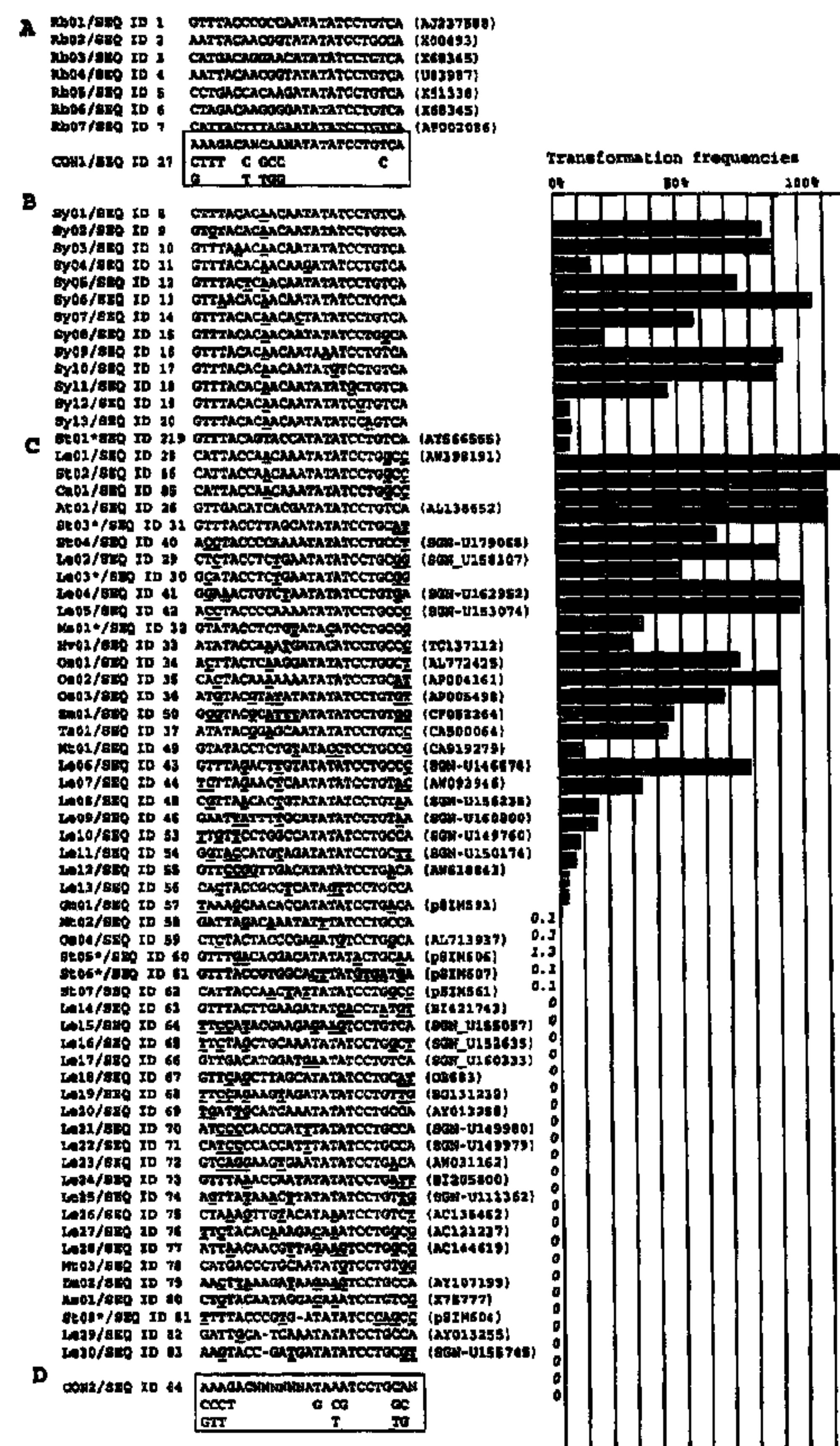




(86) Date de dépôt PCT/PCT Filing Date: 2005/09/07  
(87) Date publication PCT/PCT Publication Date: 2006/03/16  
(45) Date de délivrance/Issue Date: 2013/01/29  
(85) Entrée phase nationale/National Entry: 2007/03/08  
(86) N° demande PCT/PCT Application No.: US 2005/031537  
(87) N° publication PCT/PCT Publication No.: 2006/029076  
(30) Priorités/Priorities: 2004/09/08 (US60/607,586);  
2005/05/26 (US60/684,525); 2005/07/14 (US60/698,938)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01),  
*A01H 1/00* (2006.01)  
(72) Inventeurs/Inventors:  
ROMMENS, CAIUS M. T., US;  
BOUGRI, OLEG V., US;  
YAN, HUA, US  
(73) Propriétaire/Owner:  
J.R. SIMPLOT COMPANY, US  
(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : ELEMENTS GENETIQUES PROPRES A DES PLANTES ET CASSETTES DE TRANSFERT POUR LA TRANSFORMATION DE PLANTES  
(54) Title: PLANT-SPECIFIC GENETIC ELEMENTS AND TRANSFER CASSETTES FOR PLANT TRANSFORMATION



(57) Abrégé/Abstract:

The present invention provides nucleic acid molecules and sequences, particularly those identified and obtained from plants, that are useful for transferring and integrating one polynucleotide into another via plant transformation techniques.



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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
16 March 2006 (16.03.2006)

PCT

(10) International Publication Number  
**WO 2006/029076 A2**

## (51) International Patent Classification:

C12N 15/82 (2006.01) A01H 1/00 (2006.01)

## (21) International Application Number:

PCT/US2005/031537

## (22) International Filing Date:

7 September 2005 (07.09.2005)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

|            |                               |    |
|------------|-------------------------------|----|
| 60/607,586 | 8 September 2004 (08.09.2004) | US |
| 60/684,525 | 26 May 2005 (26.05.2005)      | US |
| 60/698,938 | 14 July 2005 (14.07.2005)     | US |

(71) Applicant (for all designated States except US): **J.R. SIMPLOT COMPANY** [US/US]; One Capital Center, 999 Main Street, Suite 1300, P.O. Box 27, Boise, ID 83702 (US).

## (72) Inventors; and

(75) Inventors/Applicants (for US only): **ROMMENS, Caius**

**M. T.** [NL/US]; 3395 East Windsong Drive, Boise, ID 83712 (US). **BOUGRI, Oleg V.** [RU/US]; 2017 Mortimer Drive, Boise, Idaho 83712 (US). **YAN, Hua** [CN/US]; 6102 S. Settlement Way, Boise, Idaho 83716 (US).

(74) Agent: **PEET, Richard C.**; FOLEY & LARDNER LLP, 3000 K Street, Suite 500, Washington, DC 20007-5143 (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, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

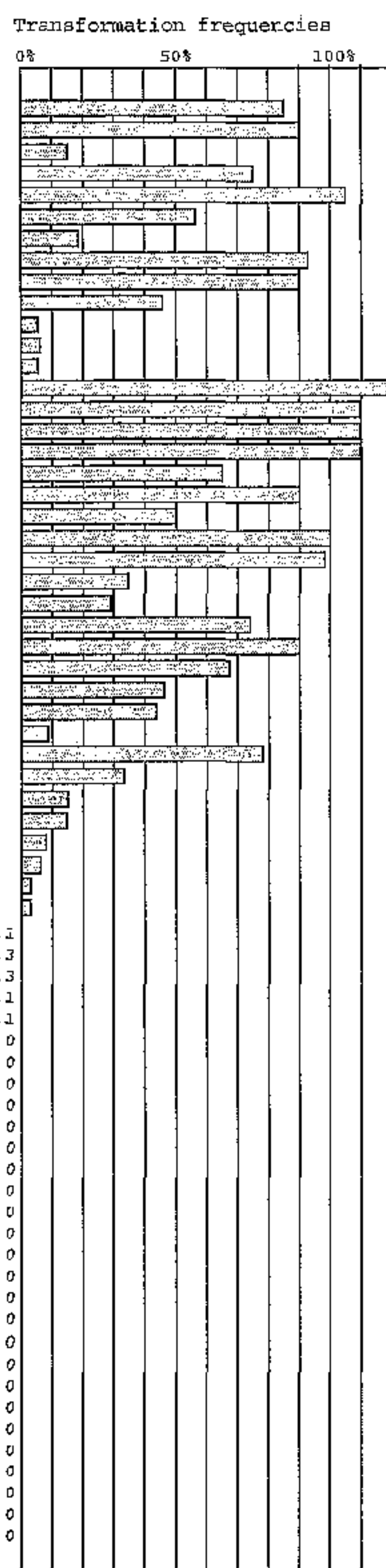
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

## (54) Title: PLANT-SPECIFIC GENETIC ELEMENTS AND TRANSFER CASSETTES FOR PLANT TRANSFORMATION

A Rb01/SEQ ID 1 GGGGACCCGCGGAGATATATCTGCA (AJ227588)  
Rb02/SEQ ID 2 GGGGACCAAGGAGATATATCTGCA (XU0493)  
Rb03/SEQ ID 3 GATACACAGAGATATATCTGCA (X83481)  
Rb04/SEQ ID 4 AATTACACAGATATATCTGCA (U83987)  
Rb05/SEQ ID 5 CCTGACCAAGATATATCTGCA (X83338)  
Rb06/SEQ ID 6 CTAGACAGGAGATATATCTGCA (X66345)  
Rb07/SEQ ID 7 CATTACTTAGATATATCTGCA (APC02086)  
CON1/SEQ ID 27 CTT C GCC C  
G T TGG

B Sy01/SEQ ID 8 GTTACACACATATATCTGCA  
Sy02/SEQ ID 9 GTTACACACATATATCTGCA  
Sy03/SEQ ID 10 GTTACACACATATATCTGCA  
Sy04/SEQ ID 11 GTTACACACATATATCTGCA  
Sy05/SEQ ID 12 GTTACACACATATATCTGCA  
Sy06/SEQ ID 13 GTTACACACATATATCTGCA  
Sy07/SEQ ID 14 GTTACACACATATATCTGCA  
Sy08/SEQ ID 15 GTTACACACATATATCTGCA  
Sy09/SEQ ID 16 GTTACACACATATATCTGCA  
Sy10/SEQ ID 17 GTTACACACATATATCTGCA  
Sy11/SEQ ID 18 GTTACACACATATATCTGCA  
Sy12/SEQ ID 19 GTTACACACATATATCTGCA  
Sy13/SEQ ID 20 GTTACACACATATATCTGCA  
St01/SEQ ID 21 GTTACACACATATATCTGCA (AY566555)  
Le01/SEQ ID 28 CATTACACACATATATCTGCA (AB398192)  
St02/SEQ ID 86 CATTACACACATATATCTGCA  
Ca01/SEQ ID 85 CATTACACACATATATCTGCA  
At01/SEQ ID 26 GTTACACACATATATCTGCA (AT136862)  
St03/SEQ ID 31 GTTACACACATATATCTGCA  
St04/SEQ ID 40 GTTACACACATATATCTGCA (SGN-U179688)  
Le02/SEQ ID 29 GTTACACACATATATCTGCA (SGN-U158307)  
Le03/SEQ ID 30 GTTACACACATATATCTGCA  
Le04/SEQ ID 41 GTTACACACATATATCTGCA (SGN-U162952)  
Le05/SEQ ID 42 GTTACACACATATATCTGCA (SGN-U153074)  
Me01/SEQ ID 32 GTTACACACATATATCTGCA  
Hv01/SEQ ID 33 ATATACACACATATATCTGCA (TCL37112)  
Gm01/SEQ ID 34 GTTACACACATATATCTGCA (AL772425)  
Cm02/SEQ ID 35 GTTACACACATATATCTGCA (AB004161)  
Cm03/SEQ ID 36 GTTACACACATATATCTGCA (AP005498)  
Zm01/SEQ ID 50 GTTACACACATATATCTGCA (CP082664)  
Ta01/SEQ ID 37 ATATACACACATATATCTGCA (CA500064)  
Mt01/SEQ ID GTTACACACATATATCTGCA (CA919279)  
Le06/SEQ ID 43 GTTACACACATATATCTGCA (SGN-U146678)  
Le07/SEQ ID 44 GTTACACACATATATCTGCA (AN092946)  
Le08/SEQ ID 45 GTTACACACATATATCTGCA (SGN-U156338)  
Le09/SEQ ID 46 GTTACACACATATATCTGCA (SGN-U160800)  
Le10/SEQ ID 53 GTTACACACATATATCTGCA (SGN-U149760)  
Le11/SEQ ID 54 GTTACACACATATATCTGCA (SGN-U150174)  
Le12/SEQ ID 55 GTTACACACATATATCTGCA (AW618443)  
Fe13/SEQ ID 36 CATTACACACATATATCTGCA (pSIM593)  
Zm01/SEQ ID TAAACACACATATATCTGCA (pSIM593)  
Me02/SEQ ID 58 GTTACACACATATATCTGCA (AL713997)  
Ca04/SEQ ID 59 GTTACACACATATATCTGCA (pSIM606)  
St05/SEQ ID 60 GTTACACACATATATCTGCA (pSIM607)  
St06/SEQ ID 61 GTTACACACATATATCTGCA (pSIM611)  
St07/SEQ ID 62 GTTACACACATATATCTGCA (BI421743)  
Le14/SEQ ID 53 GTTACACACATATATCTGCA (SGN-U155057)  
Le15/SEQ ID 54 GTTACACACATATATCTGCA (SGN-U152638)  
Le16/SEQ ID 55 GTTACACACATATATCTGCA (SGN-U160333)  
Le17/SEQ ID 56 GTTACACACATATATCTGCA (O86823)  
Le18/SEQ ID 57 GTTACACACATATATCTGCA (BU131232)  
Le19/SEQ ID 58 GTTACACACATATATCTGCA (AY013255)  
Le20/SEQ ID 59 GTTACACACATATATCTGCA (SGN-U149980)  
Le21/SEQ ID 70 GTTACACACATATATCTGCA (SGN-U149979)  
Le22/SEQ ID 71 GTTACACACATATATCTGCA (AW031462)  
Le23/SEQ ID 72 GTTACACACATATATCTGCA (B1205808)  
Le24/SEQ ID 73 GTTACACACATATATCTGCA (SGN-U113362)  
Le25/SEQ ID 74 GTTACACACATATATCTGCA (AC135462)  
Le26/SEQ ID 75 GTTACACACATATATCTGCA (AC121237)  
Le27/SEQ ID 76 GTTACACACATATATCTGCA (AC146119)  
Le28/SEQ ID 77 GTTACACACATATATCTGCA  
Mt02/SEQ ID 78 GTTACACACATATATCTGCA (AY107199)  
Zm02/SEQ ID 79 GTTACACACATATATCTGCA (X75777)  
Rb08/SEQ ID 80 GTTACACACATATATCTGCA (pSIM604)  
St08/SEQ ID 81 GTTACACACATATATCTGCA (AY013255)  
Le29/SEQ ID 82 GTTACACACATATATCTGCA (SGN-U155748)  
Le30/SEQ ID 83 GTTACACACATATATCTGCA



(57) Abstract: The present invention provides nucleic acid molecules and sequences, particularly those identified and obtained from plants, that are useful for transferring and integrating one polynucleotide into another via plant transformation techniques.

WO 2006/029076 A2

# WO 2006/029076 A2



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).

*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.*

**Published:**

— *without international search report and to be republished upon receipt of that report*

## **PLANT-SPECIFIC GENETIC ELEMENTS AND TRANSFER CASSETTES FOR PLANT TRANSFORMATION**

### **FIELD OF THE INVENTION**

**[0001]** Described herein are nucleic acid molecules and sequences, particularly those identified and obtained from plants, that are useful for transferring and integrating one polynucleotide into another via bacterial-mediated transformation.

