vitamins and organics, carbohydrates, undefined media components such as casein hydrolysates, an appropriate gelling agent such as a form of agar, a low melting point agarose or Gelrite if desired. Those of skill in the art are familiar with the variety of tissue culture media, which when supplemented appropriately, support plant tissue growth and development and are suitable for plant transformation and regeneration. These tissue culture media can either be purchased as a commercial preparation, or custom prepared and modified. Examples of such media would include but are not limited to Murashige and Skoog (Mursahige and Skoog, Physiol. Plant, 15:473-497, 1962), N6 (Chu et al., Scientia Sinica 18:659, 1975), Linsmaier and Skoog (Linsmaier and Skoog, Physio. Plant., 18:100, 1965), Uchimiya and Murashige (Uchimiya and Murashige, Plant Physiol. 15:473, 1962), Gamborg's B5 media (Gamborg et al., Exp. Cell Res., 50:151, 1968), D medium (Duncan et al., Planta, 165:322-332, 1985), Mc-Cown's Woody plant media (McCown and Lloyd, HortScience 6:453, 1981), Nitsch and Nitsch (Nitsch and Nitsch, Science 163:85-87, 1969), and Schenk and Hildebrandt (Schenk and Hildebrandt, Can. J. Bot. 50:199-204, 1972) or derivations of these media supplemented accordingly. Those of skill in the art are aware that media and media supplements such as nutrients and growth regulators for use in transformation and regeneration and other culture conditions such as light intensity during incubation, pH, and incubation temperatures can be varied.

Those of skill in the art are aware of the numerous modifications in selective regimes, media, and growth conditions that can be varied depending on the plant system and the selective agent. Typical selective agents include but are not limited to antibiotics such as geneticin (G418), kanamycin, paromomycin or other chemicals such as glyphosate or other herbicides. Consequently, such media and culture conditions disclosed in the present invention can be modified or substituted

with nutritionally equivalent components, or similar processes for selection and recovery of transgenic events, and still fall within the scope of the present invention.

Mini-chromosome Delivery without selection

5 The Mini-chromosome is delivered to plant cells or tissues, e.g., plant cells in suspension to obtain stably modified callus clones for inheritance assay. Suspension cells are maintained in a growth media, for example Murashige and Skoog (MS) liquid medium containing an auxin such as 2,4-dichlorophenoxyacetic acid (2,4-D). Cells are bombarded using a particle bombardment process, such as the helium-driven PDS-1000/He system, and propagated in the same liquid medium to
10 permit the growth of modified and non-modified cells. Portions of each bombardment are monitored for formation of fluorescent clusters, which are isolated by micromanipulation and cultured on solid medium. Clones modified with the mini-chromosome are expanded and homogenous clones are used in inheritance assays, or assays measuring mini-chromosome structure or autonomy.

15 Mini-chromosome transformation with selectable marker gene

Isolation of mini-chromosome-modified cells in bombarded calluses or explants can be facilitated by the use of a selectable marker gene. The bombarded tissues are transferred to a medium containing an appropriate selective agent for a particular selectable marker gene. Such a transfer usually occurs between 0 and about
20 7 days after bombardment. The transfer could also take place any number of days after bombardment. The amount of selective agent and timing of incorporation of such an agent in selection medium can be optimized by using procedures known in the art. Selection inhibits the growth of non-modified cells, thus providing an advantage to the growth of modified cells, which can be further monitored by tracking
25 the presence of a fluorescent marker gene or by the appearance of modified explants (modified cells on explants may be green under light in selection medium, while surrounding non- modified cells are weakly pigmented). In plants that develop through shoot organogenesis (*e.g. Brassica*, tomato or tobacco), the modified cells can form shoots directly, or alternatively, can be isolated and expanded for
30 regeneration of multiple shoots transgenic for the mini-chromosome. In plants that develop through embryogenesis (*e.g. corn* or soybean), additional culturing steps may be necessary to induce the modified cells to form an embryo and to regenerate in the appropriate media.

Useful selectable marker genes are well known in the art and include, for example, herbicide and antibiotic resistance genes including but not limited to neomycin phosphotransferase II (conferring resistance to kanamycin, paramomycin and G418), hygromycin phosphotransferase (conferring resistance to hygromycin), 5-
5 enolpyruvylshikimate-3-phosphate synthase (EPSPS, conferring resistance to glyphosate), phosphinothricin acetyltransferase (conferring resistance to phosphinothricin/bialophos), MerA (conferring resistance to mercuric ions). Selectable marker genes may be transformed using standard methods in the art.

The first step in the production of plants containing novel genes
10 involves delivery of DNA into a suitable plant tissue (described in the previous section) and selection of the tissue under conditions that allow preferential growth of any cells containing the novel genes. Selection is typically achieved with a selectable marker gene present in the delivered DNA, which may be a gene conferring resistance to an antibiotic, herbicide or other killing agent, or a gene allowing utilization of a
15 carbon source not normally metabolized by plant cells. For selection to be effective, the plant cells or tissue need to be grown on selective medium containing the appropriate concentration of antibiotic or killing agent, and the cells need to be plated at a defined and constant density. The concentration of selective agent and cell density are generally chosen to cause complete growth inhibition of wild type plant tissue that
20 does not express the selectable marker gene; but allowing cells containing the introduced DNA to grow and expand into adchromosomal clones. This critical concentration of selective agent typically is the lowest concentration at which there is complete growth inhibition of wild type cells, at the cell density used in the experiments. However, in some cases, sub-killing concentrations of the selective
25 agent may be equally or more effective for the isolation of plant cells containing mini-chromosome DNA, especially in cases where the identification of such cells is assisted by a visible marker gene (e.g., fluorescent protein gene) present on the mini-chromosome.

In some species (*e.g.*, tobacco or tomato), a homogenous clone of
30 modified cells can also arise spontaneously when bombarded cells are placed under the appropriate selection. An exemplary selective agent is the neomycin phosphotransferase II (nptII) marker gene, which is commonly used in plant biotechnology and confers resistance to the antibiotics kanamycin, G418 (geneticin)

and paramomycin. In other species, or in certain plant tissues or when using particular selectable markers, homogeneous clones may not arise spontaneously under selection; in this case the clusters of modified cells can be manipulated to homogeneity using the visible marker genes present on the mini-chromosomes as an indication of which
5 cells contain mini-chromosome DNA.

Regeneration of adchromosomal plants from explants to mature, rooted plants

For plants that develop through shoot organogenesis (*e.g. Brassica*, tomato and tobacco), regeneration of a whole plant involves culturing of regenerable explant tissues taken from sterile organogenic callus tissue, seedlings or mature plants
10 on a shoot regeneration medium for shoot organogenesis, and rooting of the regenerated shoots in a rooting medium to obtain intact whole plants with a fully developed root system. These plants are potted in soil and grown to maturity in a greenhouse.

For plant species, such corn and soybean, regeneration of a whole plant
15 occurs via an embryogenic step that is not necessary for plant species where shoot organogenesis is efficient. In these plants the explant tissue is cultured on an appropriate media for embryogenesis, and the embryo is cultured until shoots form. The regenerated shoots are cultured in a rooting medium to obtain intact whole plants with a fully developed root system. These plants are potted in soil and grown to
20 maturity in a greenhouse.

Explants are obtained from any tissues of a plant suitable for regeneration. Exemplary tissues include hypocotyls, internodes, roots, cotyledons, petioles, cotyledonary petioles, leaves and peduncles, prepared from sterile seedlings or mature plants.

25 Explants are wounded (for example with a scalpel or razor blade) and cultured on a shoot regeneration medium (SRM) containing Murashige and Skoog (MS) medium as well as a cytokinin, *e.g.*, 6-benzylaminopurine (BA), and an auxin, *e.g.*, α -naphthaleneacetic acid (NAA), and an anti-ethylene agent, *e.g.*, silver nitrate (AgNO_3). For example, 2 mg/L of BA, 0.05 mg/L of NAA, and 2 mg/L of AgNO_3
30 can be added to MS medium for shoot organogenesis. The most efficient shoot regeneration is obtained from longitudinal sections of internode explants.

Shoots regenerated via organogenesis are rooted in a MS medium containing low concentrations of an auxin such as NAA. Plants are potted and grown in a greenhouse to sexual maturity for seed harvest.

To regenerate a whole plant with a mini-chromosome, explants are
5 pre-incubated for 1 to 7 days (or longer) on the shoot regeneration medium prior to bombardment with mini-chromosome (see below). Following bombardment, explants are incubated on the same shoot regeneration medium for a recovery period up to 7 days (or longer), followed by selection for transformed shoots or clusters on the same medium but with a selective agent appropriate for a particular selectable marker gene
10 (see below).

Method of co-delivering growth inducing genes to facilitate isolation of adchromosomal plant cell clones

Another method used in the generation of cell clones containing mini-chromosomes involves the co-delivery of DNA containing genes that are capable of
15 activating growth of plant cells, or that promote the formation of a specific organ, embryo or plant structure that is capable of self-sustaining growth. In one embodiment, the recipient cell receives simultaneously the mini-chromosome, and a separate DNA molecule encoding one or more growth promoting, organogenesis-promoting, embryogenesis-promoting or regeneration-promoting genes. Following
20 DNA delivery, expression of the plant growth regulator genes stimulates the plant cells to divide, or to initiate differentiation into a specific organ, embryo, or other cell types or tissues capable of regeneration. Multiple plant growth regulator genes can be combined on the same molecule, or co-bombarded on separate molecules. Use of these genes can also be combined with application of plant growth regulator
25 molecules into the medium used to culture the plant cells, or of precursors to such molecules that are converted to functional plant growth regulators by the plant cell's biosynthetic machinery, or by the genes delivered into the plant cell.

The co-bombardment strategy of mini-chromosomes with separate DNA molecules encoding plant growth regulators transiently supplies the plant
30 growth regulator genes for several generations of plant cells following DNA delivery. During this time, the mini-chromosome may be stabilized by virtue of its centromere, but the DNA molecules encoding plant growth regulator genes, or organogenesis-promoting, embryogenesis-promoting or regeneration-promoting genes will tend to be

lost. The transient expression of these genes, prior to their loss, may give the cells containing mini-chromosome DNA a sufficient growth advantage, or sufficient tendency to develop into plant organs, embryos or a regenerable cell cluster, to outgrow the non- modified cells in their vicinity, or to form a readily identifiable structure that is not formed by non- modified cells. Loss of the DNA molecule encoding these genes will prevent phenotypes from manifesting themselves that may be caused by these genes if present through the remainder of plant regeneration. In rare cases, the DNA molecules encoding plant growth regulator genes will integrate into the host plant's genome or into the mini-chromosome.

Under a different embodiment of this invention, the genes promoting plant cell growth may be genes promoting shoot formation or embryogenesis, or giving rise to any identifiable organ, tissue or structure that can be regenerated into a plant. In this case, it may be possible to obtain embryos or shoots harboring mini-chromosomes directly after DNA delivery, without the need to induce shoot formation with growth activators supplied into the medium, or lowering the growth activator treatment necessary to regenerate plants. The advantages of this method are more rapid regeneration, higher transformation efficiency, lower background growth of non- modified tissue, and lower rates of morphologic abnormalities in the regenerated plants (due to shorter and less intense treatments of the tissue with chemical plant growth activators added to the growth medium).

Determination of mini-chromosome structure and autonomy in adchromosomal plants and tissues

The structure and autonomy of the mini-chromosome in adchromosomal plants and tissues can be determined by methods including but not limited to: conventional and pulsed-field Southern blot hybridization to genomic DNA from modified tissue subjected or not subjected to restriction endonuclease digestion, dot blot hybridization of genomic DNA from modified tissue hybridized with different mini-chromosome specific sequences, mini-chromosome rescue, exonuclease activity, PCR on DNA from modified tissues with probes specific to the mini-chromosome, or Fluorescence In Situ Hybridization to nuclei of modified cells. Table below summarizes these methods.

Assay	Assay details	Potential outcome	Interpretation
Southern blot	Restriction digest of genomic DNA* compared to purified mini-C	Native sizes and pattern of bands	Autonomous or integrated via CEN fragment
		Altered sizes or pattern of bands	Integrated or rearranged
CHEF gel Southern blot	Restriction digest of genomic DNA compared to purified mini-C	Native sizes and pattern of bands	Autonomous or integrated via CEN fragment
		Altered sizes or pattern of bands	Integrated or rearranged
	Native genomic DNA (no digest)	Mini-C band migrating ahead of genomic DNA	Autonomous circles or linears present in plant
		Mini-C band co-migrating with genomic DNA	Integrated
		>1 mini-C bands observed	Various possibilities
Exonuclease assay	Exonuclease digestion of genomic DNA followed by detection of circular mini-chromosome by PCR, dot blot, or restriction digest (optional), electrophoresis and southern blot (useful for circular mini-chromosomes)	Signal strength close to that w/o exonuclease	Autonomous circles present
		No signal or signal strength lower than w/o exonuclease	Integrated
Mini-chromosome rescue	Transformation of plant genomic DNA into <i>E. coli</i> followed by selection for antibiotic resistance gene on mini-C	Colonies isolated only from mini-C plants with mini-C not from controls; mini-C structure matches that of parental mini-C	Autonomous circles present native mini-C structure
		Colonies isolated only from mini-C plants with mini-C not from controls; mini-C structure different from parental mini-C	Autonomous circles present rearranged mini-C structure OR mini-Cs integrated via centromere fragment
		Colonies observed both in mini-C-modified plants and in controls	Various possibilities
PCR	PCR amplification of various parts of the mini-chromosome	All mini-c parts detected by PCR	Complete mini-C sequence present in plant
		Subset of mini-c parts detected by PCR	Partial mini-C sequences present in plant
FISH	Detection of mini-chromosome sequences in mitotic or meiotic nuclei by fluorescence in situ hybridization	Mini-C sequences detected free of genome	autonomous
		Mini-C sequences detected associated with genome	integrated
		Mini-C sequences detected both free and associated with genome	Both autonomous and integrated mini-C sequences present
		No mini-C sequences detected	Mini-C DNA not visible by FISH

* Genomic DNA refers to total DNA extracted from plants containing a mini-chromosome

Furthermore, mini-chromosome structure can be examined by characterizing mini-chromosomes 'rescued' from adchromosomal cells. Circular mini-chromosomes that contain bacterial sequences for their selection and propagation in bacteria can be rescued from an adchromosomal plant or plant cell and re-introduced into bacteria. If no loss of sequences has occurred during replication of the mini-chromosome in plant cells, the mini-chromosome is able to replicate in bacteria and confer antibiotic resistance. Total genomic DNA is isolated from the adchromosomal plant cells by any method for DNA isolation known to those skilled in the art, including but not limited to a standard cetyltrimethylammonium bromide (CTAB) based method (Current Protocols in Molecular Biology (1994) John Wiley & Sons, N.Y., 2.3) The purified genomic DNA is introduced into bacteria (e.g., *E. coli*) using methods familiar to one skilled in the art (for example heat shock or electroporation). The transformed bacteria are plated on solid medium containing antibiotics to select bacterial clones modified with mini-chromosome DNA. Modified bacterial clones are grown up, the plasmid DNA purified (by alkaline lysis for example), and DNA analyzed by restriction enzyme digestion and gel electrophoresis or by sequencing. Because plant-methylated DNA containing methylcytosine residues will be degraded by wild-type strains of *E. coli*, bacterial strains (e.g. DH10B) deficient in the genes encoding methylation restriction nucleases (e.g. the *mcr* and *mrr* gene loci in *E. coli*) are best suited for this type of analysis. Mini-chromosome rescue can be performed on any plant tissue or clone of plant cells modified with a mini-chromosome.

Mini-chromosome autonomy demonstration by In Situ Hybridization (ISH)

To assess whether the mini-chromosome is autonomous from the native plant chromosomes, or has integrated into the plant genome, In Situ Hybridization is carried out (Fluorescent In Situ Hybridization or FISH is particularly well suited to this purpose). In this assay, mitotic or meiotic tissue, such as root tips or meiocytes from the anther, possibly treated with metaphase arrest agents such as colchicines is obtained, and standard FISH methods are used to label both the centromere and sequences specific to the mini-chromosome. For example, a *Zea* centromere is labeled using a probe from a sequence that labels all *Zea* centromeres, attached to one fluorescent tag (Molecular Probes Alexafluor 568, for example), and sequences specific to the mini-chromosome are labeled with another fluorescent tag

(Alexafluor 488, for example). All centromere sequences are detected with the first tag; only mini-chromosomes are detected with both the first and second tag.

Chromosomes are stained with a DNA-specific dye including but not limited to DAPI, Hoechst 33258, OliGreen, Giemsa YOYO, and TOTO. An autonomous mini-
5 chromosome is visualized as a body that shows hybridization signal with both centromere probes and mini-chromosome specific probes and is separate from the native chromosomes.

Determination of gene expression levels

The expression level of any gene present on the mini-chromosome can
10 be determined by methods including but not limited to one of the following. The mRNA level of the gene can be determined by Northern Blot hybridization, Reverse Transcriptase- Polymerase Chain Reaction, binding levels of a specific RNA-binding protein, in situ hybridization, or dot blot hybridization.

The protein level of the gene product can be determined by Western
15 blot hybridization, Enzyme-Linked Immunosorbant Assay (ELISA), fluorescent quantitation of a fluorescent gene product, enzymatic quantitation of an enzymatic gene product, immunohistochemical quantitation, or spectroscopic quantitation of a gene product that absorbs a specific wavelength of light.