### **BACKGROUND OF THE INVENTION**

**[0002]** Bacterial-mediated transformation via, for example, *Agrobacterium* or *Rhizobium*, entails the transfer and integration of a polynucleotide from a bacterial plasmid into the genome of a eukaryotic organism. The region of DNA within the bacterial plasmid that is designated for such manipulation is called the transfer DNA (“T-DNA”).

**[0003]** A T-DNA region is delimited by left and right “border” sequences, which are each about twenty-five nucleotides in length and oriented as imperfect direct repeats of the other. T-DNA transfer is initiated by an initial single stranded nick at the so-called right border site and is terminated by a subsequent secondary nick at the left border site. It is the resultant single-stranded linear DNA molecule that is transported, by the activity of other proteins, into the plant cell and ultimately integrated into the plant genome.

[0005] After initial cleavage at the right border, virD2 covalently binds to the 5'-side, and the DNA unwinds towards the left border where a second cleavage reaction occurs. The released single stranded DNA, traditionally referred to as the "T-strand," is coated with virE2 and processed for transfer via type IV type secretion (Lessl and Lanka, (1994) Cell 77: 321-324, 1994; Zupan and Zambryski, Plant Physiol 107: 1041-1047, 1997).

[0006] Since border sequences alone do not support a highly effective DNA transfer, extended border regions, generally comprising about 200 or more base pairs of *Agrobacterium* tumor-inducing (Ti) plasmid DNA, are used to transform plant cells. Two non-border sequences that are located within these extended border regions have been shown to promote DNA transfer, namely the 'overdrive' domain of pTi15955 (van Haaren *et al.*, Nucleic Acids Res. 15: 8983-8997, 1987) and a DNA region containing at least five repeats of the 'enhancer' domain of pRiA4 (Hansen *et al.*, Plant Mol. Biol., 20:113-122, 1992).

[0007] One issue associated with the use of conventional *Agrobacterium* border regions is the infidelity of DNA transfer. For instance, primary cleavage reactions at the right border are often not followed by secondary cleavage reactions at the left border. This "border skipping" leads to the transfer of T-DNAs that are still connected to the rest of the plasmid. Such plasmid backbone transfer is undesirable because these sequences typically comprise antibiotic resistance genes. Plasmid backbone transfer can also be a consequence of inadvertent right border activity at the left border.

[0008] A second issue concerns the use of conventional and poorly characterized *Agrobacterium* border regions, which permit only very little optimization of transfer frequencies. This leads to poor transformation rates, and high input costs for the production of large numbers of transformed plants.

[0009] Furthermore, the presence of foreign T-DNA sequences in food crops is often perceived as undesirable, and the application of genetic engineering has therefore been limited to a small number of crops that are destined for feed, oil, fibers,

and processed ingredients. Public concerns were addressed through development of an all-native approach to making genetically engineered plants, as disclosed by Rommens *et al.* in WO2003/069980, US-2003-0221213, US-2004-0107455, and WO2005/004585. Rommens *et al.* teach the identification and isolation of genetic elements from plants that can be used for bacterium-mediated plant transformation. Thus, Rommens teaches that a plant-derived transfer-DNA (“P-DNA”), for instance, can be isolated from a plant genome and used in place of an *Agrobacterium* T-DNA to genetically engineer plants.

**[0009]** The concept of P-DNA mediated transformation has previously been demonstrated in potato. A 400-base pair potato P-DNA delineated by regions that share sequence identity with the left border of nopaline strains and the right border of octopine strains was effectively transferred from *Agrobacterium* to plant cells (Rommens *et al.*, Plant Physiol 135: 421-431, 2004).

**[0010]** The potato P-DNA was subsequently used to introduce a silencing construct for a tuber-specific polyphenol oxidase (PPO) gene into potato. Resulting intragenic plants displayed tolerance against black spot bruise sensitivity in impacted tubers.

**[0011]** The present invention provides new plant-specific DNA elements that replace bacterial borders, and are particularly useful for all-native DNA transformation methods.

**[0012]** The present invention also reveals the organization of the extended regions that are involved in the initiation of DNA transfer by mediating primary DNA cleavage, and describes the sequence requirements and spacing of genetic elements that support high activity of the described elements. Furthermore, the invention shows how manipulations of regions that surround enzyme cleavage sites can enhance the fidelity of DNA transfer.

## SUMMARY OF THE INVENTION

[0014] One aspect of the present invention is a DNA sequence, comprising a polynucleotide sequences, designated as a “cleavage sites”, that comprise the consensus sequence depicted in SEQ ID NO: 84 and which are not identical to an *Agrobacterium* transfer-DNA border sequence, nor to a previously isolated border or border-like sequence.

[0015] In one embodiment, a cleavage site is selected from the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 190, 194-196, and 198. In one embodiment, the cleavage site represents a synthetic sequence, and is selected from the group consisting of SEQ ID NOs: 8,9 and 11-13. The present invention contemplates a transformation cassette that comprises two cleavage sites. One of those sites may be termed the “primary cleavage site,” while the other may be a “secondary cleavage site.” See Figure 4.

[0016] In another embodiment, the cleavage site is generated by substituting at least one nucleotide of a cleavage site or cleavage site-like sequence selected from the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-86, 190, and 193-198.

[0017] In another embodiment, the cleavage site represents a contiguous sequence of a plant genome, and is selected from the group consisting of SEQ ID NOs: 15-17, 28-37, 38-50, and 85-86.

[0018] In yet another embodiment, the cleavage site is derived from a variant of a sequence selected from the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 190, 194-196. That is, a variant of any one of these particular sequences is encompassed by the present invention so long as the variant sequence permits cleavage by a pertinent transformation enzyme and/or enzyme complex involved in bacterium-mediated transformation. Hence, a variant sequence may share about 99%, about 98%, about 97%, about 96%, about 95%, about 94%, about 93%, about 92%, about 91%, about 90%, about 89%, about 88%, about 87%, about 86%, about 85%, about 84%, about 83%, about 82%, about 81%, about 80%,

about 79%, about 78%, about 77%, about 76%, about 75%, about 74%, about 73%, about 72%, about 71%, about 70%, about 69%, about 68%, about 67%, about 66%, about 65%, about 64%, about 63%, about 62%, about 61%, about 60%, about 59%, about 58%, about 57%, about 56%, about 55%, about 54%, about 53%, about 52%, about 51%, or about 50%, or about less than 50% sequence identity with of any one of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 190, 194-196, so long as the variant sequence can still be cleaved according to the present invention.

[0019] Another aspect of the present invention is a transfer cassette, comprising such a cleavage site positioned upstream from a desired polynucleotide. In one embodiment, the cleavage site in the transfer cassette is selected from the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-50, 85-86, 189, 190, and 194-196.

[0020] In one embodiment, the transfer cassette comprises two cleavage sites defined by a first polynucleotide and a second polynucleotide, whereby the first polynucleotide may comprise a sequence for an “initial cleavage site” that is positioned upstream from the desired polynucleotide. The second polynucleotide may comprise a sequence for a “final cleavage site” that is positioned downstream from the desired polynucleotide. The two cleavage sites may be positioned as perfect or imperfect direct repeats.

[0021] The transfer cassette may further comprise a nucleotide sequence downstream from the initial cleavage site, whereby this “DI region” is a DNA sequence that (a) comprises at least about 30 base pairs immediately downstream from the initial cleavage site, (b) comprises a sequence that shares at least 70% sequence identity with the DR domain depicted in SEQ ID NO: 107, that is positioned within about 60 base pairs from the initial cleavage site, (c) optionally contains multiple sequences that are identical or inverse complementary to SEQ ID NO: 115, (d) is not identical to a region that flanks a T-DNA right border in *Agrobacterium* Ti or Ri plasmids, and (e) supports cleavage activity. The DI region may enhance the

initial cleavage activity by at least 25% compared to the corresponding sequence of the Ti or Ri plasmid, which does not comprise the same DI region.

[0022] In one embodiment the DI region may share at least 70% sequence identity with SEQ ID NO: 22, 108-114.

[0023] In one embodiment, the transfer cassette further comprises a nucleotide sequence upstream from the final cleavage site, whereby this “UF region” is a DNA sequence that (a) comprises at least 40 base pairs immediately upstream from the final cleavage site, (b) comprises at least 55% adenine or thymine residues (AT-rich), (c) comprises a sequence that has at least 70% sequence identity to either the UL domain depicted in SEQ ID NO: 120 or the inverse complement of SEQ ID NO: 120 within a distance of about 50 base pairs from the final cleavage site, (d) optionally comprises a putative binding site for integration host factor that has at least 70% sequence identity to the consensus sequence [A/T]-ATCAANNNTT-[A/G] (SEQ ID NO: 129) or has at least 70% sequence identity to the inverse complement of SEQ ID NO: 129, and that is positioned within 200 base pairs from the final cleavage site or left border, (e) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids, and (f) supports initial cleavage site activity. In one embodiment, the UF region enables transformation frequencies that are increased, such as by at least 25%, compared to the corresponding sequence of a Ti or Ri plasmid.

[0024] In one embodiment, the UF region may share at least 70% sequence identity to the sequences depicted in SEQ ID NO: 184-186 and 211-214.

[0025] In another embodiment, the transfer cassette further comprises both a DI and UF element.

[0026] Another aspect of the present invention is a transformation vector comprising any one of such transfer cassettes, wherein the region of the plasmid backbone that is “upstream from the initial cleavage” (UI region) comprises at least a 48-nucleotide sequence that contains adenine-rich trinucleotides interspaced by

nucleotides that represent, in at least six cases, a cytosine or thymine (pyrimidine) residue, whereby the most downstream pyrimidine represents either the first base of the initial cleavage site or the base at position -4 relative to the initial cleavage site. The UI region is not identical to a region that flanks a T-DNA border of an *Agrobacterium* or binary plasmid. The UI region supports initial cleavage activity and may enable transformation frequencies that are increased, such as by at least 25%, compared to the corresponding sequence of a Ti or Ri plasmid.

[0027] In one embodiment, the UI region of the transformation vector comprises a nucleotide sequence that has greater than 70% sequence identity to the sequence depicted in SEQ ID NOs: 199-208.

[0028] In another embodiment, the region of the plasmid backbone that is associated with the final cleavage site (AF region) is a DNA sequence that (a) comprises at least part of the final cleavage site or left border and at about two to 40 base pairs flanking downstream DNA, (b) comprises at least four tightly linked clusters of two or more cytosine bases separated by 1-11 other nucleotides, CCN1-11CCN1-11CCN1-11CC (SEQ ID NO: 122), (c) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids, and (d) supports initial cleavage activity. In one embodiment, the AF region enables transformation frequencies that are, for example, at least 25% compared to the corresponding sequence of a Ti or Ri plasmid.

[0029] In one embodiment, the AF region of the transformation vector comprises a nucleotide sequence that has greater than 70% sequence identity to the sequence depicted in SEQ ID NOs: 187, 188, and 215-218.

[0030] The present invention is not limited to the percentage by which initial or final cleavage activity is enhanced by any particular transformation element described herein. For instance, any of the transformation elements described herein may enhance the initial or final cleavage activity by 100% or more than 100%, or about 99%, about 98%, about 97%, about 96%, about 95%, about 94%, about 93%, about 92%, about 91%, about 90%, about 89%, about 88%, about 87%, about 86%,

about 85%, about 84%, about 83%, about 82%, about 81%, about 80%, about 79%, about 78%, about 77%, about 76%, about 75%, about 74%, about 73%, about 72%, about 71%, about 70%, about 69%, about 68%, about 67%, about 66%, about 65%, about 64%, about 63%, about 62%, about 61%, about 60%, about 59%, about 58%, about 57%, about 56%, about 55%, about 54%, about 53%, about 52%, about 51%, about 50%, about 49%, about 48%, about 47%, about 46%, about 45%, about 44%, about 43%, about 42%, about 41%, about 40%, about 39%, about 38%, about 37%, about 36%, about 35%, about 34%, about 33%, about 32%, about 31%, about 30%, about 29%, about 28%, about 27%, about 26%, about 25%, about 24%, about 23%, about 22%, about 21%, about 20%, about 15%, or about 5% or at least about 1%, compared to a control that does not comprise the desired transformation element.

[0031] The present invention also contemplates transformation cassettes and plasmids, whereby not every transformation element in the construct enhances cleavage activity. Thus, not every element in a cassette described herein must enhance cleavage activity or transformation efficiency in order for it to be useful.

[0032] In another aspect of the present invention, a transformation vector is provided, which comprises (A) a transfer cassette, which comprises, from 5' to 3', (i) an initial cleavage site, (ii) a DI region, (iii) a UF region, and (iv) a final cleavage site, and (B) in the transformation plasmid backbone, at least one of (i) a UI region, and (ii) a AF region.

[0033] In one aspect, the relevant sequences for DNA transfer of such a transformation vector are shown in SEQ ID NO: 131 and 132.

[0034] In one embodiment, the transformation vector further comprises a desired polynucleotide positioned between DI and UF region.

[0035] In another embodiment, the transformation vector contains at least one *Agrobacterium* border as alternative to a cleavage site.

[0036] In one embodiment, a putative cleavage site is identified by screening DNA databases using programs such as BLASTN or a similar program and search motifs such as depicted in SEQ ID NO: 130.

[0037] In another embodiment, a putative cleavage site is isolated by applying PCR-based methods described in the Examples.

[0038] In yet another embodiment, a DI region or UF region is identified by screening DNA databases with programs such as BLASTN (Altschul *et al.*, Nucleic Acids Res. 25: 3389-3402, 1997) using desired domains as queries.

[0039] In one embodiment, a method of identifying a functionally active cleavage site is provided comprising the steps: (a) identifying a putative cleavage site, (b) annealing two primers in such a way that a double strand DNA sequence is generated comprising the putative cleavage site, optionally flanked by the sticky ends of specific restriction enzyme sites, (c) ligating this DNA fragment with a linearized plasmid that contains replication origins for both *E. coli* and *Agrobacterium*, (d) introducing the new plasmid into *Agrobacterium*, (e) infecting explants of a plant that is amenable to *Agrobacterium*-mediated transformation with the resulting *Agrobacterium* strain, (f) applying tissue culture methods for transformation, proliferation, and, if necessary, regeneration (g) allowing callus and/or shoot formation, (h) counting the average number of calli and/or shoots per explant, and comparing the resulting frequencies with those of conventional controls, (i) selecting putative cleavage sites that support transformation.

[0040] In one embodiment, the putative cleavage site may be found to enhance the transformation efficiency in comparison to an identical plasmid, which does not contain the putative cleavage site. For instance, a putative cleavage site may enhance the transformation efficiency by about 100% or more than 100%, or about 99%, about 98%, about 97%, about 96%, about 95%, about 94%, about 93%, about 92%, about 91%, about 90%, about 89%, about 88%, about 87%, about 86%, about 85%, about 84%, about 83%, about 82%, about 81%, about 80%, about 79%, about 78%, about 77%, about 76%, about 75%, about 74%, about 73%, about 72%, about

71%, about 70%, about 69%, about 68%, about 67%, about 66%, about 65%, about 64%, about 63%, about 62%, about 61%, about 60%, about 59%, about 58%, about 57%, about 56%, about 55%, about 54%, about 53%, about 52%, about 51%, about 50%, about 49%, about 48%, about 47%, about 46%, about 45%, about 44%, about 43%, about 42%, about 41%, about 40%, about 39%, about 38%, about 37%, about 36%, about 35%, about 34%, about 33%, about 32%, about 31%, about 30%, about 29%, about 28%, about 27%, about 26%, about 25%, about 24%, about 23%, about 22%, about 21%, about 20%, about 15%, or about 5% or at least about 1%, compared to a control that does not comprise the putative cleavage site.

**[0041]** In one embodiment, a method of identifying a functionally active DI or UF region is provided comprising the steps; (a) identifying a putative DNA region, (b) isolating the region from plant DNA using methods such as PCR, (c) using this region to replace the functional region of a transformation vector, (d) introducing the modified plasmid into *Agrobacterium*, (e) infecting explants of a plant that is amenable to *Agrobacterium*-mediated transformation with the resulting *Agrobacterium* strain, (f) applying tissue culture methods for transformation and proliferation, (g) allowing callus formation, (h) counting the average number of calli per explant, and comparing the resulting frequencies to those obtained with a conventional control plasmid that does not comprise the putative DNA region, and (i) identifying a DNA region that supports transformation.

**[0042]** In one embodiment, a putative DNA region may be found to enhance the transformation efficiency in comparison to an identical plasmid, which does not contain the putative DNA region. For instance, a putative DNA region may enhance the transformation efficiency by about 100% or more than 100%, or about 99%, about 98%, about 97%, about 96%, about 95%, about 94%, about 93%, about 92%, about 91%, about 90%, about 89%, about 88%, about 87%, about 86%, about 85%, about 84%, about 83%, about 82%, about 81%, about 80%, about 79%, about 78%, about 77%, about 76%, about 75%, about 74%, about 73%, about 72%, about 71%, about 70%, about 69%, about 68%, about 67%, about 66%, about 65%, about 64%, about 63%, about 62%, about 61%, about 60%, about 59%, about 58%, about 57%, about

56%, about 55%, about 54%, about 53%, about 52%, about 51%, about 50%, about 49%, about 48%, about 47%, about 46%, about 45%, about 44%, about 43%, about 42%, about 41%, about 40%, about 39%, about 38%, about 37%, about 36%, about 35%, about 34%, about 33%, about 32%, about 31%, about 30%, about 29%, about 28%, about 27%, about 26%, about 25%, about 24%, about 23%, about 22%, about 21%, about 20%, about 15%, or about 5% or at least about 1%, compared to a control that does not comprise the putative DNA region.

**[0043]** In one embodiment, the step of identifying the putative DNA region may be accomplished by hybridization studies, where a random or degenerate nucleic acid probe or oligonucleotide is used to identify sequences from a genome that can be subsequently tested for transformation efficacy. For instance, such a probe may be employed in a Southern blot of genomic DNA isolated from a plant, where the probe is essentially based on one of the transformation elements described herein, e.g., a UF region of the present invention.

**[0044]** Alternatively, a preparation of DNA may be subjected to PCR using primers that are specific to a particular transformation element described herein. On the other hand, the primers may be random primers or degenerate primers based on a desired transformation element, that are employed in a PCR reaction of DNA. The subsequently amplified PCR product(s) can be isolated by standard procedures, e.g., via excising it from an electrophoretic gel, and then tested according to the present invention for transformation efficacy.

**[0045]** In one embodiment, at least one, if not all, of the nucleotide sequences of the transfer cassette are endogenous to a plant. That is, in one embodiment, at least one, if not all, of the nucleotide sequences in the transfer cassette are native to a plant, or are isolated from the same plant, the same plant species, or from plants that are sexually interfertile with the plant to be transformed. In one embodiment, the plant is a monocotyledonous plant and selected from the group consisting of wheat, turf grass, maize, rice, oat, wheat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, and palm.

[0046] In another embodiment, the plant is a dicotyledonous plant and selected from the group consisting of potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, pea, bean, cucumber, grape, brassica, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, and cactus.

[0047] Another aspect of the present invention is a method for transforming a plant cell, comprising introducing a transformation vector, which comprises any one of the transfer cassettes described herein, into a plant cell.

[0048] In one embodiment, the plant cell is located in a plant. In another embodiment, the plant is selected from the group consisting of wheat, turf grass, maize, rice, oat, wheat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, palm, potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, and cactus.

[0049] In another embodiment, the transformation plasmid is introduced into the plant cell via a bacterium. In one embodiment, the bacterium is from *Agrobacterium*, *Rhizobium*, or *Phyllobacterium*. In a further embodiment, the bacterium is selected from the group consisting of *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, and *MesoRhizobium loti*.

[0050] In a preferred embodiment, at least one, if not all, of the nucleotide sequences in the transfer cassette are isolated from the same plant, the same plant species, or plants that are sexually interfertile. In one embodiment all of the nucleotide sequences are isolated from the same plant, the same plant species, or from plants that are sexually interfertile.

[0051] In one embodiment, a cassette is provided, which comprises (1) a first polynucleotide, comprising a sequence that is (i) nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and (ii) not identical to a

bacterial border sequence; (2) a second polynucleotide, which may be (i) an imperfect or perfect repeat of the first polynucleotide, or (ii) a bacterial T-DNA border; (3) a desired polynucleotide; and (4) at least one of (a) UI region, (b) DI region, (c) UF region, and (d) AF region.

**[0052]** In one embodiment, the first polynucleotide comprises a sequence that is native to a plant genome. In another embodiment, the first polynucleotide consists essentially of a sequence that is native to a plant genome.

**[0053]** In a preferred embodiment, the first polynucleotide is targeted by a vir gene-encoded protein. In one embodiment, the vir gene-encoded protein is VirD2.

**[0054]** In another embodiment, the first polynucleotide conforms to the consensus sequence depicted in SEQ ID NO: 84. In a preferred embodiment, the first polynucleotide comprises a sequence depicted in any one of the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 190, 194-196, and 198.

**[0055]** In another embodiment, the first polynucleotide comprises a sequence with at least 70% sequence identity to the sequence of any one of SEQ ID NO: 28, 85, or 86. In a further embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with a sequence depicted in any one of SEQ ID NOs: 28-30.

**[0056]** In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in SEQ ID NO: 32.

**[0057]** In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in SEQ ID NO: 33.

**[0058]** In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in any one of SEQ ID NOs: 34-36.

[0059] In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in SEQ ID NO: 37.

[0060] In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in any one of SEQ ID NOs: 195-196.

[0061] In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in any one of SEQ ID NOs: 51 and 194.

[0062] In one embodiment, the first polynucleotide comprises a sequence that shares at least 70% sequence identity with the sequence depicted in any one of SEQ ID NOs: 189-190.

[0063] In one embodiment, the first polynucleotide comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more nucleotides that are different in comparison to an *Agrobacterium* T-DNA border sequence.

[0064] In one embodiment, the first polynucleotide is greater than 70% identical in sequence to an *Agrobacterium* T-DNA border sequence.

[0065] In another embodiment, the UI region comprises a sequence that shares at least 70% sequence identity with at least one of SEQ ID NOs: 199-208.

[0066] In another embodiment, the DI region element comprises a sequence that shares at least 70% sequence identity with at least one of SEQ ID NOs: 22, 108-114.

[0067] In another embodiment, the UF region element comprises a sequence that shares at least 70% sequence identity with at least part of at least one of SEQ ID NOs: 184-186 and 211-214. In another embodiment, the AF region comprises a sequence that shares at least 70% sequence identity with at least one of SEQ ID NOs: 187, 188, or 215-218.

[0068] The present invention encompasses variant sequences of the transformation elements described herein and is not limited to the percentage sequence identity that any particular transformation element may share with any particular sequence described herein. Thus, the present invention encompasses sequences for any of the transformation elements described herein, e.g., a UI region, DI region, UF region, or AF region, that shares about 99%, about 98%, about 97%, about 96%, about 95%, about 94%, about 93%, about 92%, about 91%, about 90%, about 89%, about 88%, about 87%, about 86%, about 85%, about 84%, about 83%, about 82%, about 81%, about 80%, about 79%, about 78%, about 77%, about 76%, about 75%, about 74%, about 73%, about 72%, about 71%, about 70%, about 69%, about 68%, about 67%, about 66%, about 65%, about 64%, about 63%, about 62%, about 61%, about 60%, about 59%, about 58%, about 57%, about 56%, about 55%, about 54%, about 53%, about 52%, about 51%, about 50%, about 49%, about 48%, about 47%, about 46%, about 45%, about 44%, about 43%, about 42%, about 41%, about 40%, about 39%, about 38%, about 37%, about 36%, about 35%, about 34%, about 33%, about 32%, about 31%, about 30%, about 29%, about 28%, about 27%, about 26%, about 25%, about 24%, about 23%, about 22%, about 21%, about 20%, about 15%, or about 5% or at least about 1% sequence identity with a corresponding sequence identified herein.

[0069] Another aspect of the present invention contemplates transformation elements such as a UI region, DI region, UF region, or AF region, that does not comprise a nucleotide sequence that is identical to a corresponding region from a bacterium plasmid, such as from a tumor-inducing plasmid from *Agrobacterium* or *Rhizobium*.

[0070] In another embodiment, the AF region element comprises at least 70% sequence identity with at least part of at least one of SEQ ID NO: 187, 188, and 215-218.

[0071] In another embodiment, the desired polynucleotide is positioned between the first and second polynucleotides, and wherein the desired polynucleotide

is located downstream from a first polynucleotide cleavage site that functions in initial cleavage.

[0072] In a preferred embodiment, the cassette comprises a UI region positioned upstream from the first polynucleotide cleavage site and a AF region that is downstream from the second polynucleotide cleavage site.

[0073] In one particular embodiment, the portion of the cassette that comprises the UI and DI regions comprise the sequence depicted in SEQ ID NO: 131. In one embodiment, the portion of the cassette that comprises the UF and AF regions comprises the sequence depicted in SEQ ID NO: 132.

[0074] In one preferred embodiment, all of the DNA sequences between the first and second polynucleotides are plant DNA. In this regard, the plant DNA is endogenous to (1) a monocotyledonous plant selected from the group consisting of wheat, turf grass, maize, rice, oat, wheat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, and palm; or (2) a dicotyledonous plant selected from the group consisting of potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, and cactus, cucumber, melon, canola, apple, or pine.

[0075] In another embodiment, the cassette further comprises at least one of (1) an overdrive element, comprising a sequence that is at least 70% identical in sequence to SEQ ID NO: 88; (2) a pyrimidine-rich element, comprising a sequence that shares at least 70% sequence identity with any one of SEQ ID NOs: 199-208 but which is not identical to an *Agrobacterium* plasmid sequence that flanks a right border; (2) an AT-rich element, comprising a sequence that shares at least 70% sequence identity to at least part of any one of SEQ ID NOs: 184-186 and 211-214; and (4) a cytosine cluster, comprising a sequence at least 70% sequence identity to at least part of any one of SEQ ID NOs: 187-188 and 215-218.

**[0076]** The present invention also provides a plant transformation cassette, which comprises at least one of (1) a polynucleotide comprising a sequence depicted in any one of the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-50, 85, 86, and 190 or any other cleavage site sequence disclosed herein, wherein the 3'-end of the polynucleotide abuts a cytosine cluster, e.g., wherein the sequence comprising the 3'-end of the polynucleotide and DNA downstream thereof, comprises the sequence depicted in SEQ ID NO: 122; and (2) a polynucleotide comprising a sequence depicted in any one of the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-50, 85, and 86 or any other cleavage site disclosed herein, wherein the 5'-end of the polynucleotide abuts a UI region.

**[0077]** In one embodiment, the cytosine cluster comprises a sequence that shares at least 70% sequence identity with any one of the sequences in SEQ ID NOs: 187-188.

**[0078]** In another embodiment, the UI region comprises a sequence that shares at least 70% sequence identity with any one of the sequences in SEQ ID NOs: 199, 209, and 210.

**[0079]** In another embodiment, a plant transformation cassette is provided, which comprises at least one of (1) a polynucleotide comprising a sequence depicted in any one of the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-50, 85, 86, and 190, wherein the 3'-end of the polynucleotide abuts a cytosine cluster; (2) a polynucleotide comprising (i) a sequence depicted in any one of the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 194-196, and 198, and (ii) a DNA sequence positioned downstream of the sequence of (i), wherein the sequences of (i) and (ii) together comprise a cytosine cluster; and (3) a polynucleotide comprising a sequence depicted in any one of the group consisting of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 194-196, and 198, wherein the 5'-end of the polynucleotide abuts a pyrimidine-rich element. In one embodiment, the cytosine cluster comprises a sequence that shares at least 70% sequence identity with any one of the sequences in SEQ ID NOs: 187-188. In another embodiment, the pyrimidine-

rich element comprises a sequence that shares at least 70% sequence identity with any one of the sequences in SEQ ID NOs: 21 and 199-208.

[0080] Another aspect of the present invention is a method for transforming a plant cell, which comprises introducing any one of the cassettes or plant transformation cassettes described herein into a plant cell. Such a cassette may be positioned within a plant transformation plasmid, such as a Ti- or Ri-plasmid.

[0081] Thus, in one particular embodiment, a cassette of the present invention is placed in a vector, which is derived from a tumor-inducing cassette from an *Agrobacterium*, *Rhizobium*, or *Phyllobacterium* bacterium, and which is suitable for plant transformation.

[0082] In one embodiment, the bacterium is selected from the group consisting of *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, and *MesoRhizobium loti*.

[0083] In another embodiment of this method, the vector housing the desired cassette is maintained in a strain of one of these bacteria and it is the bacterium strain that is used to infect the plant cell and thereby introduce the cassette or plant transformation cassette into the plant cell.

[0084] In one embodiment, the plant cell is located in either (1) a monocotyledonous plant or explant thereof selected from the group consisting of wheat, turf grass, maize, rice, oat, wheat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, and palm; or (2) a dicotyledonous plant or explant thereof selected from the group consisting of potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, and cactus.

[0085] In one particular embodiment, a tomato plant is transformed using a cassette in which the first polynucleotide in the cassette comprises a sequence that

shares at least 70% sequence identity with any one of the sequences of SEQ ID NO: 28-30.

[0086] In another embodiment, an alfalfa plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to the sequence depicted in SEQ ID NO: 32.

[0087] In another embodiment, a barley plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to the sequence depicted in SEQ ID NO: 33.

[0088] In another embodiment, a rice plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to the sequence depicted in SEQ ID NOs: 34-36.

[0089] In another embodiment, a wheat plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to the sequence depicted in SEQ ID NO: 37.

[0090] In another embodiment, a soybean plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to the sequence depicted in any one of SEQ ID NOs: 195-196.

[0091] In another embodiment, a maize plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to the sequence depicted in any one SEQ ID NOs: 51 and 194.

[0092] In another embodiment, a *Brassica* plant is transformed using a cassette in which the first polynucleotide comprises a sequence that shares at least 70% sequence identity to one of the sequences depicted in SEQ ID NOs: 189 or 198. In one embodiment, the plant to be transformed is a *Brassica* plant.

[0093] The present invention does not limit which polynucleotide sequence can be used to transform a particular plant. Thus, a first polynucleotide that comprises a sequence that shares at least 70% sequence identity to the sequence

depicted in any one of SEQ ID NOs: 51 and 194, can be used to transform a potato plant, instead of maize. Hence, the present invention contemplates various permutations of transformation elements and their usefulness in transforming a variety of plants and organisms. According to the present invention, an animal cell may be transformed using any of the cassettes or plasmids described herein. Hence, in one embodiment, an animal cell may be transformed with genetic elements that are native to the animal and its species, thereby providing an “all-native” approach to transforming animal cells and animals.

[0094] In one particular embodiment, the monocotyledonous or dicotyledonous explant is a seed, germinating seedling, leaf, root, stem, cutting, or bud.

[0095] According to these methods, the bacterium that is used to perform the plant transformation can be an *Agrobacterium*, *Rhizobium*, or *Phyllobacterium* bacterium. In one embodiment, the bacterium is selected from the group consisting of *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, and *MesoRhizobium loti*.

[0096] In one embodiment, the bacterial T-DNA border of the cassette described herein is from *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, or *MesoRhizobium loti*.