Use of exonuclease to isolate circular mini-chromosome DNA from genomic DNA:

20 Exonucleases may be used to obtain pure mini-chromosome DNA, suitable for isolation of mini-chromosomes from *E. coli* or from plant cells. The method assumes a circular structure of the mini-chromosome. A DNA preparation containing mini-chromosome DNA and genomic DNA from the source organism is treated with exonuclease, for example lambda exonuclease combined with *E. coli*
25 exonuclease I, or the ATP-dependent exonuclease (Qiagen Inc). Because the exonuclease is only active on DNA ends, it will specifically degrade the linear genomic DNA fragments, but will not affect the circular mini-chromosome DNA. The result is mini-chromosome DNA in pure form. The resultant mini-chromosome DNA can be detected by a number of methods for DNA detection known to those skilled in
30 the art, including but not limited to PCR, dot blot followed by hybridization analysis, and southern blot followed by hybridization analysis. Exonuclease treatment followed by detection of resultant circular mini-chromosome may be used as a method to determine mini-chromosome autonomy.

Structural analysis of mini-chromosomes by BAC-end sequencing:

BAC-end sequencing procedures, known to those skilled in the art, can be applied to characterize mini-chromosome clones for a variety of purposes, such as structural characterization, determination of sequence content, and determination of the precise sequence at a unique site on the chromosome (for example the specific sequence signature found at the junction between a centromere fragment and the vector sequences). In particular, this method is useful to prove the relationship between a parental mini-chromosome and the mini-chromosomes descended from it and isolated from plant cells by mini-chromosome rescue, described above.

Methods for scoring meiotic mini-chromosome inheritance

A variety of methods can be used to assess the efficiency of meiotic mini-chromosome transmission. In one embodiment of the method, gene expression of genes encoded by the mini-chromosome (marker genes or non-marker genes) can be scored by any method for detection of gene expression known to those skilled in the art, including but not limited to visible methods (*e.g.* fluorescence of fluorescent protein markers, scoring of visible phenotypes of the plant), scoring resistance of the plant or plant tissues to antibiotics, herbicides or other selective agents, by measuring enzyme activity of proteins encoded by the mini-chromosome, or measuring non-visible plant phenotypes, or directly measuring the RNA and protein products of gene expression using microarray, northern blots, in situ hybridization, dot blot hybridization, RT-PCR, western blots, immunoprecipitation, Enzyme-Linked Immunosorbant Assay (ELISA), immunofluorescence and radio-immunoassays (RIA). Gene expression can be scored in the post-meiotic stages of microspore, pollen, pollen tube or female gametophyte, or the post-zygotic stages such as embryo, seed, or progeny seedlings and plants. In another embodiment of the method, the mini-chromosome can be directly detected or visualized in post-meiotic, zygotic, embryonal or other cells in by a number of methods for DNA detection known to those skilled in the art, including but not limited to fluorescence in situ hybridization, in situ PCR, PCR, southern blot, or by mini-chromosome rescue described above.

FISH analysis of mini-chromosome copy number in meiocytes, roots or other tissues of adchromosomal plants

The copy number of the mini-chromosome can be assessed in any cell or plant tissue by In Situ Hybridization (Fluorescent In Situ Hybridization or FISH is

particularly well suited to this purpose). In an exemplary assay, standard FISH methods are used to label the centromere, using a probe which labels all chromosomes with one fluorescent tag (Molecular Probes Alexafluor 568, for example), and to label sequences specific to the mini-chromosome with another fluorescent tag (Alexafluor 488, for example). All centromere sequences are detected with the first tag; only mini-chromosomes are detected with both the first and second tag. Nuclei are stained with a DNA-specific dye including but not limited to DAPI, Hoechst 33258, OliGreen, Giemsa YOYO, and TOTO. Mini-chromosome copy number is determined by counting the number of fluorescent foci that label with both tags.

10 **Induction of callus and roots from adchromosomal plants tissues for inheritance assays**

Mini-chromosome inheritance is assessed using callus and roots induced from transformed plants. To induce roots and callus, tissues such as leaf pieces are prepared from adchromosomal plants and cultured on a Murashige and Skoog (MS) medium containing a cytokinin, *e.g.*, 6-benzylaminopurine (BA), and an auxin, *e.g.*, α -naphthaleneacetic acid (NAA). Any tissue of an adchromosomal plant can be used for callus and root induction, and the medium recipe for tissue culture can be optimized using procedures known in the art.

Clonal propagation of adchromosomal plants

20 To produce multiple clones of plants from a mini-chromosome-transformed plant, any tissue of the plant can be tissue-cultured for shoot organogenesis using regeneration procedures described under the section regeneration of plants from explants to mature, rooted plants (see above). Alternatively, multiple auxiliary buds can induced from a mini-chromosome-modified plant by excising the shoot tip, which can be rooted and subsequently be grown into a whole plant; each auxiliary bud can be rooted and produce a whole plant.

Scoring of antibiotic- or herbicide resistance in seedlings and plants (progeny of self- and out-crossed transformants)

Progeny seeds harvested from mini-chromosome-modified plants can be scored for antibiotic- or herbicide resistance by seed germination under sterile conditions on a growth media (for example Murashige and Skoog (MS) medium) containing an appropriate selective agent for a particular selectable marker gene. Only seeds containing the mini-chromosome can germinate on the medium and

further grow and develop into whole plants. Alternatively, seeds can be germinated in soil, and the germinating seedlings can then be sprayed with a selective agent appropriate for a selectable marker gene. Seedlings that do not contain mini-chromosome do not survive; only seedlings containing mini-chromosome can survive and develop into mature plants.

Genetic methods for analyzing mini-chromosome performance:

In addition to direct transformation of a plant with a mini-chromosome, plants containing a mini-chromosome can be prepared by crossing a first plant containing the functional, stable, autonomous mini-chromosome with a second plant lacking the mini-chromosome.

Fertile plants modified with mini-chromosomes can be crossed to other plant lines or plant varieties to study mini-chromosome performance and inheritance, in the first embodiment of this method, pollen from an adchromosomal plant can be used to fertilize the stigma of a non-adchromosomal plant. Mini-chromosome presence is scored in the progeny of this cross using the methods outlines in the preceding section. In the second embodiment, the reciprocal cross is performed by using pollen from a non-adchromosomal plant to fertilize the flowers of a adchromosomal plant. The rate of mini-chromosome inheritance in both crosses can be used to establish the frequencies of meiotic inheritance in male and female meiosis. In the third embodiment of this method, the progeny of one of the crosses just described are back-crossed to the non-adchromosomal parental line, and the progeny of this second cross are scored for the presence of genetic markers in the plant's natural chromosomes as well as the mini-chromosome. Scoring of a sufficient marker set against a sufficiently large set of progeny allows the determination of linkage or co-segregation of the mini-chromosome to specific chromosomes or chromosomal loci in the plant's genome. Genetic crosses performed for testing genetic linkage can be done with a variety of combinations of parental lines; such variations of the methods described are known to those skilled in the art.

It should be understood that various changes and modifications to the presently preferred embodiments described herein will be apparent to those skilled in the art. Such changes and modifications can be made without departing from the scope of the present invention and without diminishing its intended

advantages. It is therefore intended that such changes and modifications be covered by the appended claims.

Example 1 Corn Centromere Discovery

5 BAC library construction

Two Bacterial Artificial Chromosome (BAC) libraries were constructed from corn genomic DNA. The corn genomic DNA was isolated from corn variety B73 and digested with the restriction enzymes *Bst*YI or *Mbo*I. These enzymes were chosen because they are methylation insensitive and therefore can be
10 used to enrich BAC libraries for centromere DNA sequences.

Probe identification and selection

Twenty-three groups of corn repetitive genomic or plastid sequences, including specific centromere-localized sequences, were initially compiled as candidate probes for hybridization with the BAC libraries (Table 3). These probes
15 represented various classes of corn repetitive sequences including satellite repeats (heterochromatic / centromere-specific), retroelements, rDNA, B chromosome-specific repeats, chloroplast and mitochondrion DNA, hypermethylated or hypomethylated DNA fractions, and telomeric DNA.

Table 3 Maize Repetitive Sequences and Bac Library Probes

Class	Class Name	Primers	Description	Reference	Comment	GenBank accession
1	CR (centromere retrotransposon) element	CRJM-001 and 002	gypsy-type localized to centromeres in all cereals. Centromere and CRM co-IP with CEN H3	Aragon-Alcaide et al 1996, Jiang et al 1996, Zhong et al 2002	aka CRM, pSau3A9 (from sorghum), CR (from rice)	AY1290008
2	CentA	CHR 15 and 1	centromere retrotransposon, includes MCS1A and B		AF082532 Similar sequence	AF078917
3	Huck	CRJM-005 and 006	Ty3/gypsy	Meyers et al 2000	(most frequent)	AF050438
4	Grande	CRJM-056 and 057	Ty3/gypsy	Meyers et al 2000		AF050437
5	Cinful	CRJM-007 and 008	Ty3/gypsy	Meyers et al 2000		AF049110
6	Ji/Prem2	LTR-5 CRJM-011 and 012 gag CRJM-013 and 014	Ty1/copia	Meyers et al 2000		from alpha zein seq
7	Opie		Ty1/copia	Meyers et al 2000	5' LTR	AF050453
8	Tekay	CRJM-009 and 010			3' LTR	AF050452
9	alpha zein					AF090447

WO 2007/030510

PCT/US2006/034669

10	adh					AF123535
11	bz					AF448416
12	knob 180	CHR 11 and 1			many sequences!	gi 168710 gb M2521.1 MZEZNA
13	MZEHETRO	CRJM-015 and 016	maize heterochromatic repeat (knob)	Peacock et al PNAS. 78, 4490-4494 (1981)		M35408
14	TR-1(knob 36)	CHR 13 and 1	Knob-specific	Hsu et al 2002	3 lengths, multiple types. Type 1 BLASTs to a 3. Cuts w/RI	AF071126
15	CentC	CHR 17 and 1	156 bp	Ananiev et al 19	all match well	AY321491 (CentC27)
						AF078923 158
		CRJM-019 and 020				AF078922 156
16	Cent4	CRJM-021 and 022	Chromosome 4 repeat homologous to B-chromosome cent repeat	Page et al, 2001		AF242891
17	pZmBs and K	S67586	B-specific repeat B73 has no B chromosomes	Alfenito and Birchler 1993; Kaszas and Birchler 1993, 1998		AY173950
18	rDNA	CRJM-023 and 024	maize intergenic spacer			AF013103
		CRJM-025 and 026	maize 5S			AF273104
		CRJM-027 and 028	maize 17S			K0220
19	chloroplast	CHHZ211 and 212	Arabidiosis			
		CRJM-030 and 031	maize xpl rDNA			X01365
20	mito	CHHZ214 and 215	Arabidiosis			
		CRJM-032 and 033	maize mito 26S rDNA			K01868
21	hypermethylated fraction	purified				complex mixture
22	hypomethylated fraction	purified				complex mixture
23	telomere		sub-telomeric repeat		U39641	U39642

Twelve probes were picked to interrogate the BAC libraries. These probes represent different groups of commonly found repetitive sequences in the corn genome. The twelve probes selected are shown in Tables 3 and 4 and were: CentC (#15), Cent4 (#16), MZEHETRO (#13) , TR-1 (#14), CentA (#2), CR (#1), Huck

WO 2007/030510**PCT/US2006/034669**

(#3), Grande (#4), 17S rDNA (#18), 5S rDNA (#18); B cen (#17), and xplmito (#19 and #20). The primers used to amplify these probes are identified in Table 4. Probes were prepared and labeled with standard molecular methods.

Table 4 Classification of maize BAC Clones Containing Centromeric DNA														
		Probe Hybridization Range												
Class	Class Properties	CentC	CentA	CR	Huck	Grande	17S rDNA	Cent4	TR-1	MZE HETRO	5S rDNA	B cen	xplmito	# clones identified
I	HiC LoA	>= 7	< 7	< 7	< 7	< 6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	61
II	HiC HiA	>= 7	>= 6	< 7	<= 10	<= 10	N/A	N/A	N/A	N/A	N/A	N/A	N/A	61
III	HiCR HiC	>= 7	< 6	>= 6	<= 10	<= 10	N/A	N/A	N/A	N/A	N/A	N/A	N/A	30
IV	HiA HiC HiCR	>= 7	> 6	>= 6	<= 10	<= 10	N/A	N/A	N/A	N/A	N/A	N/A	N/A	30
V	HiC Hi17s	>= 7	> 0	> 0	> 0	> 0	> 5	N/A	N/A	N/A	N/A	N/A	N/A	30
VI	Hi4	> 0	> 0	> 0	> 0	N/A	N/A	> 5	N/A	N/A	N/A	N/A	N/A	17
VII	HiTr1 LoHe	> 0	> 0	N/A	N/A	N/A	> 0	N/A	> 6	< 6	N/A	N/A	N/A	31
VIII	LoTr1 HiHe	> 0	> 0	N/A	N/A	N/A	> 0	N/A	< 5	> 7	N/A	N/A	N/A	31
IX	HiTr1 HiHe	> 0	> 0	N/A	N/A	N/A	> 0	N/A	> 6	> 6	N/A	N/A	N/A	24
Total														315

* Values represent hybridization intensities of an individual BAC to each probe on a scale of 1 to 10. Values were normalized.

Library interrogation and data analysis

The BAC clones from the libraries were spotted onto filters for further analysis. The filters were hybridized with each of the 12 probes to identify specific BAC clones that contain DNA from the group of sequences represented by the probe(s).

- 5 Exemplary hybridization conditions: 0.5 x SSC 0.25% SDS at 65 degrees for 15 minutes, followed by a wash at 65 degrees for a half hour.

A total of 92,160 BAC clones from the two libraries (36,864 BAC clones from 2 filters from the BstYI library and 55,296 clones from 3 filters from the MboI library) were interrogated with each of the 12 probes described above, and the hybridization intensities of the BAC clones with each probe were scanned to quantitate hybridization intensity for each clone. Scores of 1 to 10 (based on the hybridization intensities, with 10 being the strongest hybridization) were imported into a relational database, for classification. The database contained a total of 24 tables, 12 from each library used in the interrogation. Each table contained the hybridization scores of each BAC clone from the BstY1 or MboI library, to one of the 12 probes. Data analysis was carried out using standard SQL (Structured Query Language) routines to find BACs that contain different groups of repetitive sequences.

Classification and selection of BAC clones for mini-chromosome construction

BAC clones containing centromeric/heterochromatic DNA were identified by their hybridization scores to different probes. The goal was to select BAC clones that contained a diverse set of various repetitive sequences. Nine classes of centromeric BAC clones were eventually chosen to cover the broadest possible range of centromeric/heterochromatic sequences for mini-chromosome construction. Detailed descriptions of each class and probe hybridization values for each class are shown in Table 4.

Class I (HiC LoA) BAC clones had strong hybridization to probe CentC, but low hybridization to CentA, CR, Huck and Grande. Class II (HiC HiA) BAC clones had strong hybridization to both CentC and CentA, but low hybridization to CR. Class III (HiCR HiC) BAC clones had strong hybridization to both CentC and CR, but low hybridization to CentA. Class IV (HiA HiC HiCR) BAC clones had strong hybridization to CentC, CentA, and CR. Class V (HiC Hi17s) BAC clones had strong hybridization to CentC and 17S rDNA. Class VI (Hi4) BAC clones had strong hybridization to Cent4. Class VII (HiTr1 LoHet) BAC clones had strong hybridization to TR-1 but low hybridization to MZEHETRO. Class VIII (LoTr1 HiHet) BAC clones had strong hybridization to MZEHETRO but low

hybridization to TR-1. Class IX (HiTr1 HiHet) BAC clones had strong hybridization to both TR-1 and MZEHETRO.

5 A number of representative clones from each class were chosen to yield a total of 315 BAC clones for further analysis by restriction digest fingerprinting. The number of clones chosen in each class is shown in Table 4.

10 The 315 BAC clones were fingerprinted based on restriction sites found in the centromere specific sequence(s). Fingerprinting was used to evaluate the sequence composition of the large numbers of BAC clones and to compare their similarity to each other by comparing the restriction enzyme digest fragment patterns. A sequence with a tandem repeated sequence will show a single intense band of unit repeat size when digested with a restriction enzyme that cuts within the unit repeat. Second, BAC clones with similar sequences will show similar patterns of restriction fragments in a digest.

15 BAC DNA was extracted from bacteria using methods familiar to those in the art. Restriction enzymes *HpaII* and *MspI* were used to digest BAC clones in Classes I through VI, and restriction enzyme *NdeI* was used to digest BAC clones in classes VII through IX.

20 *Z. mays* BACs ZB19 and ZB113 were deposited with the American Type Culture Collection (ATCC) on February 22, 2005 and assigned accession nos. PTA-6604 and PTA-6605. ZB19 was classified as "class 1" or "HiCLoA when characterized with the restriction endonucleases *HpaII*, *MspI* and fingerprint class CL/SL, sm. ZB113 was classified as "class 4" or "HiA, HiC and HiCR and fingerprint class CL/SL.

Example 2

Construction of Maize Mini-chromosomes

25 The 315 BAC clones identified in Example 1 were grown up and DNA was extracted for mini-chromosome construction using NucleoBond™ Purification Kit (Clontech). To determine the molecular weight of centromere fragments in the BAC libraries, a frozen sample of bacteria harboring a BAC clone was grown in selective liquid media and the BAC DNA harvested using a standard alkaline lysis method. The recovered BAC DNA was restriction digested and resolved on an agarose gel. Centromere fragment size was 30 determined by comparing to a molecular weight standard.

For each BAC, two types of mini-chromosomes were generated, differing only by the promoter used to express the DsRed gene. Corn ADH promoter was used to express

DsRed in mini-chromosomes constructed with pCHR667 and the Arabidopsis UBQ10 promoter was used to express DsRed in mini-chromosomes constructed with pCHR758. Mini-chromosome genetic elements within the pCHR667 and pCHR758 vectors are set out in Table 5 and 6, respectively.