[0097] Another aspect of the present invention is a cassette, which comprises (1) a first polynucleotide, comprising a sequence that is nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and; (2) a second polynucleotide that has greater than 70% sequence identity to any one of SEQ ID NOs: 133-137. In one embodiment, the cassette further comprises a desired polynucleotide. In another embodiment the first polynucleotide is a bacterial T-DNA right border sequence. In another embodiment, the first polynucleotide is not identical in sequence to a bacterial T-DNA right border sequence. The sequence of

the first polynucleotide may comprise the sequence depicted in any one of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-50, 85, 86, 189, 190, and 194-196.

**[0098]** In another aspect, a transposase-transposon, plant transformation cassette is provided, which comprises (i) left and right transfer-DNA border sequences; (ii) a non-autonomous transposable element; and (iii) a transposase gene, wherein the non-autonomous transposable element and the transposase gene are positioned between the left and right border sequences.

**[0099]** In one embodiment, the plant transformation cassette comprises at least one of the border sequences comprising a sequence that is (i) nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and (ii) is not identical to a bacterial border sequence. The sequence of the first polynucleotide may comprise the sequence depicted in any one of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-50, 85, 86, 189, 190, and 194-196.

**[0100]** In one embodiment, in this cassette, at least one of the border sequences is a bacterial T-DNA border. In another embodiment, the cassette further comprises a desired polynucleotide positioned within the non-autonomous transposable element.

**[0101]** In one embodiment, the terminal ends of the non-autonomous transposable element are those from maize transposable element Ac.

**[0102]** In a further embodiment, the desired polynucleotide is positioned at least 80-200 nucleotides from either terminal end of the non-autonomous transposable element, such as an Ac element. In one embodiment, one terminal end of the Ac element comprises the sequence depicted in SEQ ID NO: 139 and wherein the other terminal end of the Ac element comprises the sequence depicted in SEQ ID NO: 140. In one embodiment, SEQ ID NO: 139 is at the 5'-end of the Ac element, while SEQ ID NO: 140 is at the 3'-end of the Ac element.

**[0103]** In a preferred embodiment, the non-autonomous transposable element is an Ac, Spm, or Mu transposable element.

[0104] In one embodiment, the transposase gene is operably linked to a regulatory elements that can express the transposase gene.

[0105] This transposase-transposon cassette may be in a plasmid that is present in a bacterium strain selected from the group consisting of *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, and *MesoRhizobium loti*. Hence, one method of the present invention is a method for transforming a plant with a desired polynucleotide, comprising infecting a plant with such a bacterium strain that contains the transposase-transposon cassette.

[0106] Another aspect of the present invention is a method for transforming a plant, comprising infecting a plant with any one of the transposon-transposase cassettes of the present invention.

[0107] Another aspect of the present invention is a method for transforming a plant, comprising (1) transforming a plant with a transformation plasmid that is suitable for bacterium-mediated plant transformation, wherein the plasmid comprises a transfer-DNA that is delineated by (i) left and right transfer-DNA border sequences, and which comprises (ii) a non-autonomous transposable element, which comprises a desired polynucleotide, and a (iii) a transposase gene, wherein the non-autonomous transposable element and the transposase gene are positioned between the left and right border sequences, and (2) selecting a plant that stably comprises in its genome the non-autonomous transposable element but not the transfer-DNA.

[0108] In one embodiment, at least one of the border sequences of this method comprises a sequence that is (i) nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and (ii) not identical to a bacterial border sequence.

[0109] In another embodiment, the sequence of at least one of the border sequences comprises the sequence depicted in any one of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 190, 194-196, and 198.

[0110] In another embodiment, the step of selecting a plant comprises positively selecting for a plant that comprises the non-autonomous transposable element and counter-selecting against a plant that comprises the transfer-DNA. In another embodiment, the non-autonomous transposable element comprises the terminal ends of any one of an Ac, Spm, or Mu transposable element. In one embodiment, one terminal end of the Ac element comprises the sequence depicted in SEQ ID NO: 139 and wherein the other terminal end of the Ac element comprises the sequence depicted in SEQ ID NO: 140. In another embodiment, the transposase gene is operably linked to regulatory elements that permit expression of the transposase gene in a plant cell.

[0111] In another embodiment, the plasmid that is used to infect the plant is maintained in a bacterium strain selected from the group consisting of *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, and *MesoRhizobium loti*. Accordingly, the present invention also encompasses a method for transforming a plant with a desired polynucleotide, comprising infecting a plant with one of these bacterium strains that contains the transposon-transposase plasmid.

[0112] In another embodiment, a cassette is provided, which comprises (1) a first polynucleotide, comprising a sequence that is (i) nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and (ii) not identical to a bacterial border sequence; (2) a second polynucleotide, which may be (i) an imperfect or perfect repeat of the first polynucleotide, or (ii) a bacterial T-DNA border; and (3) a region comprising a virC2 gene, which may be flanked by regulatory sequences.

[0113] In one embodiment, the region that comprises the virC2 gene, comprises the sequence depicted in SEQ ID NO: 167. In another embodiment, the cassette is in a plasmid suitable for bacterium-mediated transformation.

[0114] Another aspect of the present invention is a method for transforming a plant with a desired polynucleotide, comprising infecting the plant with a bacterium strain comprising any plasmid described herein, wherein the bacterium strain selected

from the group consisting of *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, and *MesoRhizobium loti*.

[0115] In one embodiment, one or more of the polynucleotides, regions, elements, or domains described herein are not 100% identical in nucleotide sequence to a corresponding bacterium sequence. For instance, a polynucleotide comprising a sequence for a cleavage site according to the present invention, is not 100% identical across its length to an *Agrobacterium* right border sequence.

[0116] A transformation cassette may comprise, therefore, sequences that facilitate plant transformation, some, if not all, of which may or may not be identical to a corresponding bacterium sequence. Alternatively, the transformation cassette may comprise one or more bacterial sequences. Thus, the present invention contemplates various permutations of nucleic acid molecules that cover transformation cassettes with no bacterial sequences as well as those that do. For instance, a plant-derived cleavage site might be used in conjunction with a left border sequence from an *Agrobacterium* T-DNA.

[0117] Another aspect of the present invention, is a method for identifying a polynucleotide sequence that is involved in bacterium-mediated plant transformation, comprising:

[0118] (i) isolating a candidate sequence from a source of genetic material;

[0119] (ii) operably replacing one of (a) the first or second polynucleotide, (b) the UI region, (c) the DI region, (d) the UF region, or (e) the AF region of the cassette of claim 1, with the candidate sequence;

[0120] (iii) infecting a plant with the cassette using bacterium-mediated transformation; and

[0121] (iv) determining whether the plant is stably transformed with the desired polynucleotide, wherein a plant that is transformed with the desired

polynucleotide indicates that the candidate sequence is involved in bacterium-mediated plant transformation.

**[0121.1]** Another aspect of the present invention is a vector, comprising: (1) a first polynucleotide, comprising the consensus sequence [A/C/G]-[A/C/T]-[A/C/T]-[G/T]-A-[C/G]-NNNNNN-A-[G/T]-A-[A/C/T]-[A/G]-TCCTG-[C/G/T]-[A/C/G]-N that is consistent with SEQ ID NO:84, which is (i) nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and (ii) not identical to a bacterial border sequence; (2) a second polynucleotide, which is a bacterial T-DNA border; and (3) a desired polynucleotide positioned between the first and second polynucleotides.

**[0121.2]** In an embodiment, the above-mentioned sequence of the first polynucleotide is native to a plant genome. In a further embodiment, the plant sequence is endogenous to (1) a monocotyledonous plant, wherein the monocotyledonous plant is wheat, turf grass, maize, rice, oat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, or palm; or (2) a dicotyledonous plant, wherein the dicotyledonous plant is potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, pea, bean, cucumber, grape, brassica, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, cactus, cucumber, melon, canola, apple, or pine.

**[0121.3]** Another aspect of the present invention is a method for transforming a plant cell, comprising introducing the above-mentioned vector into a plant cell.

**[0121.4]** In an embodiment, the above-mentioned plant cell is located in either (1) a monocotyledonous plant or explant thereof, wherein the monocotyledonous plant or explant thereof is wheat, turf grass, maize, rice, oat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, or palm; or (2) a dicotyledonous plant or explant thereof, wherein the dicotyledonous plant or explant thereof is potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, pea, bean, cucumber,

grape, brassica, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, or cactus.

**[0121.5]** In an embodiment, the above-mentioned vector is a plasmid which is maintained in a bacterial strain, wherein the bacterial strain is *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, or *MesoRhizobium loti*.

**[0121.6]** Another aspect of the present invention is a method for transforming a plant with a desired polynucleotide, comprising infecting a plant with a bacterial strain, wherein the bacterial strain is *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, or *MesoRhizobium loti*, wherein said bacterial strain comprises the above-mentioned vector, and wherein said vector is a plasmid.

**[0121.7]** In an embodiment, the above-mentioned sequence of the consensus sequence of the first polynucleotide is from a potato plant and is SEQ ID NO: 31, SEQ ID NO: 40, or SEQ ID NO: 86.

**[0121.8]** In an embodiment, the above-mentioned sequence of the consensus sequence of the first polynucleotide is from a tomato plant and is SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 42 or SEQ ID NO: 43.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0122]** Figure 1. Sequence requirements for 25-bp cleavage sites. Mismatches to the consensus of *Agrobacterium* Right Borders (CON1) are bold and underlined. Horizontal bars show transformation frequencies compared to those supported by the conventional Right Border Rb02 and the synthetic control cleavage site Ct01, and represent the mean of at least three experiments. The accession numbers of sequences identified in public databases are shown between parentheses. Sequences that were isolated by employing PCR/inverse PCR approaches are

indicated with asterisks. (A) *Agrobacterium* Right Borders, indicated as Rb, are derived from plasmids of *A. tumefaciens* (Rb01, Rb02), *A. rhizogenes* (Rb03, Rb04, Rb05, Rb06 and Rb07), and *A. vitis* (Rb04). (B) Synthetic elements are indicated with Sy. (C) The sequences of plant-derived cleavage sites or cleavage site-like sequences are designated with the initials of the species name followed by a number. (D) The overall consensus for both functional Right Borders and cleavage sites is indicated by CON2.

**[0123]** Figure 2. Sequences flanking right border alternatives.

(A) Upstream sequences display a conserved organization of cytosine/thymine residues separated by adenine-rich trinucleotide spacers. The overdrive sequence of pTi15955 is underlined (dotted). Direct repeats are indicated with grey arrows. Transformation efficacies are shown between parentheses as percentages of controls, and represent the mean  $\pm$  SE of three experiments. “+1” indicates the position of the first base of the right border or right border alternative. ND = not determined.

(B) Helical stability profile (kcal/mol) across the extended 2-kb St02 region of pSIM551 with 60-bp step size and 120-bp window size. (C) Downstream sequences comprise a DR domain (bold) at a distance of one to 27 nucleotides from the border. Plasmids pSIM781, 793, and 843 contain DNA fragments from a potato homolog of AY566555, a potato homolog of AY972080, and an alfalfa homolog of Medicago

truncatula AC131026, respectively. Plasmid pSIM582 contains Le01 flanked by the same tomato DNA sequence that flanks the element in its original genomic context. The 5'-GCCC motif is underlined. Transformation frequencies are shown between parentheses as percentages of controls, and represent the mean  $\pm$  SE of three experiments.

[0124] Figure 3. DNA sequences flanking left borders and left border alternatives. Upstream DNA is italicized with UL domain indicated in bold. Left borders and left border alternatives are highlighted in grey. Cytosine clusters are boxed. Frequencies of transgenic plants containing the designated transfer DNA delineated by borders or border alternatives ('T'), the transfer DNA still attached to backbone sequences ('TB'), and backbone-only ('B') are shown on the right and represent the mean  $\pm$  SE of three experiments. ND = not determined.

[0125] Figure 4. General organization of extended border regions. Putative sites for DnaA and IHF are indicated with open vertical arrows. The primary cleavage and secondary cleavage sites are represented by open boxes. The cleavage sites could be considered to correspond to transfer-DNA right and left borders, respectively. The direction in which DNA unwinds is indicated with a dashed horizontal arrow.

[0126] Figure 5. Schematic of a transposon-transposase construct of the present invention.

[0127] Figure 6. Plasmid maps: (A) pSIM551, pSIM578, pSIM579, pSIM580, and pSIM581; (B) pSIM843B, pSIM108, pSIM831, pSIM829, pSIM401, and pSIM794; (C) pSIM1026, pSIM1008, pSIM781, pSIM844, and pSIM827. "Ori Ec" denotes an origin of replication from bacteria, including *E. coli*. "Ori At" denotes an origin of replication from bacteria, including *Agrobacterium tumefaciens*.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0128] The present invention provides a variety of DNA sequences that are capable of initiating and facilitating the transfer of one polynucleotide into another via standard plant transformation methods. Also identified by the present invention are particular elements within these sequences that help to improve the frequency and integrity of DNA integration. It is an aspect of the present invention that the DNA sequences for any or all of the described transformation elements originate from, or are endogenous to, a plant genome. These transformation elements can be generically described as follows below.

[0129] Cleavage site: a function of the cleavage site is to serve as a recognition site for nuclease proteins or protein complexes that may include virD2 and catalyze a single strand DNA nick within the element during *Agrobacterium*-mediated processing.

[0130] A desired polynucleotide of interest, which is destined for integration into another nucleic acid molecule, may be linked to at least one of such cleavage sites. For example, the desired polynucleotide may be inserted into a plasmid that can be maintained in *Agrobacterium* and has been engineered to contain these elements, such that the desired polynucleotide is ultimately flanked by one or two cleavage sites.

[0131] When there exist two cleavage sites, one may be regarded as being mainly involved in initial cleavage, while the other may be regarded as typically supporting final cleavage. The cleavage sites may be identical in sequence, whereby their functional difference is mediated by specific characteristics of flanking DNA. The transfer DNA contains the initial cleavage site upstream from the final cleavage site. Upstream, with respect to the position of a nucleic acid sequence, means 5'- to the 5'-end of any particular nucleic acid sequence. Downstream, with respect to the position of a nucleic acid sequence, means 3'- to the 3'-end of any particular nucleic acid sequence. All sequences described in this invention refer to the DNA strand that corresponds to the transfer DNA. The non-transfer strand contains the inverse

complement of the final cleavage site upstream from the inverse complement of the initial cleavage site.

[0132] When a desired polynucleotide is flanked by upstream and downstream elements, it is advantageous for the elements to be oriented as either perfect or imperfect direct repeats of each other.

[0133] The sequence of the cleavage site may conform to a consensus sequence, such as that depicted in SEQ ID NO: 84 whereby the sequence of the cleavage site is not identical to an *Agrobacterium* Right Border or Left Border.

[0134] [A/C/G]-[A/C/T]-[A/C/T]-[G/T]-A-[C/G]-NNNNNN-A-[G/T]-A-[A/C/T]-[A/G]-TCCTG-[C/G/T]-[A/C/G]-N (SEQ ID NO: 84)

[0135] The consensus sequence analysis indicates that a DNA sequence that is useful for transferring one polynucleotide into another can accommodate nucleotide degeneracy, especially at its 5'-terminus.

[0136] According to the consensus sequence, a cleavage site may be 25 nucleotides in length. The present invention is not limited to this length, however, but also contemplates longer and shorter cleavage sites that function as described herein. That is, regardless of their length, the cleavage sites should facilitate cleavage for subsequent integration of a desired polynucleotide to which it is linked into another nucleic acid molecule. Accordingly, elements that are 15 nucleotides, 16 nucleotides, 17 nucleotides, 18 nucleotides, 19 nucleotides, 20 nucleotides, 21 nucleotides, 22 nucleotides, 23 nucleotides, 24 nucleotides, 26 nucleotides, 27 nucleotides, 28 nucleotides, 29 nucleotides, and 30 nucleotides elements are envisioned as variants to the 25 nucleotide-long consensus elements described herein.