5 **Table 5 Donor Components of pCHR667**

Genetic Element	Size (base pair)	Location (bp)	Details
ADH Corn Promoter	1189	14-1202	PCR amplified maize promoter alcohol dehydrogenase 1 (ADH-1) for expression of DsRed in maize (used primers CRJM-42/43)
Maize ADH Intron	579	1216-1794	PCR amplified maize ADH intron with AUG mutation for stabilization of DsRed2 gene transcript and increase protein expression level (used primers CRJM-72/73)
DsRed2 + NLS	780	1817-2596	Nuclear localized red fluorescent protein from <i>Discosoma sp.</i> (Matz, M. et. al <i>Nat Biotechnol</i> 1999 Dec; 17(12): 1227).
ADH Terminator	203	2725-2927	Amplified maize terminator using primers CRJM-46/47
Bacterial Kanamycin	817	3066-3882	Bacterial kanamycin selectable marker
Rps16A terminator	489	4065-4553	Amplified from <i>Arabidopsis thaliana</i> 40S ribosomal protein S16 (At2g0999) for termination of NptII gene
NPTII	795	4617-5411	Neomycin phosphotransferase II plasmid selectable marker
UBQ10 intron	359	5439-5798	PCR amplified <i>Arabidopsis thaliana</i> intron from UBQ10 gene (At4g05320) for stabilization of NptII gene transcript and increase protein expression level
YAT1 yeast promoter	2000	5812-7811	PCR amplified YAT1 promoter from chromosome I of <i>Saccharomyces cerevisiae</i> for expression of NptII in maize
LoxP	34	10341-10374 and 7829-7862	Recombination site for Cre mediated recombination (Arenski et. al 1983, Abremski et. al 1984)

Table 6 Donor Components of pCHR758

Genetic Element	Size (base pair)	Location (bp)	Details
UBQ10 promoter	2038	14-2051	<i>Arabidopsis thaliana</i> polyubiquitin promoter (At4g05320)
DsRed2 + NLS	780	2088-2867	Nuclear localized red fluorescent protein from <i>Discosoma sp.</i> (Matz, M et. al <i>Nat Biotechnol</i> 1999 Dec; 17(12): 1227).
Pyruvate kinase terminator	332	3002-3333	<i>Arabidopsis thaliana</i> pyruvate kinase terminator (At5g52920)
Bacterial Kanamycin	817	3478-4294	Bacterial kanamycin selectable marker
Rps16A terminator	489	4477-4965	Amplified from <i>Arabidopsis thaliana</i> 40S ribosomal protein S16 (At2g0999) for termination of NptII gene
NPTII	795	5029-5823	Neomycin phosphotransferase II plant selectable marker
UBQ10 intron	359	5851-6210	PCR amplified <i>Arabidopsis thaliana</i> intron from UBQ10 gene (At4g05320) for stabilization of NptII gene transcript and increase protein expression level
YAT1 yeast promoter	2000	6224-8223	PCR amplified YAT1 promoter from chromosome I of <i>Saccharomyces cerevisiae</i> for expression of NptII in maize
LoxP	34	8243-8276 & 10751-10788	Recombination site for Cre mediated recombination (Arenski et. al 1983, Abremski et. al 1984)

Corn mini-chromosomes were constructed by following a two-step procedure:

Step 1: Preparation of donor DNA for retrofitting with BAC centromere vectors and Step 2:
 5 Cre-Lox Recombination-BAC and Donor DNA to generate the mini-chromosome. A total of 230 corn mini-chromosomes were constructed using this assembly process, and were subsequently tested in several different corn cell lines.

Preparation of donor DNA for retrofitting

10 *Cre* recombinase-mediated exchange was used to construct mini-chromosomes by combining the plant centromere fragments cloned in pBeloBAC11 with a donor plasmid (*i.e.* pCHR667 or pCHR758, Tables 7 & 8). The recipient BAC vector carrying the plant centromere fragment contained a *loxP* recombination site; the donor plasmid contained two such sites, flanking the sequences to be inserted into the recipient BAC.

15 *Cre* recombinase-mediated exchange was used to construct mini-chromosomes by combining the plant centromere fragments cloned in pBeloBAC11 with a

donor plasmid (*i.e.* pCHR667 & pCHR758, Table 5 & 6). The recipient BAC vector carrying the plant centromere fragment contained a *loxP* recombination site; the donor plasmid contained two such sites, flanking the sequences to be inserted into the recipient BAC. Mini-chromosomes were constructed using a two-step method. First, the donor plasmid was
5 linearized to allow free contact between the two *loxP* site; in this step the backbone of the donor plasmid is eliminated. In the second step, the donor molecules were combined with centromere BACs and were treated with Cre recombinase, generating circular mini-chromosomes with all the components of the donor and recipient DNA. Mini-chromosomes were delivered into *E. coli* and selected on medium containing kanamycin and
10 chloramphenicol. Only vectors that successfully *cre* recombined and contained both selectable markers survived in the medium. Mini-chromosomes were extracted from bacteria and restriction digested to verify DNA composition and calculate centromere insert size.

To determine the molecular weight of the centromere fragments in the mini-chromosomes, three bacterial colonies from each transformation event were independently
15 grown in selective liquid media and the mini-chromosome DNA harvested using a standard alkaline lysis method. The recovered mini-chromosome was restriction digested and resolved on an agarose gel. Centromere fragment size was determined by comparing to a molecular weight standard. If variation in centromere size was noted, the mini-chromosome with the largest centromere insert was used for further experimentation. Selection of Corn Cell
20 Clones Stably Containing Mini-chromosome DNA

Functional Testing of Mini-chromosomes Using Transient Assays

Maize mini-chromosomes were tested in several corn cell lines including PC1117, HiII, and BMS, and the procedure was optimized for antibiotic selection, cell pre-treatments, and bombardment conditions. All assays were transient and fluorescent cells
25 were counted at several time points. Preliminary results identified several mini-chromosomes that successfully generated fluorescent cell clusters.

Example 3

Mini-chromosome Delivery into Maize Cells

Various methods have been used to deliver DNA into plant cells. These
30 include biological methods, such as viruses, physical methods such as biolistic particle bombardment and silicon carbide whiskers, electrical methods such as electroporation, and chemical methods such as the use of poly-ethylene glycol and other compounds known to

stimulate DNA uptake into cells. Biolistic particle bombardment have been the methods that have found most widespread use in plant biotechnology.

Biolistic particle delivery of mini-chromosomes

A biolistic particle delivery method was used to transfer corn mini-chromosomes into a number of different corn tissues including suspension cells, plate-grown calli, and immature embryos. For the purpose of transient delivery or selection of stable cell culture modified with a corn mini-chromosome, suspension cells were used for delivery using wet or dry gold delivery methods. An example of such a suspension culture is the publicly available line, PC1117.

Wet Bombardment

A biolistic delivery method using wet gold particles kept in an aqueous DNA suspension was adapted from the teachings of Milahe and Miller (Biotechniques 16: 924-931, 1994) and used to transform corn cells. To prepare the wet gold particles for bombardment, 1.0 μm gold particles were washed by mixing with 100% ethanol on a vortex followed by spinning the particles in a microfuge at 4000 rpm in order to remove supernatant. Subsequently, the gold particles were washed with sterile distilled water three times, followed by spinning in a microfuge to remove supernatant. The washed gold particles are resuspend in sterile distilled water at a final concentration of 90 mg per ml and stored at 4 °C until use. For bombardment, the gold particle suspension (90 mg/ml) was then mixed rapidly with 1 $\mu\text{g}/\mu\text{l}$ DNA solution (in dH₂O or TE), 2.5 M CaCl₂, and 1 M spermidine. DNA/gold mixture was left at room temperature and used for bombardment within 2-4 hours.

For bombardment of corn cells, the cells were harvested by centrifugation (1200 rpm for 2 minutes) on the day of bombardment. The cells were plated onto 50 mm circular polyester screen cloth disks placed on petri plates with solid medium. The solid medium used was the same medium that the cells are normally grown in, plus 0.26% gelrite, or 0.6% tissue culture agar, added before autoclaving. Approximately 1.5 ml packed cells were placed on each filter disk, and spread out in a very even spot approximately 1 inch in diameter.

Bombardment of the cells was carried out in the BioRad PDS-1000/He Biolistic Particle Delivery System (BioRad). The DNA/gold suspension was resuspended and immediately inserted onto the grid of the filter holder. A 50 mm circular polyester screen cloth disk with the cells was placed into a fresh 60 mm petri dish with the same medium and the cells were covered with a 10x10 cm square of sterile nylon or Dacron chiffon netting. A

metal cylinder was inserted into the petri dish and used to push the netting down to the bottom of the dish. This weight prevents the cells from being dislodged from the plate during bombardment. The petri dish containing the cells was then placed onto the sample holder, and positioned in the sample chamber of the gene gun and bombarded with the DNA/gold suspension. After the bombardment, the cells were scraped off the filter circle in the petri dish containing solid medium with a sterile spatula and transferred to fresh medium in a 125 ml blue-capped glass bottle. The bottles were transferred onto a shaker and grown while shaking at 150 rpm.

Suspensions of the maize cell line PC117 were bombarded with wet gold particles containing DNA from BAC clones ZB10, ZB18, ZB19 and ZB99. After bombardment, all cells were returned to liquid culture and allowed to grow for three days prior to plating in selection media. Subsequently, the transfected cells were grown in selection medium containing various concentrations of antibiotics. The selection media contained either an increasing concentration of kanamycin (25, 50, 75, 100, 125 and 150 µg/ml) or G418 (10, 20, 35, 50, 75 and 100 µg/ml). The growth of clones in the selection medium indicated expression of the selection gene within the mini-chromosome and suggests a functional centromere within the mini-chromosome. These results are summarized in Table 7.

Table 7

Construct	# bombardments	# clones isolated
ZB10R2-1	2	0
ZB18R3-1	2	0
ZB19R2-1	12	9
ZB99R1-1	12	1

Dry Bombardment

A biolistic delivery method using dry gold particles was also carried out to deliver mini-chromosomes to corn embryos. For this method, 5 µg of mini-chromosome DNA was precipitated onto 3 mg of sterilized and washed 0.6µ gold particles. The DNA-containing gold particles were resuspended in cold sterile water containing 2.5 M CaCl₂. The mixture was lightly vortexed, and then filter-sterilized 0.1 M Spermidine (free base) was added to the mixture. Subsequently, the mixture was lightly vortexed and allowed to precipitate on ice for an hour, with vortexing about every 10 minutes. The precipitated DNA

was then washed with 100% ethanol, resuspended in 100% ethanol which was allowed to fully evaporate prior to bombardment.

Immature embryos were excised onto N6 based medium (Chu's N6 medium with 25 μ M silver nitrate) 3-5 days prior to day of bombardment. The embryos were
5 osmotically adjusted approximately 4 hours prior to bombardment. This osmotic medium is composed of Chu's N6 Basal medium with the addition of 25 μ M silver nitrate, 36.4 g/l sorbitol, and 36.4 g/l mannitol. Embryos were arranged scutellar side up in an open ring that had the same diameter as the plate stage in the gun.

The embryos were bombarded using the BioRad PDS-1000/He Biolistic
10 Particle Delivery System. For this bombardment, the rupture disk rating was 1100 psi with one shot per plate of embryos. The distance from the rupture disk to the macrocarrier was 1/4 inch. After bombardment, the plates of embryos were incubated in a dark incubator overnight at 27°C. The following day, the bombarded tissue was transferred to selection medium, Chu's N6 with 200-250 mg/l Paromomycin or 25-35 mg/l G418 (Geneticin), and
15 cultured in the dark. During this transfer, any emerging coleoptiles were removed from the immature embryos.

Approximately 2-3 weeks after bombardment, all tissue was transferred to fresh selective medium at a higher selection pressure of 250-300 mg/l Paromomycin or 35-50 mg/l G418. At this transfer, the callus was separated into approximately 2-3 mm segments.
20 The callus that was proliferating and showed dsRed activity after at least two subcultures was regenerated. Regeneration was initiated when the amount of healthy callus suggested that a minimum of three plants can be regenerated from that event.

For regeneration, the callus was transferred to R1 medium (MS medium with 20 g/l sucrose and 5 mg/l 6-benzyl-aminopurine). Plates were then incubated at 27°C in the
25 dark for 3-7 days. Tissue was then moved to R2 medium (MS medium with 60 g/l sucrose) with either 10 mg/l G418 or 50 mg/l Paromomycin and placed under low light at 26°C. When leaf tissue reached the top of the petri dish, developing plantlets were transferred to R3 medium (MS medium with 15 g/l sucrose) with either 10 mg/l G418 or 50 mg/l under higher light intensity at 26°C to continue plant growth and allow substantial root development.
30 Plantlets were then transferred into moistened soilless mix under a humi-dome to maintain high humidity in a growth chamber for one week prior to being transplanted into the greenhouse.

Example 4

Selection of Corn Cell Clones Stably Containing Mini-chromosome DNA

Use of visible marker genes

The presence of visible marker genes allowed for visual selection of Corn cells stably containing mini-chromosome DNA because any modified cells or cell clusters were readily identified by virtue of fluorescent protein expression. In addition, the use of fluorescent protein expression allowed for the use of sub-killing concentrations of selective agent during growth of plant tissue on selective medium. This flexibility allowed for the use of a wider range of antibiotic concentrations than possible in the absence of a visible marker gene, without having to consider the amount of background growth observed in wild type plant tissue. As a result, the adchromosomal cell clones were isolated with use of certain selectable marker genes, and under conditions that might not be effective in standard selection experiments as practiced in the industry. These selections were typically done at lower antibiotic concentrations than practiced elsewhere, and resulted in higher levels of background growth. Fluorescent cell clusters can be visually identified after one to several weeks of growth on selective media. Clusters of cells stably containing mini-chromosomes were identified by visual observation of fluorescence in the cells in a darkened room.

Manipulation of adchromosomal tissue to homogeneity

After identifying clusters of fluorescent cells, physical manipulations were carried out to allow for the preferential expansion of cells harboring the delivered mini-chromosomes. Non-fluorescent tissue surrounding the fluorescent clusters was trimmed to avoid overgrowth of fluorescent cells by non-fluorescent ones, while retaining a minimum tissue size capable of rapid growth. These manipulations were performed under sterile conditions with the use of a fluorescence stereomicroscope that allows for visualization of the fluorescent cells and cell clumps in the larger pieces of tissue. In between the mechanical purification steps, the tissue was allowed to grow on appropriate media, either in the presence or absence of selection. Over time, a pure population of fluorescent cells was obtained.

Method of co-delivering growth inducing genes to facilitate isolation of adchromosomal plant cell clones

Another method used in the generation of cell clones containing mini-chromosomes involved the co-delivery of DNA containing genes that are capable of activating growth of plant cells. In this method, the cell receiving DNA receives simultaneously the mini-chromosome, and a separate NA molecule encoding one or more growth promoting genes. Following DNA delivery, expression of the plant growth regulator

genes stimulates the plant cells to divide, or to initiate differentiation into a specific organ, embryo, or other cell types or tissues capable of regeneration. Multiple plant growth regulator genes are combined on the same molecule, or co-bombarded on separate molecules. Use of these genes can also be combined with application of plant growth regulator molecules into the medium used to culture the plant cells, or of precursors to such molecules that are converted to functional plant growth regulators by the plant cell's biosynthetic machinery, or by the genes delivered into the plant cell.

The co-bombardment strategy of mini-chromosomes with separate DNA molecules encoding plant growth regulators transiently supplies the plant growth regulator genes for several generations of plant cells following DNA delivery. During this time, the mini-chromosome may be stabilized by virtue of its centromere, but the DNA molecules encoding plant growth regulator genes will tend to be lost. In rare cases, the DNA molecules encoding plant growth regulator genes will integrate into the host plant's genome or into the mini-chromosome.

15

Example 5

Regeneration of Adchromosomal Corn Plants

A total of 125 corn mini-chromosomes were prepared as described herein and are shown in Table 8.

Table 8

BAC Number	Mini-chromosome Number			Bac Number	Mini-chromosome Number	
ZB	667 donor vector	758 donor vector		ZB	667 donor vector	758 donor vector
5	ZB5R1-1			30		ZB130R2-1
6	ZB6R1-1	ZB6R2-1		31		ZB131R2-2
7	ZB7R1-1	ZB7R2-2		37		ZB137R2-3
8	ZB8R1-2	ZB8R2-1		44		ZB144R2-1
9	ZB9R1-1			45		ZB145R2-2
10	ZB10R2-1	ZB10R3-1		46		ZB146R2-1
13	ZB13R1-1	ZB13R2-1		47		ZB147R2-2
14	ZB14R1-1	ZB14R2-1				ZB150R2-1

BAC Number	Mini-chromosome Number			Bac Number	Mini-chromosome Number	
				50		
18	ZB18R2-1	ZB18R3-1		56	ZB156R1-1	ZB156R2-1
19	ZB19R1-1	ZB19R2-1		57	ZB157R1-2	
20	ZB20R1-1			58	ZB158R1-2	ZB158R2-3
21	ZB21R2-1			67		ZB167R2-1
24	ZB24R2-1			75	ZB175R1-1	ZB175R2-1
25		ZB25R2-1		77	ZB177R1-1	ZB177R2-1
29	ZB29R1-1			78	ZB178R1-1	ZB178R2-1
32		ZB32R3-1		99	ZB199R1-1	ZB199R2-1
34		ZB34R3-1		07	ZB207R1-1	
44		ZB44R2-2		11	ZB211R3-1	
49		ZB49R2-1		32	ZB232R1-1	ZB232R2-1
64	ZB64R1-1	ZB64R2-2		33	ZB233R1-1	ZB233R2-1
65	ZB65R1-1			35	ZB235R1-1	ZB235R2-1
66	ZB66R1-1			38		ZB238R2-1
71	ZB71R1-3			43		ZB243R2-1
72	ZB72R1-2			48		ZB248R2-1
73	ZB73R1-3	ZB73R2-1		53		ZB253R2-1
80		ZB80R2-1		58	ZB258R2-1	ZB258R3-2
81		ZB81R2-1		59	ZB259R2-2	
82	ZB82R1-2	ZB82R2-1		60	ZB260R2-2	
94	ZB94R1-1	ZB94R2-1		61	ZB261R2-1	
96	ZB96R1-1	ZB96R2-1		65	ZB265R2-1	
98	ZB98R1-3	ZB98R2-1				ZB271R3-2

BAC Number	Mini-chromosome Number			Bac Number	Mini-chromosome Number	
				71		
99	ZB99R1-1	ZB99R2-1		79		ZB279R3-1
100	ZB100R1-2	ZB100R2-3		82		ZB282R2-2
101	ZB101R1-2	ZB101R2-2		91		ZB291R3-1
		ZB104R2-1		93	ZB293R1-1	
105	ZB105R1-1	ZB105R2-1		95	ZB295R1-3	ZB295R2-1
106	ZB106R1-1	ZB106R2-2		96	ZB296R1-2	ZB296R2-1
108	ZB108R1-2	ZB108R2-1		97	ZB297R1-3	ZB297R2-2
109	ZB109R1-1	ZB109R2-1		98	ZB298R1-1	
113	ZB113R1-1	ZB113R2-1		05	ZB305R1-2	ZB305R2-1
120	ZB120R1-1			08	ZB308R1-1	ZB308R2-2
122	ZB122R1-3	ZB122R2-1				
123	ZB123R1-1					
124	ZB124R1-1					
129		ZB129R2-2				

The biolistic delivery method described above was used to deliver the mini-chromosomes into a number of different corn tissues including suspension cells, plate-grown calli, and immature embryos. For the purpose of transient delivery or selection of stable cell culture modified with a corn mini-chromosome, suspension cells were used for delivery using wet or dry gold delivery methods. An example of such a suspension culture is the publicly available line, PC1117.

To obtain trans-chromosomal corn plants modified with corn mini-chromosomes, standard protocols for corn tissue culture and transformation are followed. Such protocols include the Maize Embryo/Callus Bombardment Protocols available at Iowa State University, College of Agriculture web site.

The transformation process involves the preparation of regenerable tissues such as immature embryos from corn cultivars such as HiII, pre-culture of embryos on an auxin-enriched medium, delivery of miniC's into immature embryos or embryogenic calli, selection and isolation of fluorescent cell clusters, expansion of cell clusters and
 5 formation of transchromosomal embryos, maturation and regeneration of embryos into whole plants.

Example 6

Sequence Analysis of Centromeres

Two BAC clones (ZB19 and ZB113) were sequenced and the centromere
 10 sequences were analyzed using conventional methods. Briefly, the BAC DNA was purified from *E. coli*, sheared and cloned into standard cloning vectors to create a shotgun library. Clones in the library were sequenced as reads 500-900 bp in length. Individual reads were trimmed to remove sequence of poor quality (phred score of < 20) and to remove sequences derived from the cloning vector used to generate the shotgun library.
 15 The remaining sequence information was then filtered to remove *E. coli* sequences, which inevitably contaminate the BAC DNA prep, and sequences corresponding to the known vector component of each mini-chromosome.

The filtered reads and sequences were then analyzed with a variety of tools to establish sequence content and to locate repetitive DNA sequences. Contig assemblies
 20 were recomputed with phredPhrap. The following programs were used extensively: phred/phrap and consed, available on the internet at www.phred.org; and RepeatMasker (available at the Institute of Systems Biology website). The following databases were used to identify maize sequences: Genbank, RepeatMasker Libraries (repeatmaskerlibraries20050523.tar.gz), TIGR databases "characterized_02202004.fasta",
 25 "uncharacterized_02202004.fasta", "RECON_prediction_02202004.fasta" which are accessible at the TIGR web site.

As described in detail below, repeat CentC is highly represented in the sequence of both ZB19 and ZB113. These fingerprint analysis classified BAC clone ZB19 as "class

- 119a -

1" or "HiCLoA" and BAC clone ZBl13 as "class 4" or "HiA, HiC and HiCR" (see Table 4 above). The repeated sequence CRM was also highly represented in ZBl13.

The full length sequence of CentC is set out in GenBank Accession No. AY321491 (SEQ ID NO: 76). The full length sequence of CRM is set out in GenBank
5 Accession No. AY129008 (SEQ ID NO: 77). The full length sequence of CentA is set out in Genbank Accession No. AF078917 (SEQ ID NO: 78).

Characterization of ZB19

The nucleotide sequence of ZB19 was assembled into 31 contigs with a combined trimmed length of 64 kb. ZB19 contigs numbered 1-31 correspond to SEQ ID NOS: 21-51, respectively.

5 When examining all contigs, only the two largest contigs, 30 and 31, showed significant numbers of high and low quality matches among various sequencing reads. Alternatively, all but three contigs (16, 17 and 22) show nearly complete matches to TIGR maize database entries. Large numbers of sequence regions within contig 30 have significant matches to sequence regions in contig 31. Given the small number of inconsistent
10 forward/reverse pairs, this does not suggest a misassembly but rather that both contigs 30 and 31 share large numbers of common maize sequence. Other distinct sequence similarities were evident between contigs 7 and 29, and contigs 17 and 22.

The sequence analysis of ZB19 indicated that 0.47% of the sequence is simple repeats and low complexity sequence (e.g. AT-rich, (CGA)_n, GA-rich and CT-rich), 14%
15 vector sequence, 1.15% E. coli sequence, 83% sequence is present in the TIGR maize database, 1.10% uncharacterized sequence and 28.91% CentC repeat. About 19.4 kb of the sequence was true repeat sequences, meaning those sequences are repeated within the BAC ZB19 sequence.

ZB19 has 39 simple repeat bases (0.06%) and 257 low complexity bases
20 (0.39%) contained within contigs 16, 24, 25, and 28. This low simple repeat content is summarized in Table 9.

Table 9 ZB19 Simple Repeat Content

Contig (length)	Contig Match				Simple repeat
	begin	end	length	% diverge	
ZB19.Contig16 (2303)	1572	1597	25	0	AT_rich
ZB19.Contig24 (2816)	708	747	39	17.5	(CGA) _n
ZB19.Contig25 (2997)	60	87	27	3.6	AT_rich
ZB19.Contig25 (2997)	2518	2552	34	11.4	GA-rich
ZB19.Contig28 (3308)	1121	1292	171	32.4	CT-rich
			296		
			0.47%		

25 The ZB19 contigs are set out as SEQ ID NOS: 21-51 respectively. These contigs were compared to the NCBI database at the National Institute of Health Web Site using BLAST. Results of the BLAST comparison are set out in Table 10.

Table 10 ZB19 Genbank Homology

Contig (length)	Contig Alignment			Genbank		
	begin	end	length	% id	begin	end
ZB19.Contig1 (1500)	1	1500	1501	97.07	191344	189846
ZB19.Contig2 (1708)	545	1451	918	86.06	14246	13335
ZB19.Contig3 (118)	1	118	118	100	4877	4760
ZB19.Contig4 (194)	1	194	194	100	980	1173
ZB19.Contig5 (1176)	28	1148	1122	97.15	8181	7060
ZB19.Contig6 (731)						
ZB19.Contig7 (1325)	560	1311	756	92.33	3387	2633
ZB19.Contig8 (77)						
ZB19.Contig9 (153)						
ZB19.Contig10 (1424)	23	1412	1396	91.55	42422	41034
ZB19.Contig11 (78)						
ZB19.Contig12 (1743)	561	1532	974	90.04	40272	41239
ZB19.Contig13 (1528)	853	1301	449	91.31	1	448
ZB19.Contig14 (460)						
ZB19.Contig15 (234)	1	234	234	99.57	669	436
ZB19.Contig16 (2303)						
ZB19.Contig17 (1638)						
ZB19.Contig18 (1869)	132	1719	1590	84.97	37117	35528
ZB19.Contig19 (2133)	1055	2109	1055	97.63	309950	308897
ZB19.Contig20 (1536)	93	1505	1422	83.97	400455	401871
ZB19.Contig21 (1614)	261	1556	1296	96.91	238900	237606
ZB19.Contig22 (2563)						
ZB19.Contig23 (2753)	695	2625	1938	85.19	33521	35457
ZB19.Contig23 (2753)	187	680	496	82.26	32998	33492
ZB19.Contig23 (2753)	31	148	119	82.35	170387	170505
ZB19.Contig24 (2816)	748	2788	2046	96.19	308241	306197
ZB19.Contig24 (2816)	136	707	572	94.76	308852	308282
ZB19.Contig25 (2997)	31	770	746	86.6	113462	114199
ZB19.Contig25 (2997)	849	1560	720	86.67	104886	105601

Mo17 locus bz
rust resistance rp3-1
Coliphage phi-X174
Coliphage phi-X174
Mo17 locus bz
"pCHR758mcv"
40S ribosomal protein S8
"low quality"
"E coli"
putative gag-pol
"E coli"
rust resistance rp3-1
retrotransposon
"E coli"
Coliphage phi-X174
"pCHR758mcv"
"pCHR758mcv"
Mo17 locus 9008
alpha zein gene cluster
Mo17 locus 9009
Mo17 locus 9008
"pCHR758mcv"
Mo17 locus 9008
Mo17 locus 9008
Mo17 locus 9008
alpha zein gene cluster
alpha zein gene cluster
rust resistance rp3-1
rust resistance rp3-1

AY664416
AY574035
J02482
J02482
AY664416
"NA"
AY530951
"NA"
"NA"
AF464738
"NA"
AY574035
AY574035
"NA"
J02482
"NA"
"NA"
AY664418
AF090447
AY664419
AY664418
"NA"
AY664418
AY664418
AY664418
AF090447
AF090447
AY574035
AY574035

Contig (length)	Contig Alignment				Genbank		
	begin	end	length	% id	begin	end	Accession # Homologous feature
ZB19.Contig26 (2897)	482	2861	2380	91.18	89258	86886	AY664416 Mo17 locus bz
ZB19.Contig26 (2897)	35	463	430	93.72	93544	93117	AY664416 Mo17 locus bz
ZB19.Contig27 (2845)	38	2822	2790	92.29	75093	72309	AY664416 Mo17 locus bz
ZB19.Contig28 (3308)	157	2297	2142	91.83	116223	118359	AY664413 B73 locus 9002
ZB19.Contig28 (3308)	2308	2372	65	92.31	118385	118446	AY664413 B73 locus 9002
ZB19.Contig29 (4998)	27	1161	1135	96.3	189722	188588	AY664416 Mo17 locus bz
ZB19.Contig29 (4998)	1161	1429	271	93.36	195008	195276	AY664416 Mo17 locus bz
ZB19.Contig29 (4998)	1430	2298	870	92.41	31346	30482	AY664416 Mo17 locus bz
ZB19.Contig29 (4998)	2094	4994	2903	92.08	30851	27961	AY664416 Mo17 locus bz
ZB19.Contig30 (8151)							"NA"
ZB19.Contig31 (10813)							"NA"

"CentC-like TIGR identified"
"CentC-like TIGR identified"

The identity, distribution and frequency of repeats within the centromere sequences of ZB19 are set out in Table 11. The repeats were identified by comparing the contigs to the TIGR maize database of the Institute of Genomic Research Web Site. Results of this comparison is summarized in Table 11. Percent divergence is defined as the percentage of a sequence (% of the total number of nucleotides) that is different from another sequence, with nucleotide mismatches are classified as differences.

Nearly all of contigs 4, 8, 11, 14, 15 and 18 match repeat elements without gaps apart from 155 bases on the 5' end of contig 2, a 454 bp gap in the middle of contig 16 and the 3' 1071 bp of contig 17. Sequence regions from ZB19 are identified by 75 named TIGR maize sequence database records. Among these, 23 records are CentC variants and many are multiply represented. The remaining 52 records are not CentC and are either uniquely represented or multiply represented by non-overlapping fragments.

Table 11 TIGR Maize Sequence Content in ZB19

Contig (length)	Contig Match			% diverge	Maize repeat DB	TIGR
	begin	end	length		identifier	homology
ZB19.Contig1 (1500)	1	1447	1446	15.2	SiTERTOOT0149	put. retrotrans.
ZB19.Contig1 (1500)	1152	1483	331	13	SgTERTOOT03898	put. retrotrans.
ZB19.Contig2 (1708)	30	205	175	17.8	SgCMCMOOT00130	centromere-related
ZB19.Contig2 (1708)	212	1474	1262	11.8	SmOTOT00101839	family_4154_C17
ZB19.Contig2 (1708)	1475	1700	225	17	SgTERTOOT30294	put. retrotrans.
ZB19.Contig2 (1708)	1475	1676	201	15.2	SgTERTOOT31072	put. retrotrans.
ZB19.Contig5 (1176)	10	121	111	9.8	SgTERTOOT01453	put. retrotrans.
ZB19.Contig5 (1176)	22	1148	1126	4.7	SiTERTOOT0208	put. retrotrans.
ZB19.Contig5 (1176)	977	1169	192	19.2	SgTERT00100386	put. retrotrans.
ZB19.Contig7 (1325)	30	606	576	4.9	SgTERTOOT19733	put. retrotrans.
ZB19.Contig7 (1325)	560	1325	765	5.6	SgTERTOOT00141	put. retrotrans.
ZB19.Contig10 (1424)	12	117	105	9.5	SgTERTOOT00426	put. retrotrans.
ZB19.Contig10 (1424)	23	1421	1398	7.1	SiTERTOOT0207	put. retrotrans.
ZB19.Contig12 (1743)	1	218	217	5.5	SiTERTOOT0090	put. retrotrans.
ZB19.Contig12 (1743)	219	1590	1371	10.9	SgTERTOOT03659	put. retrotrans.
ZB19.Contig12 (1743)	1589	1743	154	22.1	SgTERTOOT29480	put. retrotrans.
ZB19.Contig13 (1528)	21	105	84	28.2	SmOTOT00101761	family_3909_C1
ZB19.Contig13 (1528)	43	805	762	13.8	SmOTOT00200539	family_18_C52
ZB19.Contig13 (1528)	688	819	131	13.6	SmOTOT00200521	family_18_C35
ZB19.Contig13 (1528)	822	1528	706	5	SiTERTOOT0162	put. retrotrans.
ZB19.Contig18 (1869)	16	347	331	9.3	SmOTOT00201263	family_457_C3
ZB19.Contig18 (1869)	33	417	384	20	SmOTOT00100906	family_21444_C1
ZB19.Contig18 (1869)	418	1175	757	13.2	SiTERTOOT0109	put. retrotrans.
ZB19.Contig18 (1869)	418	1083	665	8.4	SgTERTOOT23750	put. retrotrans.
ZB19.Contig18 (1869)	986	1848	862	32	SgTERTOOT01238	put. retrotrans.
ZB19.Contig19 (2133)	1055	2109	1054	2.3	SiTERTOOT0192	put. retrotrans.
ZB19.Contig20 (1536)	23	1319	1296	18.9	SiTERTOOT0103	put. retrotrans.
ZB19.Contig20 (1536)	1320	1508	188	9.5	SmOTOT00101322	family_2963_C1

Contig (length)	Contig Match			% diverge	Maize repeat DB	TIGR
	begin	end	length		identifier	homology
ZB19.Contig21 (1614)	51	647	596	21.1	SgTERTOOT16518	put. retrotrans.
ZB19.Contig21 (1614)	131	1608	1477	5.3	SiTERTOOT0172	put. retrotrans.
ZB19.Contig23 (2753)	7	63	56	13.2	SmOTOT00200682	family_21_C7
ZB19.Contig23 (2753)	22	189	167	10.7	SmOTOT00100741	family_1920_C6
ZB19.Contig23 (2753)	190	547	357	26.1	SgTERTOOT18326	put. retrotrans.
ZB19.Contig23 (2753)	250	399	149	7.5	SmOTOT00200636	family_20_C20
ZB19.Contig23 (2753)	399	632	233	9	SmOTOT00201628	family_73_C1
ZB19.Contig23 (2753)	640	1305	665	32.8	SgTERTOOT02327	put. retrotrans.
ZB19.Contig23 (2753)	710	1116	406	32.4	SgTERTOOT26280	put. retrotrans.
ZB19.Contig23 (2753)	804	1364	560	31.6	SiTERTOOT0139	put. retrotrans.
ZB19.Contig23 (2753)	1050	1247	197	14.1	SmOTOT00201653	family_766_C5
ZB19.Contig23 (2753)	1271	1403	132	15.8	SmOTOT00201649	family_766_C1
ZB19.Contig23 (2753)	1405	1691	286	11.8	SmOTOT00201691	family_79_C1
ZB19.Contig23 (2753)	1756	2702	946	35.8	SgTERTOOT00119	put. retrotrans.
ZB19.Contig23 (2753)	1903	2039	136	13.1	SmOTOT00200145	family_1251_C1
ZB19.Contig23 (2753)	2230	2626	396	31.8	SgTERTOOT00404	put. retrotrans.
ZB19.Contig24 (2816)	14	2788	2774	6	SiTERTOOT0192	put. retrotrans.
ZB19.Contig25 (2997)	22	2767	2745	21.6	SiTERTOOT0310	put. retrotrans.
ZB19.Contig25 (2997)	2498	2970	472	25.4	SgTERTOOT26929	put. retrotrans.
ZB19.Contig25 (2997)	2875	2981	106	15.9	SgTERTOOT08404	put. retrotrans.
ZB19.Contig26 (2897)	28	463	435	2.8	SgTERTOOT22255	put. retrotrans.
ZB19.Contig26 (2897)	480	2871	2391	3.7	SiTERTOOT0162	put. retrotrans.
ZB19.Contig27 (2845)	7	2823	2816	6.3	SiTERTOOT0162	put. retrotrans.
ZB19.Contig28 (3308)	27	2368	2341	18.9	SiTERTOOT0157	put. retrotrans.
ZB19.Contig29 (4998)	29	1161	1132	21.8	SiTERTOOT0310	put. retrotrans.
ZB19.Contig29 (4998)	1162	1429	267	2.6	SgTERTOOT17469	put. retrotrans.
ZB19.Contig29 (4998)	1430	4994	3564	8.7	SiTERTOOT0170	put. retrotrans.
ZB19.Contig29 (4998)	2032	4998	2966	1.9	SiTERTOOT0172	put. retrotrans.
ZB19.Contig30 (8151)	74	233	159	31.8	SiTERTOOT0296	put. retrotrans.
ZB19.Contig30 (8151)	368	500	132	6	SgCMCM00200161	CentC
ZB19.Contig30 (8151)	501	655	154	2.6	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	656	810	154	5.8	SgCMCM00200282	CentC
ZB19.Contig30 (8151)	811	966	155	5.8	SgCMCM00200175	CentC
ZB19.Contig30 (8151)	967	1121	154	3.2	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	1122	1276	154	6.5	SgCMCM00200356	CentC
ZB19.Contig30 (8151)	1277	1431	154	5.8	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	1432	1587	155	3.9	SgCMCM00200282	CentC
ZB19.Contig30 (8151)	1588	1743	155	5.8	SgCMCM00200175	CentC
ZB19.Contig30 (8151)	1744	1898	154	4.5	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	1899	2053	154	3.2	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	2054	2208	154	3.9	SgCMCM00200356	CentC
ZB19.Contig30 (8151)	2209	2363	154	4.5	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	2364	2518	154	3.2	SgCMCM00200356	CentC
ZB19.Contig30 (8151)	2519	2674	155	3.9	SgCMCM00200228	CentC
ZB19.Contig30 (8151)	2675	2829	154	3.2	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	2830	2984	154	3.2	SgCMCM00200145	CentC
ZB19.Contig30 (8151)	2985	3139	154	3.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	3140	3295	155	1.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	3296	3452	156	1.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	3453	3607	154	3.9	SgCMCM00200034	CentC