[0137] The functional activity of a putative cleavage site can be tested by inserting it into a "test plasmid" described in the Examples, and using an *Agrobacterium* strain carrying the resulting vector to transform plants such as tobacco. Transformation frequencies achieved with this vector can then be compared to those of conventional benchmark vectors that contain at least one *Agrobacterium*

T-DNA Right Border to determine the efficacy of the putative cleavage site to mediate DNA transfer.

[0138] Examples of highly efficient synthetic cleavage sites are shown as SEQ ID NOs: 8, 9, 11-13, and 15-17. Similarly efficient plant-derived cleavage sites are depicted in SEQ ID NOs: 28-37 and 85-86. Additional plant-derived cleavage sites that display at least 5% of the activity of Right Borders are shown in SEQ ID NOs: 38-50.

[0139] Assessment of the functional activity of a putative cleavage site is more elaborate. Test vectors used for this purpose contain both a functional site for initial cleavage (or Right Border) and the putative site for final cleavage as described in the Examples. Upon transformation and molecular analysis, plants are separated in two different classes. One class of plants only contains the transfer DNA delineated by cleavage sites. This class of transformation events is designated “desired.” The second class of plants contains the transfer DNA still linked to plasmid backbone sequences. The smaller the percentage of events belonging to this latter “undesired” class, the better the final cleavage site functions in terminating DNA transfer.

[0140] In reference to the DNA strand that comprises the transfer DNA, the position of all DNA regions that are described herein can be identified as upstream and downstream of cleavage sites. The regions include:

[0141] (1) The UI region. A UI region may include one or more of the following characteristics:

[0142] (a) comprises the first base pair of the initial cleavage site and at least about 47 base pairs immediately upstream from this cleavage site,

[0143] (b) is part of a larger sequence that can be predicted by using methods described by, e.g., Huang and Kowalski, 2003, to contain a helical stability that is below the average helical stability, i.e., the sequence may typically requires less energy for unwinding than a random DNA sequence comprising the same number of base pairs,

[0144] (c) is part of an adenine-rich (>25% adenine residues) sequence,

[0145] (d) comprises at least one adenine-cytosine dinucleotide.

[0146] (e) comprises a 45-nucleotide sequence that contains adenine-rich (>25%) trinucleotides interspaced by nucleotides that represent, in at least six cases, a cytosine or thymine (pyrimidine) residue, whereby the most downstream pyrimidine represents either the first base of the initial cleavage site or the base at position -4 relative to the initial cleavage site. See also SEQ ID NOs: 90-97 and 99, and Figures 2A and B.

[0147] (f) may comprise a sequence that shares at least 70% sequence identity with the overdrive depicted in SEQ ID NO: 88,

[0148] (g) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids.

[0149] The UI region may support or enhance any level of initial cleavage activity. For instance, a UI region may enhance the initial cleavage activity by at least 25% compared to the corresponding sequence of the Ti or Ri plasmid.

[0150] (2) The DI region. A DI region may include one or more of the following characteristics:

[0151] (a) comprises at least 45 base pairs immediately downstream from the initial cleavage site,

[0152] (b) comprises a DR domain at a distance of 0-50 base pairs from the initial cleavage site, wherein the DR domain may comprise the sequence depicted in SEQ ID NO: 107,

[0153] (c) optionally contains multiple sequences that are identical or inverse complementary to SEQ ID 115 (CCCG),

[0154] (d) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids, and

[0155] (e) supports or enhances any level of initial cleavage activity. For instance, a DI region may enhance the initial cleavage activity by at least 25% compared to the corresponding sequence of the Ti or Ri plasmid.

[0156] (3) The UF region. A UF region may include one or more of the following characteristics:

[0157] (a) comprises at least 40 base pairs immediately upstream from the final cleavage site,

[0158] (b) comprises at least 55% adenine or thymine residues (AT-rich),

[0159] (c) comprises a sequence that shares at least 70% sequence identity to the UL domain depicted in SEQ ID NO: 120 or to its inverse complement within a distance of about 50 base pairs from the final cleavage site,

[0160] (d) optionally comprises a putative binding site for integration host factor with the consensus sequence [A/T]-ATCAANNNTT-[A/G] (SEQ ID NO: 129),

[0161] (e) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids, and

[0162] (f) supports or enhances any level of initial cleavage activity. For instance, a UF region may enhance the initial cleavage activity by at least 25% compared to the corresponding sequence of the Ti or Ri plasmid.

[0163] (4) the AF region. An AF region may include one or more of the following characteristics:

[0164] (a) comprises at least part of the final cleavage site and at about two to 40 base pairs flanking downstream DNA,

[0165] (b) comprises at least four tightly linked clusters of two or more cytosine bases separated by 1-11 other nucleotides, CCN1-11CCN1-11CCN1-11CC (SEQ ID NO: 122),

[0166] (c) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids, and

[0167] (d) supports or enhances any level of initial cleavage activity. For instance, an AF region may enhance the initial cleavage activity by at least 25% compared to the corresponding sequence of the Ti or Ri plasmid.

[0168] The cytosine cluster domain is thought to form into tertiary quadruplexes at slightly acid or neutral pH, in a similar manner as described for mammalian cytosine clusters. See Zarudnaya *et al.*, Nucleic Acids Res 31: 1375-1386, 2003, and Neidle and Parkinson, Curr Opin Struct Biol 13: 275-283, 2003. It is possible that the specific folding associated with cytosine cluster regions either facilitates or impairs DNA unwinding and/or final cleavage.

[0169] The enzymes necessary for implementing *Agrobacterium*-mediated cleavage include virD2 nicking the top strand of this schematic representation. Figure 4 is a schematic of the transfer cassette within a plasmid for use in *Agrobacterium*-mediated transformation. The elements are oriented in a manner that corresponds to the sequences described herein. Their orientation also corresponds to the strand that is transferred from *Agrobacterium* to plant cells. It is possible to apply the mirror image of this arrangement in combination with the inverse complement of the sequences shown herein, whereby “downstream” becomes “upstream” and vice versa. Typically, the first enzyme nick is made by virD2 and accessory proteins within the initial cleavage site. Sometimes, however, the pertinent enzyme complex does not

effectively make a second nick within the final cleavage site. In this, situation, therefore, the entire top strand of the plasmid becomes linearized, and is transferred to the plant cell.

[0170] On the other hand, effective nicking at both the initial cleavage site and the final cleavage site produces a single-stranded DNA molecule that is terminated by residual portions of the cleavage sites. It is desirable that this particular DNA molecule be integrated into a plant genome.

#### **Source of elements and DNA sequences**

[0171] Any or all of the elements and DNA sequences that are described herein may be endogenous to one or more plant genomes. Accordingly, in one particular embodiment of the present invention, all of the elements and DNA sequences, which are selected for the ultimate transfer cassette are endogenous to, or native to, the genome of the plant that is to be transformed. For instance, all of the sequences may come from a potato genome. Alternatively, one or more of the elements or DNA sequences may be endogenous to a plant genome that is not the same as the species of the plant to be transformed, but which function in any event in the host plant cell. Such plants include potato, tomato, and alfalfa plants. The present invention also encompasses use of one or more genetic elements from a plant that is interfertile with the plant that is to be transformed.

[0172] In this regard, a “plant” of the present invention includes, but is not limited to angiosperms and gymnosperms such as potato, tomato, tobacco, avocado, alfalfa, lettuce, carrot, strawberry, sugarbeet, cassava, sweet potato, soybean, pea, bean, cucumber, grape, brassica, maize, turf grass, wheat, rice, barley, sorghum, oat, oak, eucalyptus, walnut, and palm. Thus, a plant may be a monocot or a dicot. “Plant” and “plant material,” also encompasses plant cells, seed, plant progeny, propagule whether generated sexually or asexually, and descendents of any of these, such as cuttings or seed.. “Plant material” may refer to plant cells, cell suspension cultures, callus, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, seeds, germinating seedlings, and microspores.

Plants may be at various stages of maturity and may be grown in liquid or solid culture, or in soil or suitable media in pots, greenhouses or fields. Expression of an introduced leader, trailer or gene sequences in plants may be transient or permanent.

[0173] One or more traits of a tuber-bearing plant of the present invention may be modified using the transformation sequences and elements described herein. A “tuber” is a thickened, usually underground, food-storing organ that lacks both a basal plate and tunic-like covering, which corms and bulbs have. Roots and shoots grow from growth buds, called “eyes,” on the surface of the tuber. Some tubers, such as caladiums, diminish in size as the plants grow, and form new tubers at the eyes. Others, such as tuberous begonias, increase in size as they store nutrients during the growing season and develop new growth buds at the same time. Tubers may be shriveled and hard or slightly fleshy. They may be round, flat, odd-shaped, or rough. Examples of tubers include, but are not limited to ahipa, apio, arracacha, arrowhead, arrowroot, baddo, bitter casava, Brazilian arrowroot, cassava, Chinese artichoke, Chinese water chestnut, coco, cocoyam, dasheen, eddo, elephant's ear, girasole, goo, Japanese artichoke, Japanese potato, Jerusalem artichoke, jicama, lilly root, ling gaw, mandioca, manioc, Mexican potato, Mexican yam bean, old cocoyam, potato, saa got, sato-imo, seegoo, sunchoke, sunroot, sweet casava, sweet potatoes, tanier, tannia, tannier, tapioca root, topinambour, water lily root, yam bean, yam, and yautia. Examples of potatoes include, but are not limited to Russet Potatoes, Round White Potatoes, Long White Potatoes, Round Red Potatoes, Yellow Flesh Potatoes, and Blue and Purple Potatoes.

[0174] Tubers may be classified as “microtubers,” “minitubers,” “near-mature” tubers, and “mature” tubers. Microtubers are tubers that are grown on tissue culture medium and are small in size. By “small” is meant about 0.1 cm – 1 cm. A “minituber” is a tuber that is larger than a microtuber and is grown in soil. A “near-mature” tuber is derived from a plant that starts to senesce, and is about 9 weeks old if grown in a greenhouse. A “mature” tuber is one that is derived from a plant that has undergone senescence. A mature tuber is, for example, a tuber that is about 12 or more weeks old.

**[01745]** In this respect, a plant-derived transfer-DNA (“P-DNA”) border sequence of the present invention is not identical in nucleotide sequence to any known bacterium-derived T-DNA border sequence, but it functions for essentially the same purpose. That is, the P-DNA can be used to transfer and integrate one polynucleotide into another. A P-DNA can be inserted into a tumor-inducing plasmid, such as a Ti-plasmid from *Agrobacterium* in place of a conventional T-DNA, and maintained in a bacterium strain, just like conventional transformation plasmids. The P-DNA can be manipulated so as to contain a desired polynucleotide, which is destined for integration into a plant genome via bacteria-mediated plant transformation. See Rommens *et al.* in WO2003/069980, US-2003-0221213, US-2004-0107455, and WO2005/004585.

**[01756]** Thus, a P-DNA border sequence is different by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleotides from a known T-DNA border sequence from an *Agrobacterium* species, such as *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*.

**[0177]** A P-DNA border sequence is not greater than 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, 75%, 74%, 73%, 72%, 71%, 70%, 69%, 68%, 67%, 66%, 65%, 64%, 63%, 62%, 61%, 60%, 59%, 58%, 57%, 56%, 55%, 54%, 53%, 52%, 51% or 50% similar in nucleotide sequence to an *Agrobacterium* T-DNA border sequence.

**[01768]** Methods were developed to identify and isolate transfer DNAs from plants, particularly potato and wheat, and made use of the border motif consensus described in US-2004-0107455.

**[01779]** In this respect, a plant-derived DNA of the present invention, such as any of the sequences, cleavage sites, regions, or elements disclosed herein is functional if it promotes the transfer and integration of a polynucleotide to which it is linked into another nucleic acid molecule, such as into a plant chromosome, at a transformation frequency of about 99%, about 98%, about 97%, about 96%, about

95%, about 94%, about 93%, about 92%, about 91%, about 90%, about 89%, about 88%, about 87%, about 86%, about 85%, about 84%, about 83%, about 82%, about 81%, about 80%, about 79%, about 78%, about 77%, about 76%, about 75%, about 74%, about 73%, about 72%, about 71%, about 70%, about 69%, about 68%, about 67%, about 66%, about 65%, about 64%, about 63%, about 62%, about 61%, about 60%, about 59%, about 58%, about 57%, about 56%, about 55%, about 54%, about 53%, about 52%, about 51%, about 50%, about 49%, about 48%, about 47%, about 46%, about 45%, about 44%, about 43%, about 42%, about 41%, about 40%, about 39%, about 38%, about 37%, about 36%, about 35%, about 34%, about 33%, about 32%, about 31%, about 30%, about 29%, about 28%, about 27%, about 26%, about 25%, about 24%, about 23%, about 22%, about 21%, about 20%, about 15%, or about 5% or at least about 1%.

**[0180]** Any of such transformation-related sequences and elements can be modified or mutated to change transformation efficiency. Other polynucleotide sequences may be added to a transformation sequence of the present invention. For instance, it may be modified to possess 5'- and 3'- multiple cloning sites, or additional restriction sites. The sequence of a cleavage site as disclosed herein, for example, may be modified to increase the likelihood that backbone DNA from the accompanying vector is not integrated into a plant genome.

**[0181]** Any desired polynucleotide may be inserted between any cleavage or border sequences described herein. For example, a desired polynucleotide may be a wild-type or modified gene that is native to a plant species, or it may be a gene from a non-plant genome. For instance, when transforming a potato plant, an expression cassette can be made that comprises a potato-specific promoter that is operably linked to a desired potato gene or fragment thereof and a potato-specific terminator. The expression cassette may contain additional potato genetic elements such as a signal peptide sequence fused in frame to the 5'-end of the gene, and a potato transcriptional enhancer. The present invention is not limited to such an arrangement and a transformation cassette may be constructed such that the desired polynucleotide, while operably linked to a promoter, is not operably linked to a terminator sequence.

**[0182]** In addition to plant-derived elements, such elements can also be identified in, for instance, fungi and mammals. See, for instance, SEQ ID NOs: 173-182. Several of these species have already been shown to be accessible to *Agrobacterium*-mediated transformation. See Kunik *et al.*, Proc Natl Acad Sci USA 98: 1871-1876, 2001, and Casas-Flores *et al.*, Methods Mol Biol 267: 315-325, 2004. Thus, the new BOA elements may be used to extend the concept of all-native DNA transformation (Rommens, Trends Plant Sci 9: 457-464, 2004) to organisms, such as eukaryotes, other than plants.

**[0180]** When a transformation-related sequence or element, such as those described herein, are identified and isolated from a plant, and if that sequence or element is subsequently used to transform a plant of the same species, that sequence or element can be described as “native” to the plant genome.

**[0181]** Thus, a “native” genetic element refers to a nucleic acid that naturally exists in, originates from, or belongs to the genome of a plant that is to be transformed. In the same vein, the term “endogenous” also can be used to identify a particular nucleic acid, e.g., DNA or RNA, or a protein as “native” to a plant. Endogenous means an element that originates within the organism. Thus, any nucleic acid, gene, polynucleotide, DNA, RNA, mRNA, or cDNA molecule that is isolated either from the genome of a plant or plant species that is to be transformed or is isolated from a plant or species that is sexually compatible or interfertile with the plant species that is to be transformed, is “native” to, i.e., indigenous to, the plant species. In other words, a native genetic element represents all genetic material that is accessible to plant breeders for the improvement of plants through classical plant breeding. Any variants of a native nucleic acid also are considered “native” in accordance with the present invention. In this respect, a “native” nucleic acid may also be isolated from a plant or sexually compatible species thereof and modified or mutated so that the resultant variant is greater than or equal to 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, 75%, 74%, 73%, 72%, 71%, 70%, 69%, 68%, 67%, 66%, 65%, 64%, 63%, 62%, 61%, or 60% similar in nucleotide sequence to the

unmodified, native nucleic acid isolated from a plant. A native nucleic acid variant may also be less than about 60%, less than about 55%, or less than about 50% similar in nucleotide sequence.