ZB19.Contig30 (8151)	3608	3763	155	1.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	3764	3918	154	4.5	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	3919	4073	154	1.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	4074	4227	153	2.6	SgCMCM00200092	CentC
ZB19.Contig30 (8151)	4228	4382	154	3.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	4383	4538	155	2.6	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	4539	4693	154	3.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	4694	4848	154	2.6	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	4849	5003	154	3.9	SgCMCM00200356	CentC
ZB19.Contig30 (8151)	5004	5158	154	5.8	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	5159	5314	155	4.5	SgCMCM00200526	CentC
ZB19.Contig30 (8151)	5315	5468	153	3.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	5469	5624	155	5.8	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	5625	5779	154	3.2	SgCMCM00200228	CentC
ZB19.Contig30 (8151)	5780	5934	154	3.2	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	5935	6089	154	3.9	SgCMCM00200145	CentC
ZB19.Contig30 (8151)	6090	6244	154	3.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	6245	6399	154	4.5	SgCMCM00200145	CentC
ZB19.Contig30 (8151)	6400	6555	155	3.2	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	6556	6710	154	4.5	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	6711	6865	154	1.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	6866	7020	154	3.9	SgCMCM00200356	CentC
ZB19.Contig30 (8151)	7021	7175	154	4.5	SgCMCM00200009	CentC
ZB19.Contig30 (8151)	7176	7330	154	2.6	SgCMCM00200228	CentC
ZB19.Contig30 (8151)	7331	7485	154	5.2	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	7486	7640	154	5.2	SgCMCM00200026	CentC
ZB19.Contig30 (8151)	7642	7796	154	1.9	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	7797	7952	155	4.5	SgCMCM00200034	CentC
ZB19.Contig30 (8151)	7954	8107	153	4.6	SgCMCM00200372	CentC
ZB19.Contig31 (10813)	23	167	144	4.8	SgCMCM00200030	CentC
ZB19.Contig31 (10813)	168	322	154	3.9	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	323	477	154	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	478	632	154	3.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	633	788	155	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	789	943	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	944	1098	154	3.9	SgCMCM00200145	CentC
ZB19.Contig31 (10813)	1100	1254	154	3.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	1255	1409	154	2.6	SgCMCM00200228	CentC
ZB19.Contig31 (10813)	1410	1565	155	5.8	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	1566	1720	154	3.9	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	1721	1875	154	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	1876	2031	155	4.5	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	2032	2186	154	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	2187	2342	155	5.1	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	2343	2497	154	3.9	SgCMCM00200090	CentC
ZB19.Contig31 (10813)	2498	2649	151	3.3	SgCMCM00200260	CentC
ZB19.Contig31 (10813)	2650	2804	154	1.9	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	2805	2959	154	1.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	2960	3114	154	4.5	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	3115	3270	155	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	3271	3425	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	3426	3580	154	2.6	SgCMCM00200034	CentC

ZB19.Contig31 (10813)	3581	3734	153	3.9	SgCMCM00200159	CentC
ZB19.Contig31 (10813)	3735	3890	155	4.5	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	3891	4045	154	1.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	4047	4201	154	5.2	SgCMCM00200026	CentC
ZB19.Contig31 (10813)	4202	4356	154	3.2	SgCMCM00200090	CentC
ZB19.Contig31 (10813)	4357	4513	156	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	4514	4668	154	6.5	SgCMCM00200179	CentC
ZB19.Contig31 (10813)	4669	4823	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	4824	4978	154	5.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	4979	5133	154	3.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	5135	5289	154	5.2	SgCMCM00200026	CentC
ZB19.Contig31 (10813)	5290	5444	154	3.2	SgCMCM00200090	CentC
ZB19.Contig31 (10813)	5445	5601	156	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	5602	5756	154	6.5	SgCMCM00200179	CentC
ZB19.Contig31 (10813)	5757	5911	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	5912	6066	154	5.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	6068	6221	153	3.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	6222	6376	154	4.5	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	6377	6533	156	7.1	SgCMCM00200282	CentC
ZB19.Contig31 (10813)	6534	6686	152	5.9	SgCMCM00200026	CentC
ZB19.Contig31 (10813)	6687	6840	153	4.5	SgCMCM00200090	CentC
ZB19.Contig31 (10813)	6841	6995	154	3.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	6996	7150	154	2.6	SgCMCM00200104	CentC
ZB19.Contig31 (10813)	7151	7305	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	7306	7460	154	3.2	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	7461	7616	155	4.5	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	7617	7771	154	3.9	SgCMCM00200145	CentC
ZB19.Contig31 (10813)	7772	7925	153	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	7926	8081	155	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	8083	8236	153	4.5	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	8237	8392	155	5.8	SgCMCM00200282	CentC
ZB19.Contig31 (10813)	8393	8547	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	8548	8703	155	3.9	SgCMCM00200258	CentC
ZB19.Contig31 (10813)	8704	8859	155	3.9	SgCMCM00200058	CentC
ZB19.Contig31 (10813)	8860	9015	155	4.5	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	9016	9170	154	3.9	SgCMCM00200090	CentC
ZB19.Contig31 (10813)	9171	9324	153	3.9	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	9325	9479	154	4.5	SgCMCM00200228	CentC
ZB19.Contig31 (10813)	9480	9634	154	3.2	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	9635	9789	154	1.3	SgCMCM00200092	CentC
ZB19.Contig31 (10813)	9790	9917	127	3.1	SgCMCM00200234	CentC
ZB19.Contig31 (10813)	9914	10011	97	2.1	SgCMCM00200032	CentC
ZB19.Contig31 (10813)	10012	10165	153	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	10167	10320	153	3.2	SgCMCM00200356	CentC
ZB19.Contig31 (10813)	10325	10476	151	2.6	SgCMCM00200150	CentC
ZB19.Contig31 (10813)	10477	10631	154	2.6	SgCMCM00200034	CentC
ZB19.Contig31 (10813)	10632	10785	153	4.5	SgCMCM00200092	CentC

The contigs of ZB19 consist of sequence that is not repeated within the library apart from all of contig 31 and all but the very 5' end of contig 30 and perhaps a small ~400 base repeat in the middle of contig 29. The repeat regions extend approximately 8 and 10 kb

in contigs 30 and 31, respectively. Since the repeated regions are apparent both when compared to self and the reverse complement, the larger repeat region consists of many smaller repeat regions that occur both in the forward and reverse direction.

The consensus sequence of the CentC repeat present in ZB19 is set out as SEQ ID NO: 70. The variants of the CentC repeats present in ZB19 are set out in Table 12 where the most common base is indicated. Where the most common base occurs less than 60% of the time, the percent occurrence of each base is reported.

Table 12 CentC consensus and Variation in ZB19

1	T	60	A (A:99/T:1)	119	C (A:2/C:96/T:3)
2	G (C:4/G:96)	61	T	120	T (A:3/T:97)
3	G (A:9/G:91)	62	G (A:4/G:96)	121	A (A:90/G:1/T:9)
4	T (A:3/T:97)	63	G	122	A (A:99/T:1)
5	T (C:4/T:96)	64	G (A:3/G:89/T:8)	123	A
6	C (C:81/G:3/T:16)	65	T	124	G (C:1/G:99)
7	C (A:3/C:85/G:1/T:11)	66	G	125	T (C:1/T:99)
8	G (A:5/G:95)	67	A	126	A
9	G (A:2/G:97)	68	C (C:96/T:4)	127	G (A:4/G:96)
10	T (G:1/T:99)	69	G (A:4/G:96)	128	T (G:1/T:99)
11	G (G:98/T:2)	70	T (C:1/T:99)	129	G (A:2/G:92/T:6)
12	G (C:1/G:99)	71	G	130	G (G:72/T:28)
13	C (A:1/C:99)	72	C (C:93/G:2/T:5)	131	A
14	A (A:94/C:3/G:2)	73	G (A:4/G:90/T:6)	132	T (G:6/T:94)
15	A (A:97/C:2/G:1)	74	G (A:6/G:94)	133	T (C:1/T:99)
16	A (A:98/G:2)	75	C	134	G (C:3/G:97)
17	A	76	A (A:95/G:5)	135	G (A:2/C:3/G:95/T:1)
18	A (A:99/C:1/-:1)	77	A	136	G (G:97/T:3)
19	C (A:2/C:97/T:1)	78	C (A:1/C:99)	137	C (A:2/C:98)
20	T (A:4/C:3/T:94)	79	G (A:3/G:97)	138	A (A:97/C:2/G:1)
21	C (C:91/T:9)	80	A	139	T
22	G (A:7/G:90/T:3)	81	A	140	G (A:3/G:96/T:1)
23	T (C:2/T:98)	82	A	141	T
24	G (C:3/G:93/-:4)	83	T	142	T
25	C (A:4/C:90/G:1/T:4)	84	T (A:1/G:1/T:99)	143	C (C:94/T:6)
26	H(A:24/C:3/T:55/-:18)	85	G (A:4/G:96)	144	G (A:4/G:94/T:3)
27	T (A:2/G:3/T:95)	86	C (A:3/C:94/T:3)	145	T (T:99)
28	T (A:1/G:1/T:99)	87	G (A:3/G:97)	146	T
29	D(A:12/G:27/T:53/-:8)	88	C (A:2/C:91/G:1/T:5/-:2)	147	G
30	W(A:48/T:52)	89	G (A:4/G:96)	148	C (C:97/T:3)
31	M(A:55/C:1/-:44)	90	A (A:98/T:2)	149	G (A:2/G:97/-:1)
32	T (A:4/C:1/T:92/-:3)	91	A	150	A (A:96/C:4)
33	T(T:1/-:99)	92	A (A:99/C:1)	151	A (A:97/T:3)
34	G	93	C	152	A
35	C (C:99/T:1)	94	C (A:4/C:96)	153	A
36	A	95	A	154	A
37	C (A:3/C:97)	96	C (A:1/C:96/T:3)	155	C (A:4/C:67/G:3/T:25/-:1)
38	Y(C:26/T:3/-:71)	97	C (A:3/C:92/T:5)	156	G (A:13/G:84/-:3)
39	C (C:99/-:1)	98	C (A:1/C:99)	157	A
40	C (C:93/G:1/T:1/-:4)	99	C (A:25/C:74/-:1)	158	A (A:98/T:2)
41	C (C:79/G:2/T:2/-:18)	100	A	159	G (G:99/T:1)
42	G (A:3/C:1/G:96)	101	A	160	A
43	A (A:99/G:1)	102	C(C:3/-:97)	161	A
44	C (A:3/C:95/T:3)	103	A	162	A (A:99/G:1)
45	A	104	C (A:1/C:99/T:1)	163	T (C:1/G:1/T:97)
46	C (A:1/C:96/T:3)	105	A (A:97/T:3)	164	G (C:1/G:99)
47	C	106	A(A:1/-:99)	165	R(A:45/G:55)
48	C (C:97/G:1/T:3)	107	T	166	T

49	G(G:1/-:99)	108	G (A:1/G:94/T:4)	167	T
50	G (A:3/G:92/T:5)	109	A (A:64/C:1/G:26/T:8)	168	C
51	T (A:2/T:98)	110	G	169	Y(C:41/T:59)
52	T (T:99)	111	T (G:1/T:99)	170	G (A:30/G:70)
53	T (G:2/T:98)	112	T	171	G (A:14/G:86)
54	T (C:1/T:98/-:1)	113	T (T:98/-:2)	172	T
55	C (C:91/T:9/-:1)	114	T (T:98/-:2)		
56	G (A:1/G:99)	115	G (A:3/G:94/T:3/-:1)		
57	G (A:10/G:90)	116	G		
58	G(G:1/-:99)	117	A (A:99/G:1)	153	mean length
59	A (A:97/T:3)	118	C	3.8	std

Characterization of ZB113

The nucleotide sequence of ZB113 was assembled into 18 contigs with a combined trimmed length of 90 kb. ZB113 contigs 1-18 correspond to SEQ ID NOS: 52-69, respectively.

All but three contigs (9, 12, and 13c) of ZB113 showed significant numbers of mostly high and some low quality matches to various sequencing reads; and all but contigs 12 and 13 showed significant matches to TIGR maize database entries. With the large numbers of inconsistent forward/reverse pairs present within the contigs there may be a number of misassemblies present. Notably, contig 17 might be falsely assembled sequence fragments belonging to contigs 14, 15, and 18. Many large regions of similarity exist between the contigs. Notably, an approximately 1.3 kb region on the the 3' end of contig 18 is present several times on the 5' end of contig 18 as well as covering nearly all of contigs 15 and 17, and the 3' half of contig 14.

The sequence analysis of ZB19 indicated that 0.23% of the sequence is simple repeats and low complexity sequence (e.g. AT-rich, T-rich and (TTTTC)n), 17% vector sequence, 78% sequence is present in the TIGR maize database, 4.40% uncharacterized sequence, 47.55% CentC repeat, 0.57% CentA repeat and 31.73% of CRM repeat. About 42.3 kb of the sequence was true repeat sequence, meaning those sequences are repeated within the BAC ZB19 sequence.

ZB113 has 64 simple repeat bases (0.07%) and 145 low complexity bases (0.16%) contained within contigs 12, 13, 16, and 18. This low simple repeat content is summarized in Table 13.

Table 13 ZB113 Simple Repeat Content

Contig (length)	Contig Match				Simple repeat
	begin	end	length	% diverge	
ZB113.Contig12 (5594)	1466	1488	22	0	AT_rich
ZB113.Contig12 (5594)	2391	2423	32	0	AT_rich
ZB113.Contig13 (5111)	3730	3755	25	0	AT_rich
ZB113.Contig16 (15540)	3933	3995	62	22.2	T-rich
ZB113.Contig18 (20048)	16200	16263	63	11.3	(TTTTTC)n
204					

5 The ZB113 contigs are set out as SEQ ID NOS: 52-69, respectively. These contigs were compared to the NCBI database at the National Institute of Health Web Site using BLAST. Results of the BLAST comparison are set out in Table 14.

Table 14 ZB113 GenBank Homology

Contig (length)	Contig Alignment			Genbank		
	begin	end	length	% id	begin	end
ZB113.Contig1 (864)	366	547	182	96.15	27727	27546
ZB113.Contig2 (835)	587	703	117	99.15	21738	21622
ZB113.Contig2 (835)	743	835	93	91.4	21581	21491
ZB113.Contig2 (835)	234	279	46	100	27870	27825
ZB113.Contig2 (835)	168	215	48	97.92	28900	28853
ZB113.Contig2 (835)	317	344	28	100	27917	27890
ZB113.Contig3 (903)	137	732	598	97.99	1	597
ZB113.Contig4 (1110)					XM_367004	
ZB113.Contig5 (586)					"NA"	"CentC-like TIGR identified"
ZB113.Contig6 (857)					"NA"	"pCHR758mcv"
ZB113.Contig7 (119)					"NA"	"pCHR758mcv"
ZB113.Contig8 (1510)	20	95	76	96.05	339	339
ZB113.Contig9 (1785)					AY046113	yeast 26S ribosomal RNA
ZB113.Contig10 (867)					"NA"	"CentC-like TIGR identified"
ZB113.Contig11 (3369)					"NA"	"pCHR758mcv"
ZB113.Contig12 (5594)					"NA"	"pCHR758mcv"
ZB113.Contig13 (5111)					"NA"	Tn10
ZB113.Contig14 (8559)					"NA"	"CentC-like TIGR identified"
ZB113.Contig14 (8559)	57	4643			"NA"	"pCHR758mcv"
ZB113.Contig14 (8559)	4643	7938			"NA"	"CRM retrotrans-like TIGR identified"
ZB113.Contig14 (8559)	7937	8511			"NA"	"CentC-like TIGR identified"
ZB113.Contig15 (10771)					"NA"	"CRM retrotrans-like TIGR identified"
ZB113.Contig16 (15540)	37	9786			"NA"	"CentA-like TIGR identified"
ZB113.Contig16 (15540)	9589	10101			"NA"	"CRM retrotrans-like TIGR identified"
ZB113.Contig16 (15540)	9916	12079			"NA"	"pCHR758mcv"
ZB113.Contig16 (15540)	12080	12436			"NA"	"CRM retrotrans-like TIGR identified"
ZB113.Contig16 (15540)	12533	14274			"NA"	"pCHR758mcv"
ZB113.Contig16 (15540)	14274	15203			"NA"	"CRM retrotrans-like TIGR identified"
ZB113.Contig16 (15540)	15203	15404			"NA"	"CRM retrotrans-like TIGR identified"

The identity, distribution and frequency of repeats within the centromere sequences of ZB113 is set out in Table 15. The contigs were also compared to the TIGR maize database at the Institute of Genomic Research Web Site. Results of this comparison are summarized in Table 15. Sequence regions from ZB113 are identified by 54 named
5 TIGR maize sequence database records. Among these, 38 records are CentC variants and many are multiply represented. The remaining 16 records are not CentC and are either uniquely represented or multiply represented by non-overlapping fragments, apart from SmOTOT00200141, SmOTOT00200215, SmOTOT00200264, SmOTOT00200480, SmOTOT00201588.

Table 15 ZB113 TIGR Maize Sequence Content

Contig (length)	Contig Match		% diverge	Maize Repeat DB		TIGR
	begin	end		length	Identifier	
ZB113.Contig2 (835)	156	384	228	SmOTOT00200480	family_1868_C1	
ZB113.Contig2 (835)	390	698	308	SmOTOT00200215	family_1380_C2	
ZB113.Contig2 (835)	700	748	48	SmOTOT00200303	family_14706_C1	
ZB113.Contig2 (835)	745	835	90	SmOTOT00200264	family_1431_C3	
ZB113.Contig4 (1110)	78	233	155	SgCMCM00200034	CentC	
ZB113.Contig4 (1110)	234	369	135	SgCMCM00200034	CentC	
ZB113.Contig4 (1110)	370	524	154	SgCMCM00200228	CentC	
ZB113.Contig4 (1110)	525	678	153	SgCMCM00200173	CentC	
ZB113.Contig4 (1110)	679	833	154	SgCMCM00200034	CentC	
ZB113.Contig4 (1110)	834	988	154	SgCMCM00200228	CentC	
ZB113.Contig4 (1110)	989	1060	71	SgCMCM00200173	CentC	
ZB113.Contig8 (1510)	5	37	32	SgCMCM00200269	CentC	
ZB113.Contig8 (1510)	38	192	154	SgCMCM00200034	CentC	
ZB113.Contig8 (1510)	193	346	153	SgCMCM00200228	CentC	
ZB113.Contig8 (1510)	347	501	154	SgCMCM00200034	CentC	
ZB113.Contig8 (1510)	502	656	154	SgCMCM00200228	CentC	
ZB113.Contig8 (1510)	657	811	154	SgCMCM00200099	CentC	
ZB113.Contig8 (1510)	812	966	154	SgCMCM00200034	CentC	
ZB113.Contig8 (1510)	967	1122	155	SgCMCM00200530	CentC	
ZB113.Contig8 (1510)	1123	1277	154	SgCMCM00200058	CentC	
ZB113.Contig8 (1510)	1279	1433	154	SgCMCM00200034	CentC	
ZB113.Contig8 (1510)	1434	1471	37	SgCMCM00200214	CentC	
ZB113.Contig8 (1510)	1434	1467	33	SgCMCM00200257	CentC	
ZB113.Contig8 (1510)	1434	1460	26	SgCMCM00200350	CentC	
ZB113.Contig10 (867)	1	831	830	SmOTOT00102689	family_7207_C1	
ZB113.Contig11 (3369)	52	177	125	SgCMCM00200034	CentC	
ZB113.Contig11 (3369)	178	332	154	SgCMCM00200034	CentC	
ZB113.Contig11 (3369)	333	486	153	SgCMCM00200017	CentC	

ZB113.Contig11 (3369)	487	641	154	1.9	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	642	797	155	1.3	SgCMCM00200099	CentC
ZB113.Contig11 (3369)	798	954	156	3.8	SgCMCM00200014	CentC
ZB113.Contig11 (3369)	958	1115	157	5.8	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	1116	1270	154	2.6	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	1271	1425	154	1.9	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	1426	1579	153	1.9	SgCMCM00200228	CentC
ZB113.Contig11 (3369)	1580	1735	155	1.9	SgCMCM00200228	CentC
ZB113.Contig11 (3369)	1736	1890	154	2.6	SgCMCM00200228	CentC
ZB113.Contig11 (3369)	1891	2046	155	0.6	SgCMCM00200095	CentC
ZB113.Contig11 (3369)	2047	2201	154	1.3	SgCMCM00200228	CentC
ZB113.Contig11 (3369)	2202	2358	156	1.9	SgCMCM00200026	CentC
ZB113.Contig11 (3369)	2359	2513	154	2.6	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	2514	2669	155	2.6	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	2670	2824	154	0	SgCMCM00200228	CentC
ZB113.Contig11 (3369)	2825	2979	154	1.9	SgCMCM00200228	CentC
ZB113.Contig11 (3369)	2980	3135	155	0.6	SgCMCM00200095	CentC
ZB113.Contig11 (3369)	3136	3290	154	2.6	SgCMCM00200034	CentC
ZB113.Contig11 (3369)	3291	3328	37	13.2	SgCMCM00200214	CentC
ZB113.Contig11 (3369)	3291	3324	33	8.8	SgCMCM00200257	CentC
ZB113.Contig11 (3369)	3291	3317	26	0	SgCMCM00200095	CentC
ZB113.Contig14 (8559)	57	1862	1805	0.6	SiCMCMOOT0036	CRM
ZB113.Contig14 (8559)	1895	2046	151	1.3	SmOTOT00201588	family_6912_C1
ZB113.Contig14 (8559)	2035	2541	506	24.5	SiCMCMOOT0036	CRM
ZB113.Contig14 (8559)	2524	3249	725	1.4	SmOTOT00200264	family_1431_C3
ZB113.Contig14 (8559)	3179	3581	402	23	SmOTOT00101933	family_4330_C5
ZB113.Contig14 (8559)	3296	3604	308	1.9	SmOTOT00200215	family_1380_C2
ZB113.Contig14 (8559)	3610	4066	456	2.5	SmOTOT00200480	family_1868_C1
ZB113.Contig14 (8559)	3968	4277	309	25.4	SiCMCMOOT0036	CRM
ZB113.Contig14 (8559)	4245	4643	398	0.8	SmOTOT00200141	family_1241_C3
ZB113.Contig14 (8559)	4643	5076	433	3	SgTERTOOT02246	put. retrotrans.
ZB113.Contig14 (8559)	5077	5234	157	2.6	SgCMCM00200102	CentC
ZB113.Contig14 (8559)	5235	5390	155	0.6	SgCMCM00200099	CentC

ZB113.Contig14 (8559)	5391	5545	154	1.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	5546	5700	154	1.9	SgCMCM00200228	CentC
ZB113.Contig14 (8559)	5701	5855	154	1.9	SgCMCM00200228	CentC
ZB113.Contig14 (8559)	5856	6011	155	1.3	SgCMCM00200095	CentC
ZB113.Contig14 (8559)	6012	6166	154	3.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	6167	6321	154	1.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	6322	6476	154	0.7	SgCMCM00200228	CentC
ZB113.Contig14 (8559)	6477	6631	154	1.9	SgCMCM00200228	CentC
ZB113.Contig14 (8559)	6632	6787	155	0.6	SgCMCM00200095	CentC
ZB113.Contig14 (8559)	6788	6942	154	1.9	SgCMCM00200228	CentC
ZB113.Contig14 (8559)	6943	7097	154	1.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	7098	7252	154	1.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	7253	7409	156	2.5	SgCMCM00200026	CentC
ZB113.Contig14 (8559)	7410	7565	155	0.6	SgCMCM00200095	CentC
ZB113.Contig14 (8559)	7566	7720	154	1.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	7721	7875	154	1.9	SgCMCM00200034	CentC
ZB113.Contig14 (8559)	7876	7938	62	3.2	SgCMCM00200356	CentC
ZB113.Contig14 (8559)	7937	8511	574	11.5	SiCMCMOOT0036	put. retrotrans.
ZB113.Contig16 (15540)	37	5452	5415	1.7	SiCMCMOOT0036	put. retrotrans.
ZB113.Contig16 (15540)	5453	8994	3541	25.8	SiCMCMOOT0036	put. retrotrans.
ZB113.Contig16 (15540)	8955	9353	398	0.8	SmOTOT00200141	family_1241_C3
ZB113.Contig16 (15540)	9317	9520	203	28.6	SiCMCMOOT0033	centromeric repeat
ZB113.Contig16 (15540)	9532	9786	254	17.7	SgTERTOOT02246	put. retrotrans.
ZB113.Contig16 (15540)	9589	10101	512	17.1	SiTERTOOT0090	CentA
ZB113.Contig16 (15540)	9916	12079	2163	21.6	SiCMCMOOT0036	CRM
ZB113.Contig16 (15540)	12533	12567	34	6.1	SmOTOT00101153	family_26265_C1
ZB113.Contig16 (15540)	12537	12812	275	17.4	SmOTOT00102620	family_68_C5
ZB113.Contig16 (15540)	12815	13013	198	18.6	SmOTOT00100244	family_1167_C2
ZB113.Contig15 (10771)	62	216	154	1.9	SgCMCM00200224	CentC
ZB113.Contig15 (10771)	217	371	154	1.3	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	372	526	154	3.2	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	527	682	155	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	683	837	154	1.9	SgCMCM00200228	CentC

ZB113.Contig15 (10771)	838	994	156	1.9	SgCMCM00200026	CentC
ZB113.Contig15 (10771)	995	1150	155	1.3	SgCMCM00200095	CentC
ZB113.Contig15 (10771)	1151	1305	154	3.2	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	1306	1460	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	1461	1615	154	1.9	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	1616	1770	154	1.3	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	1771	1926	155	0.6	SgCMCM00200095	CentC
ZB113.Contig15 (10771)	1927	2081	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	2082	2236	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	2237	2391	154	2.6	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	2392	2547	155	0.6	SgCMCM00200095	CentC
ZB113.Contig15 (10771)	2548	2702	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	2703	2857	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	2858	3012	154	0.7	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	3013	3167	154	1.3	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	3168	3323	155	0.6	SgCMCM00200095	CentC
ZB113.Contig15 (10771)	3324	3478	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	3479	3633	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	3634	3788	154	0.7	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	3789	3943	154	1.3	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	3944	4099	155	1.3	SgCMCM00200095	CentC
ZB113.Contig15 (10771)	4100	4254	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	4255	4409	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	4410	4564	154	0.7	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	4565	4719	154	0.7	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	4720	4874	154	1.9	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	4875	5029	154	1.9	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	5030	5185	155	1.3	SgCMCM00200095	CentC
ZB113.Contig15 (10771)	5186	5340	154	3.2	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	5341	5495	154	1.3	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	5496	5649	153	1.3	SgCMCM00200002	CentC
ZB113.Contig15 (10771)	5650	5806	156	1.3	SgCMCM00200026	CentC
ZB113.Contig15 (10771)	5807	5961	154	1.9	SgCMCM00200034	CentC

ZB113.Contig15 (10771)	5962	6115	153	0	SgCMCM00200530	CentC
ZB113.Contig15 (10771)	6116	6270	154	4.5	SgCMCM00200058	CentC
ZB113.Contig15 (10771)	6272	6426	154	3.2	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	6427	6581	154	1.3	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	6582	6737	155	1.9	SgCMCM00200099	CentC
ZB113.Contig15 (10771)	6738	6891	153	1.3	SiCMCM00200001	CentC
ZB113.Contig15 (10771)	6892	7048	156	1.3	SgCMCM00200026	CentC
ZB113.Contig15 (10771)	7049	7203	154	1.3	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	7204	7358	154	1.3	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	7359	7513	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	7514	7668	154	1.3	SgCMCM00200017	CentC
ZB113.Contig15 (10771)	7669	7825	156	1.3	SgCMCM00200026	CentC
ZB113.Contig15 (10771)	7826	7980	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	7981	8135	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	8136	8290	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	8291	8445	154	1.3	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	8446	8600	154	1.9	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	8601	8757	156	0.6	SgCMCM00200026	CentC
ZB113.Contig15 (10771)	8758	8912	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	8913	9067	154	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	9068	9223	155	2.6	SgCMCM00200228	CentC
ZB113.Contig15 (10771)	9224	9378	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	9379	9533	154	1.9	SgCMCM00200099	CentC
ZB113.Contig15 (10771)	9534	9688	154	2.6	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	9689	9842	153	0	SgCMCM00200530	CentC
ZB113.Contig15 (10771)	9843	9997	154	3.9	SgCMCM00200058	CentC
ZB113.Contig15 (10771)	9999	10153	154	3.2	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	10154	10308	154	1.3	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	10309	10464	155	1.9	SgCMCM00200034	CentC
ZB113.Contig15 (10771)	10465	10619	154	3.2	SgCMCM00200044	CentC
ZB113.Contig15 (10771)	10620	10759	139	9.1	SgCMCM00200311	CentC
ZB113.Contig15 (10771)	10620	10726	106	3.7	SgCMCM00200152	CentC
ZB113.Contig17 (14443)	1	91	90	9.9	SgCMCM00200454	CentC

ZB113.Contig17 (14443)	13	91	78	0	SgCMCM00200530	CentC
ZB113.Contig17 (14443)	92	246	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	247	402	155	1.3	SgCMCM00200099	CentC
ZB113.Contig17 (14443)	403	561	158	2.6	SgCMCM00200102	CentC
ZB113.Contig17 (14443)	562	716	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	717	870	153	1.3	SgCMCM00200173	CentC
ZB113.Contig17 (14443)	871	1025	154	1.9	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	1026	1161	135	4.4	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	1162	1317	155	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	1318	1472	154	1.3	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	1473	1627	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	1628	1782	154	0.7	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	1783	1937	154	1.3	SgCMCM00200224	CentC
ZB113.Contig17 (14443)	1938	2093	155	2.6	SgCMCM00200014	CentC
ZB113.Contig17 (14443)	2094	2249	155	1.3	SgCMCM00200102	CentC
ZB113.Contig17 (14443)	2250	2405	155	1.9	SgCMCM00200079	CentC
ZB113.Contig17 (14443)	2406	2560	154	2.6	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	2561	2715	154	0	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	2716	2870	154	3.2	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	2871	3025	154	1.9	SgCMCM00200224	CentC
ZB113.Contig17 (14443)	3026	3180	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	3181	3337	156	1.9	SgCMCM00200026	CentC
ZB113.Contig17 (14443)	3338	3492	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	3493	3647	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	3648	3802	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	3803	3957	154	3.2	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	3958	4114	156	0.6	SgCMCM00200026	CentC
ZB113.Contig17 (14443)	4115	4269	154	0.7	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	4270	4424	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	4425	4579	154	6.5	SgCMCM00200224	CentC
ZB113.Contig17 (14443)	4580	4734	154	0.7	SgCMCM00200224	CentC
ZB113.Contig17 (14443)	4735	4894	159	0	SgCMCM00200500	CentC
ZB113.Contig17 (14443)	4895	5049	154	2.6	SgCMCM00200068	CentC

ZB113.Contig17 (14443)	5050	5204	154	0.7	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	5205	5359	154	0.7	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	5360	5515	155	1.3	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	5516	5670	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	5671	5826	155	0.6	SgCMCM00200095	CentC
ZB113.Contig17 (14443)	5827	5981	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	5982	6136	154	1.9	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	6137	6292	155	0.6	SgCMCM00200095	CentC
ZB113.Contig17 (14443)	6293	6447	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	6448	6602	154	0.7	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	6603	6757	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	6758	6912	154	1.3	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	6913	7067	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	7068	7224	156	1.3	SgCMCM00200026	CentC
ZB113.Contig17 (14443)	7225	7379	154	0	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	7380	7533	153	1.9	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	7534	7688	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	7689	7843	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	7844	7998	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	8000	8154	154	3.9	SgCMCM00200058	CentC
ZB113.Contig17 (14443)	8155	8310	155	1.3	SgCMCM00200530	CentC
ZB113.Contig17 (14443)	8311	8465	154	3.2	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	8466	8621	155	0.6	SgCMCM00200099	CentC
ZB113.Contig17 (14443)	8622	8780	158	2.6	SgCMCM00200102	CentC
ZB113.Contig17 (14443)	8781	8935	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	8936	9089	153	0	SgCMCM00200173	CentC
ZB113.Contig17 (14443)	9090	9244	154	0.7	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	9245	9399	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	9400	9554	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	9555	9709	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	9710	9864	154	0.7	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	9865	10017	152	1.3	SgCMCM00200026	CentC
ZB113.Contig17 (14443)	10018	10171	153	1.3	SgCMCM00200002	CentC