[0185] A “native” nucleic acid isolated from a plant may also encode a variant of the naturally occurring protein product transcribed and translated from that nucleic acid. Thus, a native nucleic acid may encode a protein that is greater than or equal to 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, 75%, 74%, 73%, 72%, 71%, 70%, 69%, 68%, 67%, 66%, 65%, 64%, 63%, 62%, 61%, or 60% similar in amino acid sequence to the unmodified, native protein expressed in the plant from which the nucleic acid was isolated.

[0186] As used herein, “sequence identity” or “identity” in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified region. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences which differ by such conservative substitutions are said to have “sequence similarity” or “similarity”. Means for making this adjustment are well-known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, *Computer Applic.*

*Biol. Sci.*, 4: 11-17 (1988) *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

[0187] As used herein, “percentage of sequence identity” means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0188] Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2: 482 (1981); by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970); by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85: 2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California; GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene* 73: 237-244 (1988); Higgins and Sharp, *CABIOS* 5: 151-153 (1989); Corpet, *et al.*, *Nucleic Acids Research* 16: 10881-90 (1988); Huang, *et al.*, *Computer Applications in the Biosciences* 8: 155-65 (1992), and Pearson, *et al.*, *Methods in Molecular Biology* 24: 307-331 (1994).

[0189] The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against

protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular Biology*, Chapter 19, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); Altschul *et al.*, *J. Mol. Biol.*, 215:403-410 (1990); and, Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997).

[0190] Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length  $W$  in the query sequence, which either match or satisfy some positive-valued threshold score  $T$  when aligned with a word of the same length in a database sequence.  $T$  is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters  $M$  (reward score for a pair of matching residues; always  $> 0$ ) and  $N$  (penalty score for mismatching residues; always  $< 0$ ). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity  $X$  from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters  $W$ ,  $T$ , and  $X$  determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength ( $W$ ) of 11, an expectation ( $E$ ) of 10, a cutoff of 100,  $M=5$ ,  $N=-4$ , and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength ( $W$ ) of 3, an expectation ( $E$ ) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

[0191] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance.

[0192] BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, *Comput. Chem.*, 17:149-163 (1993)) and XNU (Claverie and States, *Comput. Chem.*, 17:191-201 (1993)) low-complexity filters can be employed alone or in combination.

[0193] Multiple alignment of the sequences can be performed using the CLUSTAL method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the CLUSTAL method are KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

### **Transformation bacterium**

[0194] Bacteria species and strains other than those of *Agrobacterium*, e.g., *Agrobacterium tumefaciens*, can be used to transform a plant according to the present invention. For instance, any genera within the family Rhizobiaceae can be used in place of *Agrobacterium* to transform a plant. For instance, members of the *Rhizobium* and *Phyllobacterium* genera can be used to transform a plant according to the present invention. Examples include, but are not limited to, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, *MesoRhizobium loti* bacterial strains, which can be used to transform a plant

according to the present invention. See Broothaerts *et al.*, Nature, 433, pp. 629-633, 2005.

### **Transfer cassette embodiments**

**[01925]** The present invention does not require the presence of all of the elements described herein in the transfer cassette. Any number of permutations of these elements are envisioned. For instance, a transfer cassette may comprise a desired polynucleotide, which is flanked by cleavage sites only.

**[0196]** Alternatively, another transfer cassette may comprise a desired polynucleotide, which is flanked by cleavage sites and which also comprises one or more of the DI and UF regions. The various elements may be arranged as described herein and as depicted in Figures 4, but other arrangements are possible and envisioned by the present invention.

**[0197]** The present invention contemplates, therefore, various permutations of the transformation elements disclosed herein, as well as the use of variant forms of any of the corresponding sequences disclosed herein. See the section on “variants” below.

**[0198]** It may be desirable to select particular elements, and sequences or variant sequences that correspond to those elements, which are effective in transforming a particular plant species. That is, it is possible to use the information disclosed herein, as well as the particular sequences disclosed herein, to optimize transformation efficiency between different organisms or plants of different species.

**[0199]** In this regard, the present invention contemplates transforming a plant with one or more transformation elements that genetically originate from a plant. The present invention encompasses an “all-native” approach to transformation, whereby only transformation elements that are native to plants are ultimately integrated into a desired plant via transformation. In this respect, the present invention encompasses transforming a particular plant species with only genetic transformation elements that are native to that plant species. The native approach

may also mean that a particular transformation element is isolated from the same plant that is to be transformed, the same plant species, or from a plant that is sexually interfertile with the plant to be transformed.

[0200] On the other hand, the plant that is to be transformed, may be transformed with a transformation cassette that contains one or more genetic elements and sequences that originate from a plant of a different species. It may be desirable to use, for instance, a cleavage site, UI, DI, UF, or DF region sequence that is native to a potato genome in a transformation cassette or plasmid for transforming a tomato or pepper plant, for example.

[0201] The present invention is not limited, however, to native or all-native approach. A transformation cassette or plasmid of the present invention can also comprise sequences and elements from other organisms, such as from a bacterial species.

### **Desired polynucleotides**

[0202] The origin of the genetic sequences that make up the transformation cassette also may apply to the sequence of a desired polynucleotide that is to be integrated into the transformed plant. That is, a desired polynucleotide, which is located between the primary or initial and secondary or final cleavage site sequences of the present invention, may or may not be “native” to the plant to be transformed. As with the other transformation elements, a desired polynucleotide may be isolated from the same plant that is to be transformed, or from the same plant species, or from a plant that is sexually interfertile with the plant to be transformed. On the other hand, the desired polynucleotide may be from a different plant species compared to the species of the plant that is to be transformed. Yet, the present invention also encompasses a desired polynucleotide that is from a non-plant organism.

[0203] A desired polynucleotide of the present invention may comprise a part of a gene selected from the group consisting of a *PPO* gene, an *RI* gene, a type L or H alpha glucan phosphorylase gene, an UDP glucose glucosyltransferase gene, a *HOS1* gene, a S-adenosylhomocysteine hydrolase gene, a class II cinnamate 4-

hydroxylase gene, a cinnamoyl-coenzyme A reductase gene, a cinnamoyl alcohol dehydrogenase gene, a caffeoyl coenzyme A O-methyltransferase gene, an actin depolymerizing factor gene, a Nin88 gene, a Lol p 5 gene, an allergen gene, a P450 hydroxylase gene, an ADP-glucose pyrophosphorylase gene, a proline dehydrogenase gene, an endo- 1,4-beta-glucanase gene, a zeaxanthin epoxidase gene, a 1-aminocyclopropane-1-carboxylate synthase gene, an Rb resistance gene, a Bf2 resistance gene, a Fad2 gene, and an Ant-1 gene. Such a desired polynucleotide may be designed and oriented in such a fashion within a transformation cassette of the present invention, so as to reduce expression within a transformed plant cell of one or more of these genes. See, for instance, Rommens *et al.* in WO2003/069980, US-2003-0221213, US-2004-0107455, and WO2005/004585.

**[0197204]** Thus, a desired polynucleotide of the present invention may be used to modify a particular trait in a transformed plant that is normally manifested by an untransformed plant. For instance, a desired polynucleotide may be placed into a transformation cassette of the present invention to enhance the health and nutritional characteristics of the transformed plant or it may be used, for instance, to improve storage, enhance yield, enhance salt tolerance, enhance heavy metal tolerance, increase drought tolerance, increase disease tolerance, increase insect tolerance, increase water-stress tolerance, enhance cold and frost tolerance, enhance color, enhance sweetness, improve vigor, improve taste, improve texture, decrease phosphate content, increase germination, increase micronutrient uptake, improve starch composition, and improve flower longevity.

### **Transformation vector embodiments**

**[0205]** The present invention does not require the presence of all of the elements described herein in the transformation vector. Any number of permutations of these elements are envisioned. For instance, a transformation vector may comprise both a transfer cassette and one or more UI and AF regions. The elements may be arranged as described herein and as depicted in Figures 4, but other arrangements are possible and envisioned by the present invention.

**[0206]** Transformation of a plant is a process by which DNA is stably integrated into the genome of a plant cell. “Stably” refers to the permanent, or non-transient retention and/or expression of a polynucleotide in and by a cell genome. Thus, a stably integrated polynucleotide is one that is a fixture within a transformed cell genome and can be replicated and propagated through successive progeny of the cell or resultant transformed plant. Transformation may occur under natural or artificial conditions using various methods well known in the art. See, for instance, METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY, Bernard R. Glick and John E. Thompson (eds), CRC Press, Inc., London (1993); Chilton, Scientific American, 248(6), pp. 36-45, 1983; Bevan, Nucl. Acids. Res., 12, pp. 8711-8721, 1984; and Van Montague *et al.*, Proc R Soc Lond B Biol Sci., 210(1180), pp. 351-65, 1980. Plants also may be transformed using “Refined Transformation” and “Precise Breeding” techniques. See, for instance, Rommens *et al.* in WO2003/069980, US-2003-0221213, US-2004-0107455, WO2005/004585, US-2004-0003434, US-2005-0034188, WO2005/002994, and WO2003/079765.

**[0207]** Transformation may rely on any known method for the insertion of nucleic acid sequences into a prokaryotic or eukaryotic host cell, including the bacterium-mediated transformation protocols described herein, such as *Agrobacterium*-mediated transformation, or alternative protocols, such as by viral infection, whiskers, electroporation, heat shock, lipofection, polyethylene glycol treatment, micro-injection, and particle bombardment.

**[01988]** “Activity of the final cleavage site” is determined by comparing the number of transformed plants only containing the DNA that is positioned between initial and final cleavage site with the total number of transformed plants. The final cleavage site determines the fidelity of DNA transfer.

**[0209]** “Activity of the initial cleavage site” is assessed by determining the transformation frequency of a plasmid carrying this cleavage site. Activity is dependent on both the sequence of the initial cleavage site itself and the sequence of

flanking DNA. Activities are often expressed as a percentage of the activity of conventional Right Borders. Effective initial cleavage sites display at least 50% of the activity of Right Borders if flanked by DNA sequences that support their activity. Using methods and strains described in this invention, transformation frequencies for conventional right borders average about 10-20 calli/tobacco explant.

[0210] “Bacterium-mediated plant transformation” is the modification of a plant by infecting either that plant or an explant or cell derived from that plant with a bacterium selected of the group consisting of *Agrobacterium* sp., *Rhizobium* sp., *Phyllobacterium* sp., *SinoRhizobium* sp., and *MesoRhizobium* sp. to transfer at least part of a plasmid that replicates in that bacterium to the nuclei of individual plant cells for subsequent stable integration into the genome of that plant cell.

[0211] “Cassette” is a DNA sequence that may comprise various genetic elements.

[0212] “Cleavage site” is a DNA sequence that is structurally different but functionally similar to T-DNA borders. A cleavage site comprises a sequence that is nicked when exposed to an enzyme involved in bacterium-mediated plant transformation. It can represent a synthetic sequence that may not be present in the genome of a living organism or it can represent a sequence from a living organism such as a plant, animal, fungus, or bacterium.

[0213] “Conventional binary plasmid” is a plasmid that can be maintained in both *E. coli* and *A. tumefaciens*, and contains T-DNA right and left borders that are flanked by at least 10 base pairs of DNA that flank these elements in *Agrobacterium* Ti or Ri plasmids.

[0214] “Final cleavage site” is a DNA sequence that is structurally or sequentially different, but functionally similar to, the Left Border of *Agrobacterium* Ti plasmids by comprising a sequence mediating a second cleavage reaction and, thus, defining the end point of the transfer DNA. An effective final cleavage site allows

transfer of DNA sequences that do not include sequences downstream from the final cleavage site, i.e., plasmid backbone sequences.

[0215] “A flanking sequence” is a sequence immediately next to another sequence.

[0216] “Initial cleavage site” is a DNA sequence that is structurally different but functionally similar to the Right Border of *Agrobacterium* Ti plasmids by comprising a sequence that functions as initial cleavage site and, thus, defines the start point of the transfer DNA. An effective initial cleavage site supports or enhances plant transformation compared to a conventional Right Border.

[0217] “Non-autonomous transposable element” as used herein is a transposable element that comprises the ends that are required for transposition but which does not encode the protein that is required for transposition. Thus, a non-autonomous transposable element will transpose only if the gene encoding the protein required for transposition is expressed from either a different position in the genome or from a plasmid or DNA fragment that resides in the same plant cell.

[0218] A “terminal end of a transposable element” is a sequence at the 5’ or 3’ end of a transposable element that is required for non-autonomous transposition. Such sequences may comprise about 100 to about 300 nucleotides.

[0219] “T-DNA border” is a polynucleotide of approximately 25-base pairs in length that comprises a sequence that can be nicked when exposed to an enzyme or enzyme complex involved in bacterium-mediated plant transformation and that can define the single stranded DNA fragment that is transferred from the bacterium to the plant cell.

[0220] “UF region” is a DNA sequence that (a) comprises at least 40 base pairs immediately upstream from either the final cleavage site or left border, (b) comprises at least 55% adenine or thymine residues (AT-rich), (c) comprises a sequence which has at least 70% sequence identity to the UL domain depicted in SEQ ID NO: 120 or its inverse complement, within a distance of 50 base pairs from the

final cleavage site, (d) optionally comprises a putative binding site for integration host factor with the consensus sequence [A/T]-ATCAANNNTT-[A/G] (SEQ ID NO: 129) that is positioned within 200 base pairs from the final cleavage site or left border, (e) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids, and (f) supports or enhances activity of the initial cleavage site.

[0221] “UI region” is a DNA sequence that (a) comprises the first base pair of either the initial cleavage site or right border and at least about 47 base pairs immediately upstream from this cleavage site; (b) is part of a larger sequence that can be predicted by using methods described by, e.g., Huang and Kowalski, 2003, to contain a helical stability that is below the average helical stability, i.e., the sequence may typically requires less energy for unwinding than a random DNA sequence comprising the same number of base pairs; (c) is part of an adenine-rich (>25% adenine residues) sequence; (d) comprises at least one adenine-cytosine dinucleotide; (e) comprises a 45-nucleotide sequence that contains adenine-rich (>25%) trinucleotides interspaced by nucleotides that represent, in at least six cases, a cytosine or thymine (pyrimidine) residue, whereby the most downstream pyrimidine represents either the first base of the initial cleavage site or the base at position -4 relative to the initial cleavage site. See also SEQ ID NOs: 199-208, and Figures 2A and B; (f) may comprise a sequence with at least 70% sequence identity to the overdrive depicted in SEQ ID NO: 88; (g) is not identical to a region that flanks a T-DNA border in *Agrobacterium* Ti or Ri plasmids; and (h) supports or enhances activity of the initial cleavage site.

[0222] “UI-like region” is a sequence that resembles a UI region but differs in that it (1) represents *Agrobacterium* sequences flanking a Right Border, or (2) impairs the efficacy of a Right Border or cleavage site. The UI-like region may reduce transformation frequencies to less than that of a conventional Right order-flanking DNA sequence. For instance, it may reduce a transformation frequency to less than about 25%.

[0223] “Transformation vector” is a plasmid that can be maintained in *Agrobacterium*, and contains at least one Right Border or initial cleavage site. Infection of explants with *Agrobacterium* strains carrying a transformation vector and application of transformation procedures will produce transformed calli, shoots, and/or plants that contain at least part of the transformation vector stably integrated into their genome. The vector may comprise a selectable marker to aid identification of plants that have been stably transformed.