ZB113.Contig17 (14443)	10172	10327	155	1.3	SgCMCM00200099	CentC
ZB113.Contig17 (14443)	10328	10482	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	10483	10637	154	0.7	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	10638	10794	156	1.3	SgCMCM00200026	CentC
ZB113.Contig17 (14443)	10795	10948	153	1.3	SgCMCM00200005	CentC
ZB113.Contig17 (14443)	10949	11104	155	1.3	SgCMCM00200099	CentC
ZB113.Contig17 (14443)	11105	11259	154	1.3	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	11260	11414	154	2.6	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	11416	11570	154	3.9	SgCMCM00200058	CentC
ZB113.Contig17 (14443)	11571	11724	153	0	SgCMCM00200530	CentC
ZB113.Contig17 (14443)	11725	11879	154	1.9	SgCMCM00200034	CentC
ZB113.Contig17 (14443)	11880	12035	155	0.6	SgCMCM00200099	CentC
ZB113.Contig17 (14443)	12036	12194	158	2.6	SgCMCM00200102	CentC
ZB113.Contig17 (14443)	12195	12351	156	3.2	SgCMCM00200026	CentC
ZB113.Contig17 (14443)	12352	12506	154	0.7	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	12507	12661	154	1.3	SgCMCM00200228	CentC
ZB113.Contig17 (14443)	12662	12816	154	3.9	SgCMCM00200224	CentC
ZB113.Contig17 (14443)	12817	12911	94	3.2	SgCMCM00200025	CentC
ZB113.Contig17 (14443)	12911	13372	461	3	SgTERTOOT02573	put. retrotrans.
ZB113.Contig18 (20048)	57	212	155	0.6	SgCMCM00200095	CentC
ZB113.Contig18 (20048)	213	367	154	3.2	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	368	522	154	1.9	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	523	677	154	1.3	SgCMCM00200228	CentC
ZB113.Contig18 (20048)	678	832	154	2.6	SgCMCM00200228	CentC
ZB113.Contig18 (20048)	833	988	155	0.6	SgCMCM00200095	CentC
ZB113.Contig18 (20048)	989	1143	154	3.2	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	1144	1298	154	1.3	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	1299	1453	154	2.6	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	1454	1609	155	1.9	SgCMCM00200079	CentC
ZB113.Contig18 (20048)	1610	1765	155	1.3	SgCMCM00200102	CentC
ZB113.Contig18 (20048)	1766	1921	155	0	SgCMCM00200099	CentC
ZB113.Contig18 (20048)	1922	2077	155	2.6	SgCMCM00200014	CentC
ZB113.Contig18 (20048)	2078	2232	154	2.6	SgCMCM00200224	CentC

ZB113.Contig18 (20048)	2233	2387	154	1.9	SgCMCM00200017	CentC
ZB113.Contig18 (20048)	2388	2542	154	2.6	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	2543	2697	154	2.6	SgCMCM00200017	CentC
ZB113.Contig18 (20048)	2698	2852	154	1.9	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	2853	3007	154	0.7	SgCMCM00200228	CentC
ZB113.Contig18 (20048)	3008	3161	153	0	SgCMCM00200173	CentC
ZB113.Contig18 (20048)	3162	3317	155	1.9	SgCMCM00200156	CentC
ZB113.Contig18 (20048)	3318	3473	155	0	SgCMCM00200167	CentC
ZB113.Contig18 (20048)	3474	3629	155	3.2	SgCMCM00200014	CentC
ZB113.Contig18 (20048)	3630	3784	154	2.6	SgCMCM00200017	CentC
ZB113.Contig18 (20048)	3785	3940	155	0	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	3941	4096	155	1.3	SgCMCM00200104	CentC
ZB113.Contig18 (20048)	4097	4251	154	1.3	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	4252	4406	154	1.3	SgCMCM00200169	CentC
ZB113.Contig18 (20048)	4407	4562	155	2.6	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	4563	4718	155	2.6	SgCMCM00200102	CentC
ZB113.Contig18 (20048)	4719	4874	155	0	SgCMCM00200099	CentC
ZB113.Contig18 (20048)	4875	5027	152	2	SgCMCM00200005	CentC
ZB113.Contig18 (20048)	5028	5183	155	1.9	SgCMCM00200095	CentC
ZB113.Contig18 (20048)	5184	5339	155	1.3	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	5340	5494	154	0.7	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	5495	5649	154	0.7	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	5650	5804	154	0	SgCMCM00200169	CentC
ZB113.Contig18 (20048)	5805	5959	154	1.9	SgCMCM00200224	CentC
ZB113.Contig18 (20048)	5960	6115	155	1.3	SiCMCM0020001	CentC
ZB113.Contig18 (20048)	6116	6271	155	3.2	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	6272	6426	154	2.6	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	6427	6581	154	3.2	SgCMCM00200169	CentC
ZB113.Contig18 (20048)	6582	6737	155	1.9	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	6738	6892	154	3.2	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	6893	7047	154	1.9	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	7048	7202	154	4.5	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	7203	7388	185	1.1	SgCMCM00200225	CentC

ZB113.Contig18 (20048)	7389	7548	159	0	SgCMCM00200500	CentC
ZB113.Contig18 (20048)	7549	7703	154	0.7	SgCMCM00200224	CentC
ZB113.Contig18 (20048)	7704	7858	154	0.7	SgCMCM00200005	CentC
ZB113.Contig18 (20048)	7859	7951	92	4.3	SgCMCM00200012	CentC
ZB113.Contig18 (20048)	7952	8370	418	2	SmOTOT00200141	family_1241_C3
ZB113.Contig18 (20048)	8338	8646	308	25.2	SiCMCMOOT0036	CRM
ZB113.Contig18 (20048)	8549	8927	378	5.5	SmOTOT00200480	family_1868_C1
ZB113.Contig18 (20048)	8789	8949	160	4.5	SmOTOT00200480	family_1868_C1
ZB113.Contig18 (20048)	8955	9262	307	2.9	SmOTOT00200215	family_1380_C2
ZB113.Contig18 (20048)	8974	9371	397	25.5	SmOTOT00101933	family_4330_C5
ZB113.Contig18 (20048)	9309	10034	725	2.2	SmOTOT00200264	family_1431_C3
ZB113.Contig18 (20048)	10017	10554	537	27.1	SiCMCMOOT0036	CRM
ZB113.Contig18 (20048)	10512	10664	152	26.4	SmOTOT00201588	family_6912_C1
ZB113.Contig18 (20048)	10633	13769	3136	25.3	SiCMCMOOT0036	CRM
ZB113.Contig18 (20048)	13802	13931	129	0.8	SmOTOT00101774	family_3931_C1
ZB113.Contig18 (20048)	13932	17719	3787	2.5	SiCMCMOOT0036	CRM
ZB113.Contig18 (20048)	17718	17780	62	4.8	SgCMCM00200021	CentC
ZB113.Contig18 (20048)	17781	17936	155	1.9	SgCMCM00200095	CentC
ZB113.Contig18 (20048)	17937	18091	154	1.9	SgCMCM00200228	CentC
ZB113.Contig18 (20048)	18092	18246	154	1.9	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	18247	18401	154	1.9	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	18402	18556	154	2.6	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	18557	18711	154	0.7	SgCMCM00200034	CentC
ZB113.Contig18 (20048)	18712	18868	156	1.3	SgCMCM00200026	CentC
ZB113.Contig18 (20048)	18869	19022	153	1.3	SiCMCM0020001	CentC
ZB113.Contig18 (20048)	19023	19177	154	1.3	SgCMCM00200099	CentC
ZB113.Contig18 (20048)	19178	19333	155	2.6	SgCMCM00200014	CentC
ZB113.Contig18 (20048)	19334	19489	155	0.6	SgCMCM00200099	CentC
ZB113.Contig18 (20048)	19490	19647	157	3.9	SgCMCM00200058	CentC
ZB113.Contig18 (20048)	19648	19802	154	3.2	SgCMCM00200044	CentC
ZB113.Contig18 (20048)	19803	19958	155	0.6	SgCMCM00200099	CentC
ZB113.Contig18 (20048)	19959	20015	56	8.8	SgCMCM00200457	CentC
ZB113.Contig18 (20048)	19959	20000	41	2.4	SgCMCM00200498	CentC

The ZB113 contigs 4, 8, 11, 15, and 17 consist of nearly all repeated sequence, contig 14 has an approximately 4 kb stretch of repeated sequence within its 3' half, and contig 18 has a 5 kb and 3 kb of repeated sequence on its 5' and 3' ends. The repeated regions are apparent both when compared to self and the reverse complement, therefore the larger repeat regions consists of many smaller repeat regions that occur both in the forward and reverse direction.

The consensus sequence of the CentC repeat represented in ZB113 is set out as SEQ ID NO: 71. The variants of the CentC repeats present in ZB113 are set out in Table 16 where the most common base is indicated. Where the most common base occurs less than 60% of the time, the percent occurrence of each base is reported.

Table 16 ZB113 CentC consensus sequence and variants

1 C	60 G (A:7/C:0/G:92/T:1/-:0)	119 T(T:0/-:100)	178 C (A:2/C:97/T:1)
2 C	61 D(A:1/G:0/T:1/-:99)	120 C(C:0/-:100)	179 G (A:3/G:97)
3 A	62 T (A:0/C:0/T:99)	121 A	180 G
4 T	63 C (A:8/C:91/G:0/T:1)	122 C	181 A
5 T (G:0/T:100)	64 C (C:99/T:0/-:0)	123 C (C:93/T:7)	182 A
6 T (C:2/G:0/T:98)	65 A (A:99/C:0/G:0/-:0)	124 C (C:99/T:1)	183 C
7 C (A:0/C:92/T:8)	66 A (A:99/C:0/G:0)	125 A	184 C (C:99/T:1)
8 T (A:0/C:0/T:100)	67 A (A:100/T:0)	126 T (A:0/T:100)	185 A
9 T (C:1/G:1/T:98/-:0)	68 A	127 T (G:0/T:100)	186 T
10 C (A:1/C:84/G:3/T:12)	69 A(A:2/-:98)	128 T(T:0/-:100)	187 C
11 G (A:1/G:97/T:2)	70 C (A:0/C:99/T:1)	129 C	188 T
12 T	71 T (C:0/T:100)	130 C (C:69/G:0/T:31)	
13 T	72 C (C:100/T:0)	131 G (A:1/G:99)	
14 T (G:0/T:100)	73 A (A:100/G:0)	132 A	
15 T (A:0/T:100)	74 T (A:0/C:0/G:0/T:99)	133 A	
16 T (C:0/T:100)	75 K(G:35/T:1/-:64)	134 A	
17 K(G:0/T:0/-:99)	76 G (A:0/G:64/T:35/-:0)	135 A	
18 C (A:0/C:95/G:1/T:3/-:0)	77 T (A:0/C:0/G:0/T:64/-:35)	136 M(A:1/C:25/-:74)	
19 G (A:6/C:0/G:89/T:5)	78 T	137 C (A:1/C:72/G:18/-:10)	
20 C (A:0/C:78/G:0/T:21)	79 T (G:1/T:99/-:0)	138 G (A:1/G:96/T:3)	
21 M(A:2/C:1/-:97)	80 G (A:1/C:0/G:88/T:8/-:4)	139 G (A:5/G:95)	
22 A(A:2/-:98)	81 G (A:5/G:95)	140 G (G:75/T:8/-:17)	
23 A (A:99/C:0/-:0)	82 G (A:1/G:94/T:5)	141 T (C:0/T:100)	
24 A (A:97/C:1/-:2)	83 G (A:2/G:98)	142 G (A:24/G:76)	
25 C (C:90/G:3/T:5/-:2)	84 D(A:0/G:2/T:3/-:95)	143 T (C:0/T:100)	
26 G (A:3/G:97)	85 T (A:0/G:5/T:95)	144 C (C:97/G:2/T:0)	
27 A (A:97/G:3/T:0)	86 G (A:0/C:0/G:99)	145 G (A:14/G:86)	
28 A (A:98/G:0/T:0/-:1)	87 G (A:2/C:0/G:94/T:3)	146 G (G:83/T:11/-:6)	
29 C (A:1/C:75/G:1/-:23)	88 D(A:0/G:0/T:8/-:91)	147 G (A:4/G:96/T:0)	
30 V(A:8/C:2/G:0/-:90)	89 T	148 K(G:55/T:8/-:37)	
31 M(A:0/C:17/-:83)	90 T (G:0/T:100)	149 K(G:4/T:0/-:96)	
32 M(A:17/C:6/-:78)	91 T (C:1/T:89/-:10)	150 T (G:1/T:99)	
33 A (A:75/-:25)	92 C (A:0/C:95/G:0/T:4)	151 G (A:0/G:100)	
34 T (G:0/T:99/-:0)	93 G (A:6/C:0/G:86/T:8)	152 C (A:4/C:96)	

35	G (C:0/G:99/T:0)	94	C (A:1/C:98/G:1/T:1)	153	A (A:100/T:0)
36	C (C:100/T:0)	95	G (A:37/C:0/G:63)	154	T (A:0/C:1/T:99)
37	C (A:1/C:98/G:0/T:1)	96	C (A:1/C:99/G:0/T:0)	155	A
38	C (A:2/C:84/G:7/T:5/-:2)	97	A (A:99/T:1)	156	C (A:3/C:68/-:29)
39	B(C:1/G:0/T:0/-:98)	98	A (A:94/G:6/T:0)	157	A
40	H(A:0/C:0/T:0/-:99)	99	R(A:0/G:0/-:99)	158	A (A:96/G:4)
41	A (A:100/C:0/G:0)	100	T	159	A
42	A (A:99/T:1)	101	T (G:0/T:100)	160	G
43	T (C:3/T:97)	102	T	161	C
44	C (A:0/C:83/G:0/T:1/-:15)	103	C (A:0/C:81/G:1/T:18/-:0)	162	A (A:99/T:1)
45	C (C:99/G:0/T:0)	104	G (A:0/C:0/G:97/T:2)	163	C
46	N(A:0/C:16/G:0/T:0/-:83)	105	R(A:0/G:1/-:99)	164	G (A:1/G:99)
47	A (A:99/C:1)	106	T(T:1/-:99)	165	A (A:97/C:3)
48	C (C:97/G:2/T:1)	107	T (G:0/T:100)	166	G (A:1/G:98/-:1)
49	T (A:0/C:0/T:99)	108	T (A:0/T:100)	167	T
50	W(A:50/T:50)	109	G (C:0/G:99/T:0)	168	T (G:1/T:99)
51	A (A:99/C:1/T:1)	110	T (C:16/T:84)	169	T
52	C (C:99/T:1)	111	C (A:0/C:99/T:0)	170	T
53	T (A:0/T:100)	112	G (A:4/C:0/G:96)	171	T
54	T (A:1/T:99/-:0)	113	C (A:0/C:99/G:1)	172	G (G:99/T:1)
55	T (T:100/-:0)	114	A (A:98/G:1/T:0)	173	T(T:0/-:100)
56	W(A:1/T:1/-:98)	115	C (A:0/C:97/T:3)	174	C
57	A(A:0/-:100)	116	G (A:1/G:99)	175	C
58	A (A:99/G:1)	117	T	176	A
59	G (A:0/C:6/G:94)	118	C (C:97/T:3)	177	C (C:98/G:2)
				mean length	155
				stdev	1.5

The sequence of the CRM retrotransposon (SEQ ID NO: 77) was blasted against the contigs of ZB113 and filtered for hits with alignment lengths greater than 50 to determine the representation of CRM in ZB113. The representation of CRM within ZB113 is summarized in Table 17.

Table 17 – CRM Fragments in ZB113

Contig	CRM			Contig match	
	begin	end	% identity	begin	end
ZB113.fasta.screen.Contig14	1	930	99.3	933	1862
ZB113.fasta.screen.Contig14	1	515	99.8	7937	8451
ZB113.fasta.screen.Contig14	2796	2893	91.1	2132	2035
ZB113.fasta.screen.Contig14	5765	7571	90.9	57	1862
ZB113.fasta.screen.Contig14	6640	7156	83.7	7936	8451
ZB113.fasta.screen.Contig16	1	1434	99.4	5452	4019
ZB113.fasta.screen.Contig16	1508	5417	99.6	3947	37
ZB113.fasta.screen.Contig16	4251	4744	99.3	6565	7058
ZB113.fasta.screen.Contig16	4626	4772	80.1	12043	11897
ZB113.fasta.screen.Contig16	4945	6236	80.8	11724	10433
ZB113.fasta.screen.Contig16	4983	5342	80.0	7297	7656
ZB113.fasta.screen.Contig16	5487	5569	80.6	7801	7883
ZB113.fasta.screen.Contig16	5757	6213	85.7	8071	8527
ZB113.fasta.screen.Contig16	6529	6653	86.8	10140	10016
ZB113.fasta.screen.Contig16	6608	6658	82.4	8922	8972
ZB113.fasta.screen.Contig16	6638	7572	84.3	5455	4522
ZB113.fasta.screen.Contig18	1	1434	99.0	17719	16288

ZB113.fasta.screen.Contig18	1508	3791	99.4	16214	13932
ZB113.fasta.screen.Contig18	2796	2890	99.7	10426	10520
ZB113.fasta.screen.Contig18	4251	4744	80.3	11782	12275
ZB113.fasta.screen.Contig18	4983	5342	79.8	12514	12873
ZB113.fasta.screen.Contig18	5487	5569	80.6	13018	13100
ZB113.fasta.screen.Contig18	5757	6213	88.0	13288	13744
ZB113.fasta.screen.Contig18	6640	7572	82.1	17720	16789

Five unique repeats were identified in the nucleotide sequence of ZB113 (SmOTOT00200141, SmOTOT00200215 (2 variants), SmOTOT00200264, SmOTOT00200480, SmOTOT00201588 and analyzed for variation in a manner similar to CentC. The repeat SmOTOT00200141 was too large for for analysis with source reads matching a wide variety of locations. The consensus sequence of SmOTOT00200215 are set out as SEQ ID NO: 72 and SEQ ID NO: 73. The consensus of SmOTOT00200480 is set out as SEQ ID NO: 74. The consensus of SmOTOT00201588 is set out as SEQ ID NO: 75. The variants of the unique repeats are set out in Tables 18-20 respectively where the most common base is indicated. Where the most common base occurs less than 60% of the time, the percent occurrence of each base is reported.

The sequences were queried against GenBank, which returned no feature specific hit. SmOTOT00200215.1 (SEQ ID NO: 72) and SmOTOT00200480 (SEQ ID NO: 74) and SmOTOT00201588 (SEQ ID NO: 75) matched a clone from Zea mays (AC116034), and SmOTOT00200215.2 (SEQ ID NO: 73) returned no matches.

Table 18 SmOTOT00200215.1 Variation in ZB113

1	T	60	T	119	G
2	T	61	T	120	A
3	T	62	T	121	G
4	C	63	T	122	G
5	A	64	C	123	C
6	T	65	G	124	C
7	C	66	C	125	G
8	C	67	T	126	G
9	C	68	C	127	C
10	G	69	G	128	C
11	G	70	G	129	A
12	T	71	G	130	A
13	C	72	T	131	A
14	G	73	G	132	C
15	T	74	T	133	T
16	T	75	C	134	C (C:83/G:17)
17	T	76	C	135	A
18	T	77	A	136	C
19	T	78	A	137	C
20	A	79	A	138	T
21	G (A:17/G:83)	80	A	139	A (A:83/T:17)
22	A	81	A	140	C (C:83/T:17)
23	A	82	A	141	G

24	C	83	T	142	G
25	A	84	C	143	T
26	T	85	T	144	C
27	A	86	G	145	T
28	A	87	A	146	G
29	C	88	A	147	T
30	T	89	A	148	T
31	T	90	T	149	T
32	G	91	T	150	G
33	A	92	T	151	G
34	G	93	T	152	G
35	G	94	T	153	G
36	T	95	A	154	T
37	A	96	T (C:17/T:83)	155	T
38	C	97	A	156	C
39	C	98	G	157	G
40	T	99	G	158	A
41	T	100	A		
42	C	101	G		
43	C	102	C		
44	G	103	T		
45	T	104	A (A:83/T:17)		
46	A	105	G		
47	A	106	T		
48	A	107	T		
49	C	108	G		
50	C	109	A		
51	G	110	C		
52	G	111	A		
53	G	112	C		
54	C	113	C		
55	A	114	A		
56	T	115	T		
57	A	116	T	mean length	156
58	A	117	C	stdev	2.9
59	C	118	T (G:17/T:83)		

Table 19 SmOTOT00200215.2 Variation in ZB113

1	A	60	A (A:83/C:17)	119	T
2	C	61	G (A:17/G:83)	120	T
3	A	62	V(A:33/C:33/G:33)	121	A
4	A	63	A	122	C
5	A	64	A	123	D(A:50/G:33/T:17)
6	C	65	A	124	T
7	C	66	A	125	A (A:67/C:33)
8	G	67	A	126	A
9	A	68	A	127	A
10	G	69	A	128	A
11	T	70	A	129	A
12	C	71	A	130	A
13	A	72	A(A:33/-:67)	131	A
14	A	73	M(A:17/C:17/-:67)	132	A
15	C	74	C (A:17/C:83)	133	C
16	G	75	A	134	A
17	G	76	A	135	G
18	Y(C:50/T:50)	77	A	136	A
19	C	78	C	137	A (A:67/-:33)
20	G (G:80/T:20)	79	C	138	A (A:67/-:33)

21	T	80	G	139	C (C:67/-:33)
22	T (C:40/T:60)	81	A	140	A (A:67/-:33)
23	T (C:20/T:80)	82	G	141	A (A:67/-:33)
24	C	83	T	142	A
25	C (C:80/T:20)	84	C	143	A
26	T (G:17/T:83)	85	M(A:17/C:17/-:67)	144	A
27	T	86	A	145	A
28	G (G:67/T:33)	87	A	146	A
29	T (A:33/T:67)	88	C	147	A
30	T (G:17/T:83)	89	G (G:67/-:33)	148	A
31	T	90	A (A:67/-:33)	149	G
32	T	91	C (C:67/-:33)	150	A
33	T (C:17/T:83)	92	C (C:67/-:33)	151	A (A:67/T:33)
34	C (C:67/T:33)	93	G (G:67/-:33)	152	A
35	T (C:33/T:67)	94	G(G:17/-:83)	153	A
36	C (C:83/G:17)	95	C(C:17/-:83)	154	C
37	C (C:67/T:33)	96	C(C:17/-:83)	155	G
38	T	97	C (C:83/T:17)	156	A
39	T	98	T	157	A
40	C (C:83/G:17)	99	T (C:17/T:83)	158	G
41	G	100	C	159	G
42	G	101	C	160	A
43	T	102	T	161	G
44	T	103	T	162	A
45	A (A:67/T:17/-:17)	104	G	163	A
46	A(A:17/-:83)	105	T	164	G (A:33/G:67)
47	C(C:17/-:83)	106	T	165	G
48	C (A:17/C:67/G:17)	107	T	166	G
49	G	108	T	167	A
50	A	109	T	168	T
51	A	110	C	169	A
52	A	111	T	170	C
53	A	112	C	171	G
54	A	113	C	172	G
55	A	114	T	173	T
56	A	115	T	174	T
57	A	116	C	175	G
58	A	117	G	176	T
59	C	118	G		

153
3.6

Table 20 SmOTOT00200480 Variation in ZB113

1	G	60	A	119	C	178	C
2	A	61	Y(C:50/T:50)	120	A	179	T
3	C	62	G	121	A	180	C
4	G	63	G	122	R(A:50/G:50)	181	A
5	T	64	R(A:50/G:50)	123	K(G:50/T:50)	182	A
6	A	65	A	124	R(A:50/G:50)	183	A
7	A	66	G	125	A	184	A
8	C	67	A	126	A(A:50/-:50)	185	G
9	C	68	T	127	C	186	G
10	G	69	G	128	G	187	G
11	A	70	R(A:50/G:50)	129	C	188	T
12	A	71	T	130	G	189	G
13	G	72	G	131	A	190	C
14	G	73	G	132	A	191	T
15	A	74	C	133	A	192	G
16	G	75	G	134	G	193	G
17	A	76	G	135	C		
18	A	77	C	136	A		
19	A	78	G	137	C		
20	A	79	C	138	A		
21	A	80	T	139	C		
22	Y(C:50/T:50)	81	A	140	A		
23	A	82	G	141	A		

24	A	83	G	142	A
25	G	84	R(A:50/G:50)	143	T
26	G	85	T	144	T
27	A	86	T	145	C
28	R(A:50/G:50)	87	T	146	A
29	A	88	G	147	A
30	C	89	A	148	C
31	G	90	A	149	A
32	A	91	T	150	A
33	T	92	G	151	T
34	G	93	G	152	G
35	T	94	T	153	C
36	T	95	G	154	A
37	G	96	G	155	G
38	A	97	A	156	A
39	C	98	A	157	T
40	T	99	G	158	T
41	C	100	A	159	A
42	G	101	A	160	T
43	G	102	C	161	T
44	T	103	A	162	G
45	T	104	C	163	A
46	T	105	A	164	A
47	G	106	A	165	A
48	T	107	T	166	G
49	G	108	G	167	A
50	G	109	C	168	A
51	Y(C:50/T:50)	110	A	169	A
52	G	111	A	170	G
53	T	112	C	171	T
54	G	113	C	172	G
55	A	114	A	173	Y(C:50/T:50)
56	T	115	G	174	G
57	C	116	C	175	A
58	A	117	A	176	G
59	A	118	A	177	G

192
0.55

Summary

The sequence analysis described above demonstrates that BAC ZB19 is enriched for CentC and BAC ZB113 is enriched for CentC and CRM. The frequency of these repeats is particular to the BACs of the invention and is not a representation of the natural occurrence of these repeats in the maize genome. The relative frequency of sequences within the entire maize genome database (TIGR web site) having homology to CentC or CRM was compared to the frequency in ZB19 and ZB113. CentC hit the maize genome (300Mb) 530 times over a total aligned length of 70 kb. CRM hit the maize genome 860 times over a total aligned length of 336 kb. The proportion of CentC and CRM in ZB19 and ZB113 as compared to the maize genome is summarized in Table 21.

Table 21

	CentC	CRM
ZB19	28.9%	0.0%
ZB113	47.5%	31.7%
maize genome	0.0%	0.1%

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

The invention is claimed as follows:

1. An isolated mini-chromosome comprising a centromere, wherein the centromere comprises
 - (a) at least seven first repeated nucleotide sequences that hybridize under conditions comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour to any one of: the nucleotide sequence of SEQ ID NO: 70 or the nucleotide sequence of SEQ ID NO: 71, wherein the repeated nucleotide sequences are within about a 1.3 kb stretch of nucleotide sequence,
 - (b) at least a second nucleotide sequence comprising a nucleotide sequence that hybridizes under conditions comprising 0.5 x SSC and 0.25% SDS at 65°C for 15 minutes, followed by a wash at 65°C for a half hour to any one of: the nucleotide sequence of SEQ ID NO: 77 or the nucleotide sequence of SEQ ID NO: 78, and
 - (c) at least three exogenous nucleic acids,wherein the centromere confers the ability to segregate to daughter cells.
2. The minichromosome of claim 1, wherein the first repeated nucleotide sequence is selected from the group consisting of:
 - (a) a nucleotide sequence at least 70% identical to SEQ ID NO: 70,
 - (b) a nucleotide sequence at least 70% identical to SEQ ID NO: 71,
 - (c) a nucleotide sequence of SEQ ID NO: 70, or
 - (d) a nucleotide sequence of SEQ ID NO: 71.
3. The mini-chromosome of claim 1 or 2, wherein the second nucleotide sequence is selected from the group consisting of:
 - (a) a nucleotide sequence at least 70% identical to a fragment of SEQ ID NO: 77 at least 100 bp in length,
 - (b) a nucleotide sequence at least 85% identical to a fragment of SEQ ID NO: 77 at least 50 bp in length,

- (c) a fragment of the nucleotide sequence of SEQ ID NO: 77 at least 50 base pairs in length,
 - (d) a nucleotide sequence at least 70% identical to a fragment of SEQ ID NO: 78 at least 100 bp in length,
 - (e) a nucleotide sequence at least 85% identical to a fragment of SEQ ID NO: 78 at least 50 bp in length, and
 - (f) a fragment of the nucleotide sequence of SEQ ID NO: 78 at least 50 base pairs in length.
4. The mini-chromosome of any one of claims 1-3, wherein the second nucleotide sequence is at least 100 bp in length.
 5. The mini-chromosome of any one of claims 1-3, wherein the second nucleotide sequence is at least 150 bp in length.
 6. The mini-chromosome of any one of claims 1-3, wherein the second nucleotide sequence is at least 500 bp in length.
 7. The mini-chromosome of any one of claims 1-3, wherein the second nucleotide sequence is less than 3909 bp in length.
 8. The mini-chromosome of any one of claims 1-3, wherein the second nucleotide sequence ranges from 50 to 3909 bp in length.
 9. The mini-chromosome of any one of claims 1-3, wherein the centromere comprises at least two copies of the second nucleotide sequence.
 10. The mini-chromosome of any one of claims 1-3, wherein the centromere comprises at least 5 copies of the second nucleotide sequence.
 11. The mini chromosome of any one of claims 1-10, wherein the centromere comprises n copies of the first repeated nucleotide sequence, wherein n is less than 1000.

12. The mini chromosome of any one of claims 1-10, wherein the centromere comprises n copies of the first repeated nucleotide sequence, wherein n is less than 500.
13. The mini-chromosome of any one of claims 1-10, wherein the centromere comprises n copies of the first repeated nucleotide sequence, wherein n is at least 10.
14. The mini-chromosome of any one of claims 1-10, wherein the centromere comprises n copies of the first repeated nucleotide sequence, wherein n is at least 50.
15. The mini-chromosome of any one of claims 1-10, wherein the centromere comprises n copies of the repeated nucleotide sequence, wherein n is at least 100.
16. The mini-chromosome of any one of claims 1-10, wherein the centromere comprises at least 20 copies of the first repeated nucleotide sequence within 3.5 kb of nucleotide sequence.
17. The mini-chromosome of claim 13, wherein at least 5 first repeated nucleotides sequences are in tandem.
18. The mini-chromosome of claim 13, wherein at least 5 first repeated nucleotide sequences are consecutive.
19. The mini-chromosome of claim 13, wherein at least 5 first repeated nucleotide sequences are each separated by less than n number of nucleotides, wherein n is 50.
20. The mini-chromosome of claim 19, wherein n is 20.
21. The mini-chromosome of claim 19, wherein n is 10.
22. The mini-chromosome of any one of claims 1-21, further comprising a site for site-specific recombination.

23. The mini-chromosome of any one of claims 1-22, wherein the minichromosome comprises at least ten exogenous nucleic acids.
24. The mini-chromosome of any one of claims 1-23, wherein at least one exogenous nucleic acid is operably linked to a heterologous regulatory sequence functional in plant cells.
25. The mini-chromosome of any one of claims 1-23, wherein at least one exogenous nucleic acid is operably linked to a plant promoter.
26. The mini-chromosome of any one of claims 1-25, wherein the exogenous nucleic acid is selected from the group consisting of a herbicide resistance gene, a nitrogen fixation gene, an insect resistance gene, a disease resistance gene, a plant stress-induced gene, a nutrient utilization gene, a gene that affects plant pigmentation, a gene that encodes an antisense or ribozyme molecule, a gene encoding a secretable antigen, a toxin gene, a receptor gene, a ligand gene, a seed storage gene, a hormone gene, an enzyme gene, an interleukin gene, a clotting factor gene, a cytokine gene, an antibody gene, and a growth factor gene.
27. The mini-chromosome of claim 26, wherein the disease resistance gene confers resistance to a virus, bacteria, fungi or nematode.
28. The mini-chromosome of claim 26, wherein the enzyme gene is selected from the group consisting of a gene that encodes an enzyme involved in metabolizing biochemical wastes for use in bioremediation, a gene that encodes an enzyme for modifying pathways that produce secondary plant metabolites, a gene that encodes an enzyme that produces a pharmaceutical, a gene that encodes an enzyme that improves the nutritional content of a plant, a gene that encodes an enzyme involved in vitamin synthesis, a gene that encodes an enzyme involved in carbohydrate or starch synthesis, a gene that encodes an enzyme involved in mineral accumulation or availability, a gene that encodes a phytase, a gene that encodes an enzyme involved in fatty acid or oil synthesis, a gene that encodes an enzyme involved in synthesis of chemicals or plastics, a gene that encodes an enzyme involved in synthesis of a fuel and a gene that encodes enzyme involved in synthesis of a fragrance.

29. The mini-chromosome of any one of claims 1-25, wherein the exogenous nucleic acid encodes a protein conferring resistance to drought, heat, chilling, freezing, excessive moisture, or salt stress.
30. The mini-chromosome of any one of claims 1-25, wherein the exogenous nucleic acid is a phosphinothricin resistance gene.
31. The mini-chromosome of claim 30, wherein the phosphinothricin resistance gene is a phosphinothricin acetyltransferase.
32. The mini-chromosome of any one of claims 1-25, wherein the exogenous nucleic acid is a glyphosate resistance gene.
33. The mini-chromosome of claim 32, wherein the glyphosate resistance gene is a mutant EPSP synthase.
34. The mini-chromosome of any one of claims 1-25, wherein the exogenous nucleic acid is a *Bacillus thuringiensis* toxin gene.
35. The mini-chromosome of any one of claims 1-34, wherein the minichromosome is circular.
36. The mini-chromosome of any one of claims 1-35, wherein the minichromosome exhibits a segregation efficiency in corn cells of at least 60%.
37. The minichromosome of any one of claims 1-35, wherein the minichromosome exhibits a segregation efficiency in corn cells of at least 80%.
38. The minichromosome of any one of claims 1-35, wherein the minichromosome exhibits a segregation efficiency in corn cells of at least 90%.

39. The minichromosome of any one of claims 1-35, wherein the minichromosome exhibits a segregation efficiency in corn cells of at least 95%.
40. A corn plant cell comprising a minichromosome of any one of claims 1-39.
41. A corn plant cell comprising a minichromosome of any one of claims 1-39 that (i) is not integrated into the plant cell genome and (ii) confers an altered phenotype on the plant cell associated with at least one structural gene within the minichromosome.
42. The corn plant cell of claim 41, wherein the altered phenotype comprises increased expression of a native gene.
43. The corn plant cell of claim 41, wherein the altered phenotype comprises decreased expression of a native gene.
44. The corn plant cell of claim 41, wherein the altered phenotype comprises expression of an exogenous gene.
45. The corn plant cell of any one of claims 41-44, further comprising an integrated exogenous structural gene.
46. A method of using a plant comprising a corn plant cell of any one of claims 41-45, the method comprising the step of growing the plant to produce a recombinant protein encoded by a structural gene of the minichromosome.
47. The method of claim 46, further comprising the steps of harvesting or processing the plant and extracting the recombinant protein therefrom.
48. The method of claim 46 or 47, wherein the recombinant protein is a pharmaceutical.
49. The method of claim 46 or 47, wherein the plant produces a modified food product.

50. The method of claim 46 or 47, wherein the recombinant protein is an enzyme.
51. The method of claim 50, wherein the enzyme is selected from the group consisting of an enzyme involved in metabolizing biochemical wastes for use in bioremediation, an enzyme for modifying pathways that produce secondary plant metabolites, an enzyme that produces a pharmaceutical, an enzyme involved in vitamin synthesis, an enzyme involved in starch synthesis, an enzyme involved in mineral accumulation or availability, an enzyme involved in fatty acid synthesis, an enzyme involved in synthesis of chemicals or plastics, and enzyme involved in synthesis of a fragrance.