[0224] A “selectable marker” is typically a gene that codes for a protein that confers some kind of resistance to an antibiotic, herbicide or toxic compound, and is used to identify transformation events. Examples of selectable markers include the streptomycin phosphotransferase (spt) gene encoding streptomycin resistance, the phosphomannose isomerase (pmi) gene that converts mannose-6-phosphate into fructose-6 phosphate; the neomycin phosphotransferase (nptII) gene encoding kanamycin and geneticin resistance, the hygromycin phosphotransferase (hpt or aphiv) gene encoding resistance to hygromycin, acetolactate synthase (als) genes encoding resistance to sulfonylurea-type herbicides, genes coding for resistance to herbicides which act to inhibit the action of glutamine synthase such as phosphinothricin or basta (e.g., the bar gene), or other similar genes known in the art.

[0225] A “variant,” as used herein, such as a variant of any of the nucleic acid molecules or polypeptides described herein, is understood to mean a nucleotide or amino acid sequence that deviates from the standard, or given, nucleotide or amino acid sequence of a particular gene or protein. The terms, “isoform,” “isotype,” “homolog,” “derivative,” and “analog” also refer to “variant” forms of a nucleotide or an amino acid sequence. An amino acid sequence that is altered by the addition, removal or substitution of one or more amino acids, or a change in nucleotide sequence, may be considered such a “variant” sequence. The variant may have “conservative” changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. A variant may have “nonconservative” changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or

both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted may be found using computer programs well known in the art such as Vector NTI Suite (InforMax, MD) software.

**[0226]** The present invention encompasses a variant that has one or more point mutations compared to one of the sequenced disclosed herein. For instance, any one of the cleavage site sequences depicted by SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 194-196, may comprise one or more point mutations. That mutated variant may then be readily tested for activity or its effect on transformation efficiency, simply by replacing the original sequence with the mutated version and determining whether the sequence is cleaved and whether the efficiency of transformation is maintained, increased, or decreased.

**[0227]** Similarly, any of the sequences disclosed herein for a UI, DI, UF, or AF region may be mutated and similarly tested for activity and effect on transformation efficiency.

**[0228]** Thus, the present invention is not limited to the sequences disclosed herein that correspond to a particular transformation element. Rather, actual sequences can be used in any permutation to create useful and effective transformation cassettes and plasmids, or one or more of the component transformation elements may be mutated, tested for activity, and then incorporated into a desired transformation cassette or plasmid.

**[0229]** In this regard, a variant sequence of the present invention, such as a variant of a cleavage site or UI, DI, UF, or AF region, may be a functional homolog of a particular sequence. By this it is understood that a cleavage site that is a variant of, for instance, one of SEQ ID NOs: 8, 9, 11-13, 15-17, 28-37, 38-51, 85-86, 189, 194-196, but which still can be cleaved by an enzyme, is a functional derivative of the original sequence. By the same token, the present invention encompasses functional derivatives of any of all of the transformation elements, e.g., UI, DI, UF, and AF regions, disclosed herein.

[0230] A variant sequence of the present invention also encompasses shorter and longer sequences of those specific sequences disclosed herein. For instance, the cleavage site sequence depicted in SEQ ID NO: 8 may be positioned within a larger fragment of DNA, which may or may not be plant DNA. The subsequently larger fragment may then be inserted into a transformation cassette or plasmid. Thus, the present invention is not limited to manipulating only a polynucleotide that consists of a particular SEQ ID NO: sequence. Accordingly, one may use one of the sequences of the present invention, such as SEQ ID NO: 8, to identify and isolate another sequence homolog from a plant or any other organism genome. It may be desirable to isolate a fragment of that genomic DNA that includes sequences flanking the homolog of interest. The larger fragment, within which is included the same or similar homolog to a desired sequence described herein, may then be tested according to the methods described herein for functional activity, i.e., it may be tested to determine what effect, if any, it has on transformation efficiency in comparison to a control system that does not include the larger fragment homolog. Thus, a “variant” of any of the sequences described herein, not only that exemplified by SEQ ID NO: 8, be it a sequence for a cleavage site or for a UI, DI, UF, or AF region, for instance, encompasses longer versions of the corresponding sequences disclosed herein.

[0231] Conversely, a “variant” of the present invention also encompasses polynucleotides that are shorter than a corresponding sequence of the present invention. That is a variant polynucleotide may be “a part of” a sequence disclosed herein. It is well within the purview of the skilled person to make truncated versions of a sequence disclosed herein. For instance, the present invention contemplates truncating a cleavage site, for instance, by any number of nucleotides and then testing that cleavage site for activity. For example, one may truncate the cleavage site depicted in SEQ ID NO: 8 by removing the 5 nucleotides from the 3'-end of SEQ ID NO: 8 and then test that truncated fragment of SEQ ID NO: 8 for cleavage activity. That is, one may test to see if a pertinent enzyme can still cleave the truncated SEQ ID NO: 8, by virtue of assaying for the cleavage directly or by ascertaining the effect of the truncated SEQ ID NO: 8 on transformation efficiency compared to a control system, which employs the full-length sequence of SEQ ID NO: 8.

**[022132]** A truncation may be made at either end or within a particular sequence described herein. Thus, a variant that comprises a part of, say, SEQ ID NO: 8, may be any part of SEQ ID NO: 8. SEQ ID NO: 8 is only used here as an example. Any of the sequences disclosed herein may be truncated in such fashion and then tested for subsequent activity and/or transformation efficiency.

**[0233]** Any of the sequences described herein can be chemically synthesized. That is, it may not be necessary to physically isolate and purify a particular sequence from an organism genome prior to use. For this reason, a “truncated” version of a sequence described herein may be obtained by terminating chemical synthesis at any desired time point during manufacture.

**[0234]** Thus, a variant that is a “part of” a sequence disclosed herein may be made directly using chemical synthesis techniques rather than physically obtained from the actual polynucleotide in question. The same strategy applies for the longer variant forms: it is possible to chemically synthesize a polynucleotide, within which comprises a particular sequence described herein.

**[0235]** The following examples serve to illustrate various embodiments of the present invention and should not be construed, in any way, to limit the scope of the invention.

**[0237]** Having now fully described this invention, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters, concentrations and conditions, without undue experimentation. This application is intended to cover any variations, uses, or adaptations of the invention, following in general the principles of the invention, that include such departures from the present disclosure as come within known or customary practice

within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth.

### **EXAMPLE 1**

#### **Initial cleavage sites**

[0238] Isolated plant sequences were used as effective initial cleavage sites to mediate DNA transfer as well as effective final cleavage sites to limit the co-transfer of vector backbone sequences. In fact, backbone transfer frequencies with plant-derived cleavage sites that were linked to upstream AT-rich regions and downstream C-cluster regions were lower than obtained with conventional Left Borders. The DNA sequences described herein permits the construction of efficient all-native transfer DNAs that can be used for the production of intragenic potato, tomato, and alfalfa plants.

#### **Cleavage sites**

[0239] Initial cleavage sites function in the initiation of DNA transfer and are positioned in transformation plasmids at the junction of (i) the 5'-end of sequences destined for transfer from *Agrobacterium* to plant cells (the transfer DNA) and (ii) plasmid backbone sequences required for maintenance of the plasmid in *Agrobacterium*. Their sequences deviate from that of the *Agrobacterium* Right Borders shown in SEQ ID NOs: 1-7 denoted Rb01-Rb07, respectively. Examples of synthetic initial cleavage sites are depicted in SEQ ID NOs: 8-13, which are denoted Sy01-Sy13.

[0240] To test the functional activity of putative initial cleavage sites, such sequences were linked to (i) an upstream 109-base pair *Agrobacterium* pTi15955 sequence preceding the conventional right border (SEQ ID NO: 1), and (ii) a DI region shown in SEQ ID NO: 22. This construct was inserted into a plasmid containing an expression cassette for the neomycin phosphotransferase (nptII) selectable marker gene. *Agrobacterium* strains carrying the resulting 'single element' test vector were subsequently used to infect tobacco explants.

[0241] Two weeks after infection, the average numbers of calli per explant were compared to those produced with a control plasmid containing Rb01 ( $15.3 \pm 0.5$ ). As shown in Figure 1, all putative cleavage sites enabled DNA transfer. However, base substitutions C6A, A13C, C19G, C20G, and T21A of cleavage site Sy03, Sy07, Sy11, Sy12, and Sy13, respectively, lowered transformation frequencies more than five-fold.

[0242] Sequence requirements for initial cleavage were further determined by testing the efficacy of plant sequences that resemble the *Agrobacterium* consensus (Figure 1). In addition to the cleavage site of a previously characterized *Solanum tuberosum* (potato) P-DNA (Rommens *et al.*, Plant Physiol 135: 421-431, 2004), designated here as St01 (SEQ ID NO: 23), a large number of new elements were identified by searching publicly available databases including those maintained by "The National Center For Biotechnology Information" using, for instance, the "Motif Alignment and Search Tool" (Bailey and Gribskov, J Comput Biol 5: 211-21, 1998) and "advanced BLASTN" (Altschul *et al.*, Nucleic Acids Res 25: 3389-3402, 1997). Search motifs included CAGGATATATNNNNNGTA (SEQ ID NO: 130), using parameters such as (i) penalty for nucleotide mismatch = -1, and (ii) expect = 105. All hits were further analyzed to determine whether they uncovered sequences resembling CON1 and/or CON2. Additional databases that were searched include those covering Solanaceae ([www.sgn.cornell.edu/](http://www.sgn.cornell.edu/)), Compositae ([compositdb.ucdavis.edu/](http://compositdb.ucdavis.edu/)), and *Medicago truncatula* ([www.genome.ou.edu/medicago.html](http://www.genome.ou.edu/medicago.html)). Alternatively, border-like sequences were isolated from genomes by employing a polymerase chain reaction (PCR) approach. For this purpose, plant DNAs (2  $\mu$ g), partially digested with SauIIA, were ligated with 192-bp BamHI - EcoRV fragments of pBR322. The resulting DNAs were used as templates for amplification with a degenerate primer, SEQ ID NO: 24, and an anchor primer, SEQ ID NO: 25, with 49°C annealing temperature and 2.5-minute extension time. Subsequent PCRs were performed with the amplified DNAs ligated with pGEM-T as templates using the degenerate primer together with either SP6 or T7 primers at a slightly higher annealing temperature (52°C). The products of these reactions were

inserted into pGEM-T and sequenced to design primers for conventional inverse PCRs to determine the actual putative cleavage site sequences.

[0243] Among the new plant-derived cleavage sites, only the *Arabidopsis thaliana* At01 element (SEQ ID NO: 26) fully matched the *Agrobacterium* right border consensus.

[0244] However, this element displayed only 65% of the activity of the conventional Right Border Rb02. The lower activity of At01 suggests that the guanine base at position +4 (G4) is not as effective as T4.

[0245] Most cleavage sites contain at least one mismatch with the consensus sequence of *Agrobacterium* Right Borders (CON1) shown in Figure 1 and depicted in SEQ ID NO: 27:

[0246]

[A/C/G][A/T][A/T][G/T]AC[A/C/T]N[C/G/T][A/C/G][A/C/G][A/C/G]ATATATCC  
TG[C/T]CA (SEQ ID NO: 27)

[0247] Despite the presence of one to three mismatches with CON1, the following cleavage site displayed at least 50% activity. This result demonstrates that *Agrobacterium* appears to not have exploited the full potential of border sequence variation. See SEQ ID NOs: 28-37. Other cleavage sites include those depicted in SEQ ID NOs: 38 and 39. Cleavage sites that displayed activities between about 50% and 5% are depicted in SEQ ID NOs: 40-50.

[0248] Mismatches and/or point deletions in 31 cleavage site-like sequences from a variety of plant species resulted in either low activity (less than about 5%) or no detectable activity at all. See the sequences depicted in SEQ ID NOs: 38, 39, 52-83, 193, and 197.

[0249] By comparing tested Right Borders, cleavage sites, and cleavage site-like elements, a consensus, CON2, was identified. See Figure 1D and SEQ ID NO: 84:

[0250] 5'-[A/C/G]-[A/C/T]-[A/C/T]-[G/T]-A-[C/G]-NNNNNN-A-[G/T]-A-[A/C/T]-[A/G]-TCCTG-[C/G/T]-[A/C/G]-N (SEQ ID NO: 84).

[0251] Mismatches that reduced transformation frequencies most dramatically include, apart from those mentioned above, A5G and C6G.

[0252] The high activity of tomato Le01 prompted us to search for homologs in related plant species. Identification of identical copies in pepper (Ca01, SEQ ID NO: 85) and potato (St02, SEQ ID NO: 86) DNAs indicates that a single cleavage site can be used for all-native DNA transformation of at least three different *Solanceous* plant species, potentially facilitating the governmental approval process. We also identified a potato homolog of tomato Le05. However, the reduced efficacy of that cleavage site may limit its applicability for plant transformation.

[0253] To obtain an effective cleavage site for use in maize, we can modify Zm01 (SEQ ID NO: 50) by replacing a single base pair. Substitution of the guanine residue at position 3 by a thymine residue will yield a Zm01-derived cleavage site, designated Zm01M1 (SEQ ID NO: 51).

[0254] Similarly, an effective *Brassica* cleavage site can be obtained by modifying SEQ ID NO: 52 to create SEQ ID 189, or by modifying SEQ ID NO: 197 to produce SEQ ID NO: 198.

[0255] Efficient cleavage sites for soybean can be obtained by modifying Gm01 (SEQ ID NO: 38) and Gm02 (SEQ ID NO: 39) to create Gm01M1 (SEQ ID NO: 195) and Gm02M1 (SEQ ID NO: 196), respectively.

## **EXAMPLE 2**

### **Spacing requirements for an extended overdrive domain**

[0256] The effective test plasmid pSIM551 contained St02 linked to the sequences that contain a 31-bp fragment of pTi15955 inserted between novel sequences. The DNA region comprising this sequence and the first nucleotide of Le01 is the part of SEQ ID NO: 87 depicted in SEQ ID NO: 199, and represents a UI

region. This arrangement placed the cleavage site for potato at a distance of 12 base pairs from the overdrive, an element that was reported to promote DNA transfer (van Haaren *et al.*, 1987) and depicted in SEQ ID NO: 88.

[0257] Although the overdrive element is believed to function in a position independent manner (Shurvinton and Ream, 1991), we found that a single base pair insertion between St02 and upstream DNA (SEQ ID NO: 89) in pSIM578 reduced transformation frequencies of pSIM579 about two-fold (Figure 3A). Furthermore, the 5'-CAA trinucleotide insertion into the UI region of pSIM579 (SEQ ID NO: 90) had an even greater negative effect on the efficacy of transformation, lowering it to 35%.

[0258] To study the molecular basis of the apparent overdrive-St02 spacing requirement, we compared the UI region of pSIM551 (SEQ ID NO: 199) with corresponding T-DNA flanking regions of *Agrobacterium* plasmids (SEQ ID NOs: 91-97 shown in SEQ ID NOs: 200-206). The aligned sequences generally contained cytosine or thymine residues at conserved four-nucleotide intervals, separated by adenine-rich (46%) trinucleotide segments (Figure 3A). This arrangement resulted in a high occurrence of AC dinucleotide repeats (27%) approaching that of the overdrive element itself (42%).

[0259] Whereas the sequences upstream from (1) the Right Borders of *Agrobacterium* plasmids and (2) the UI region of pSIM551 comprised at least six pyrimidine residues at conserved positions, the impaired activity of pSIM578 and 579 was correlated with UI regions that contained five and four such residues, respectively (Figure 2A). Additional evidence for the importance of correctly spaced pyrimidines was obtained by analyzing the UI region of pSIM580, which contained the pentanucleotide 5'-ACCAA insertion between St02 and upstream DNA (part of SEQ ID NO: 98 shown in SEQ ID NO: 207). Maintenance of six pyrimidines at conserved positions in this plasmid was associated with the same DNA transfer activity as that of the original vector pSIM551 (Figure 2A).

[0260] To further test the functional significance of correctly spaced pyrimidines, the UI region of pSIM551 was replaced by a sequence that displayed

77% identity with the *Agrobacterium* pRi2659 sequences upstream from the right border (Hansen *et al.*, 1992). Immediate linkage with St02 yielded a UI region (part of SEQ ID NO: 99 shown in SEQ ID NO: 208) in pSIM844 that supported high transformation frequencies (125%) (Figure 2A). However, disruption of the pyrimidine spacing by a single base pair insertion resulted in a UI-derived region of pSIM827 (part of SEQ ID NO: 100 shown in SEQ ID NO: 207) that lowered transformation frequencies to 7%.

**[0261]** Having correlated the original spacing of pyrimidines with efficient DNA transfer, we now also tested the functional relevance of adenine-rich spacers. For this purpose, the UI region of pSIM551 was replaced with a tomato DNA fragment carrying nine pyrimidines at conserved positions but lacking a high percentage of adenine residues in the intervals (part of SEQ ID NO: 101 shown in SEQ ID NO: 210). The resulting vector pSIM581 displayed only 15% of the transformation efficacy of pSIM551, indicating that adenine-rich intervals or AC repeats play a role in the functional activity of the UI region (Figure 2A).

**[0262]** Since adenine-rich DNA is often associated with low helical stability regions, we determined the helical stability profile of pSIM551 using WEB THERMODYN™ (Huang and Kowalski, 2003). This analysis identified a 120-bp sequence immediately upstream from the St02 cleavage site and including the UI region to represent the lowest helical stability region of the pSIM551 backbone (Figure 2B and data not shown). The association of an easily unwound DNA region immediately upstream from the RBA may be functionally relevant because *Agrobacterium* Ti and Ri plasmids contain similar low helical stability regions at their Right Borders. For instance, pTiC58 contains a 120-bp region preceding the border with a stability of 116 kcal/mol. Analogous to the association of low helical stability regions with the initiation of plasmid replication (Natale *et al.*, 1993), these upstream DNAs may be involved in the initiation of DNA transfer. We conclude that the overdrive is part of a larger UI-like region that is conserved among *Agrobacterium* plasmids. This domain supports St02-mediated DNA transfer if correctly spaced relative to the initial cleavage site, and may be involved in local DNA unwinding.

The sequence that comprises the first nucleotide of the initial cleavage site and at least about 47 nucleotides of flanking upstream DNA is designated UI region.

### **EXAMPLE 3**

#### **The role of sequences downstream from initial cleavage sites**

[0263] Given that upstream DNA sequences adjacent to the border region influenced transformation efficacy, we sought to test the effect of downstream modifications. As shown in Figure 2C, analyses of the sequences downstream from Right Borders and depicted in SEQ ID NOs: 102-106 identified decamers that shared the consensus 5'-[A/C/T]-[A/C]-[A/C/T]-[A/G/T]-[A/T]-T-[A/C]-G-[G/T]-[G/T] (SEQ ID NO: 107) with the 5'-part of the overdrive, and were positioned at a distance of one to 27 nucleotides from the right border. This "downstream from right border" (DR) domain was also identified in both the potato-derived transfer DNA (Rommens *et al.*, 2004) of pSIM108 (SEQ ID 108) and DI regions of test vectors such as pSIM551 (SEQ ID NO: 109) (Figure 2C). An increase in the spacing between Le01 and DR domain from 24 nucleotides in the DI region of pSIM551 to 48 nucleotides in pSIM920 (SEQ ID NO: 110) lowered transformation frequencies by 40% (Figure 3C), indicating that the supporting function of DR domain on border activity is spacing dependent.

[0264] Because downstream DNA sequences represent the actual transfer DNA that is intended for plant transformation, we replaced the original bacterial sequences of pSIM551 with two unique potato DNA fragments. The pSIM551-derivative pSIM793 (SEQ ID NO: 113), which contained a DR domain at 27 nucleotides from Le01 yielded about the same transformation frequency as pSIM551. In contrast, the potato DNA fragment of pSIM582 (SEQ ID NO: 112), which contained a DR domain with several mismatches to the consensus, displayed only 59% activity. Interestingly, replacement of Le01-flanking DNA sequences by an alfalfa DNA fragment that contained two different DR domains (SEQ ID NO: 114) triggered unusually high transformation frequencies for the resulting vector pSIM843 (168%) (Figure 3C). This high activity may also be due, in part, to the specific

sequence of the upstream DNA of pSIM843, which contains eight 5'-GCCC (SEQ ID NO: 115) repeats. We conclude that sequences flanking right border alternatives play an important role in supporting plant DNA transfer. These sequences comprise upstream ACR and downstream DR domains.

#### **EXAMPLE 4**

##### **Substitution of left borders by right border alternatives**

[0265] The above-described studies had shown that CON2-matching 25-bp elements function as effective right border alternatives if flanked by sequences that support their activity. As shown in SEQ ID NOs: 116-119, functional differences exist, and there is divergent sequence organization, at and around, the left and right border sites. In contrast to right borders, for instance, left borders:

[0266] (1) are preceded by AT-rich DNAs each comprising an “upstream from left border” (UL) domain on either DNA strand with the consensus sequence A[C/T]T[C/G]A[A/T]T[G/T][C/T][G/T] [C/G]A[C/T][C/T][A/T] (SEQ ID NO: 120);

[0267] (2) share a more conserved consensus sequence:

[0268] 5'-[A/G]TTTACA[A/C/T][A/C/T]  
[A/C/T][C/G]AATATATCCTGCC[A/G] (SEQ ID NO: 121); and

[0269] (3) are linked to downstream plasmid backbone DNA by cytosine clusters (“C-clusters”) that conform to the consensus CCN1-11CCN1-11CCN1-11CC (SEQ ID NO: 122) (Figure 3A).

[0270] Direct evidence for the role of the C-cluster organization in supporting left border activity was obtained by comparing the fidelity of DNA transfer for pSIM831 and 829. Both vectors contained an expression cassette for the nptII gene preceded by DNA regions comprising St02 as right border alternative, and were confirmed to support the same high transformation frequencies as pSIM551 (data not shown). The vectors also contained almost identical DNA regions for secondary cleavage, shown in SEQ ID NOs: 123 and 124, respectively, which

differed only in that pSIM829 contained a 10-bp insertion in the fourth left border-associated C-cluster (Figure 3B).

[0271] The effect of this small change was assessed by classifying regenerated shoots in three groups based on PCR analyses. The first 'T' group only contained the intended transfer DNA, and would therefore be predicted to have arisen from primary cleavage events at the right border followed by secondary cleavage at the left border. Plants containing both the transfer DNA and additional backbone DNA sequences were classified in a second "TB" group, and most likely represented events where the second copy of the border alternative failed to function in terminating DNA transfer. The third 'B' group of events only contained backbone DNA, and probably arose from initial cleavage reactions at the second St02 copy. This genotype classification demonstrated that pSIM831 was more than twice as effective as pSIM829 (41% vs. 17%) in producing 'T' events (Figure 3B).

[0272] The sequence comprising at least part of the final cleavage site and at least one nucleotide of flanking downstream DNA, and comprising a C-cluster region, is designated AF region.

[0273] Efficacy of right border alternatives as sites for secondary cleavage was studied by testing pSIM108 and 843B. The vectors contained St01 and Ms01, respectively, as right border alternative. The downstream region of pSIM108, shown in SEQ ID 125, contained (1) AT-rich (62%) DNA (SEQ ID NO: 184), comprising a putative binding site for integration host factor with the consensus 5'-[A/T]-ATCAANNNTT-[A/G] (SEQ ID NO: 129), and derived from the terminator of the potato ubiquitin-3 gene (Garbarino *et al.*, 1994) containing a UL domain, and (2) a second copy of St01 associated with plasmid backbone DNA comprising five C-clusters (SEQ ID NO: 125).

[0274] Similarly, the DNA region intended for secondary cleavage in pSIM843B (SEQ ID NO: 126) contained a second copy of Ms01 preceded by an AT-rich (87%) alfalfa DNA fragment, and followed by downstream C-clusters (Figure 3B). Vector pSIM401, which contained the extended left border region of pTiC58,

was used as control. PCR genotyping demonstrated that both pSIM108 and 843B yielded even higher frequency of backbone-free transformation events (41.1 and 33.9%) than obtained with the control (26.0%), thus indicating that right border alternatives can be used to replace left borders.

**[0275]** A modification of pSIM843B that both eliminated the UL domain and altered the spacing of C-clusters yielded a UF region that lowered the frequency of desired 'T' transformation events for the resulting vector pSIM849 (SEQ ID NO: 127) to 10.2% (Figure 3B). This reduced frequency was associated with an about two-fold increased transfer of DNAs that are still attached to their vector backbones, indicating that the modifications of flanking DNA interfered with effective secondary cleavage at the second Ms01 copy. Similar alterations of the UF region of pSIM108 resulted in a sequence (SEQ ID NO: 127) that reduced transformation efficacy about four-fold (Figure 3B).

**[0276]** Sequences of UF regions of pSIM108, pSIM843B and pSIM781 are depicted in SEQ ID NOs: 184-186.

**[0277]** Collectively, this data demonstrate that right border alternatives can be used to replace left borders if associated with upstream UL domain and downstream C-clusters. Even small changes in this organization were found to have a profound effect on the frequency of backbone-free plant transformation. Replacement of the internal nptII gene expression cassette of pSIM843B by alfalfa DNA would make it possible to produce intragenic alfalfa plants.

**[0278]** The full region of pSIM843B for efficient initial cleavage comprises UI region, Ms01, and DI region, and is shown in SEQ ID NO: 131. The full region of pSIM843B for efficient final cleavage comprises UF region, Ms01, and AF region, and is shown in SEQ ID NO: 132.

## **EXAMPLE 5**

### **Cleavage sites from eukaryotes other than plants**

[0279] In addition to plant-derived cleavage sites, such elements can also be identified in, for instance, fungi and mammals. See, for instance, SEQ ID NOs. 173-182. Several of these species have already been shown to be accessible to *Agrobacterium*-mediated transformation (Kunik *et al.*, Proc Natl Acad Sci USA 98: 1871-1876, 2001; Casas-Flores *et al.*, Methods Mol Biol 267: 315-325, 2004). Thus, the new elements may be used to extend the concept of all-native DNA transformation (Rommens, Trends Plant Sci 9: 457-464, 2004) to eukaryotes other than plants.

[0280] The present invention also contemplates methods for identifying other polynucleotide sequences that can be used in place of the specific sequences described herein. For instance, it is possible to identify polynucleotide sequences that can replace cleavage sites, as well as polynucleotide sequences that can replace the regions that are upstream and downstream of the cleavage sites.

[0281] A sequence that is upstream of the cleavage site is removed and a different polynucleotide is inserted. The sequence of the different polynucleotide may or may not be known. With all the other elements in place to facilitate appropriate transformation in the transfer cassette and plasmid, the insertion is tested to determine if the different polynucleotide facilitates transformation. The assay makes it possible to identify alternative polynucleotide sequences that can be used to build an effective transfer cassette. Accordingly, one may transform a plant with a transformation plasmid in which a candidate polynucleotide sequence has been inserted in place of one of the established sequences described herein. Successful plant transformation is monitored and the inserted DNA further characterized.

[0282] Hence, various elements described herein can be replaced with candidate DNA sequences to test whether those candidate DNA sequences are useful as alternative functional elements for successful plant transformation (see Figure 4).

## **EXAMPLE 6**

### **Alternative final cleavage sites**

[0283] Instead of using Left Borders or cleavage sites that conform to SEQ ID NO: 84, it is also possible to use the sequence depicted in SEQ ID NO: 133, or a fragment thereof, as a final cleavage site. Actual single stranded DNA cleavage often occurs between the 14th and 15th nucleotide. However, it is also possible that transferred DNA comprises either more or less than 14 nucleotides of SEQ ID NO: 133.

[0284] Binary vectors that contain (1) either a Right Border or initial cleavage site upstream from a polynucleotide and (2) SEQ ID NO: 133 as final cleavage site. downstream from this polynucleotide can be used to efficiently transfer the polynucleotide, often still flanked by about three base pairs of the 3'-terminus of the Right Border or initial cleavage site and about 14 base pairs (CCCGAAAAACGGGA) (SEQ ID NO: 191) of the alternative final cleavage site. Together, the transferred sequence can be designated "transfer DNA."

[0285] Given the size of plant genomes, only plant species with very small genomes may not contain the 14 base pair sequence of SEQ ID NO: 133 that is transferred, as part of the transfer DNA, from the binary vector to the plant cell. For instance, Arabidopsis contains ACCGAAAAACGGGA (SEQ ID NO: 192) instead of SEQ ID NO: 191. The mismatch at position "1" would represent a single point mutation, which is acceptable for all-native DNA transformation because point mutations occur spontaneously in plant genomes. Furthermore, it is possible to use parts of SEQ ID NO: 133 as alternative final cleavage site. For instance, SEQ ID NO: 134 to SEQ ID NO: 137, or functional fragments thereof, may be used.

[0286] Interestingly, the fidelity of DNA transfer with vectors that contain SEQ ID NO: 133 as an alternative final cleavage site is higher than similar vectors that contain a conventional Left Border region instead. Table 1 shows the genotypes of tobacco plants derived from an infection with *Agrobacterium* LBA4404 carrying

specific plasmids. Plasmid pSIM794 contains an expression cassette for the neomycin phosphotransferase (nptII) gene inserted between a conventional Right Border and SEQ ID NO: 133. Plasmid pSIM795 contains the same plasmid except that SEQ ID NO: 133 is positioned in the inverse complementary (antisense) position. The benchmark vector contains conventional Left and Right Borders (pSIM109), and the previously discussed pSIM1008 was used as control vector. See Table 1. The use of alternative final cleavage site makes it unnecessary to use associated UF and AF regions.

### **EXAMPLE 7**

#### **T-DNA-delivered transposon-based transformation**

[0287] Instead of using either borders or cleavage sites as sequences that define the ends of the polynucleotide intended for plant transformation, it is also possible to use the termini of plant transposable elements. Until now, transposon-based transformation systems were based on either protoplast transformation (Houba-Herin *et al.*, 1994) or geminivirus vectors (Laufs *et al.*, 1990; Shen and Hohn, 1992; Wirtz *et al.*, 1997; Shen *et al.*, 1998). Both these systems are extremely inefficient, and have not been pursued for commercial purposes. In contrast to conventional transposon-based transformation, we employ the transfer DNA to deliver the transposable element into the plant nucleus. Excision from the transferred DNA, followed by integration into the plant genome, results in effective plant transformation.

[0288] The plasmid used to demonstrate the efficacy of T-DNA-delivered transposon-based (TDTB) transformation contains the conventional Left and Right Border regions of *Agrobacterium*. Between these border regions, the following elements were inserted: (1) an expression cassette for the transposase gene of the maize transposable element Ac (SEQ ID NO: 138), (2) a non-autonomous transposable element designated 'transposon' comprising an expression cassette for the neomycin phosphotransferase (nptII) gene positioned between the 5' and 3' ends of the Ac element depicted in SEQ ID NOs: 139 and 140, and (3) an expression

cassette for the cytosine deaminase (*codA*) gene. See Figure 5. Transgenic plants were created as follows:

[0289] Tobacco explants (4,500) were infected with an *Agrobacterium* strain carrying the plasmid described above. The infected explants were co-cultivated and transferred to medium containing kanamycin (100 mg/L) to select for plant cells expressing the *nptII* gene. After one month, shoots were transferred to fresh media that also contained the non-toxic 5-fluorocytosine (5-FC). Stable integration of the entire transfer DNA would result in constant expression of the *codA* gene and subsequent conversion of 5-FC into toxic 5-fluorouracil (5-FU). Thus, only transformed shoots that did not express the *codA* gene would be expected to survive this selection step. A total of 141 shoots were harvested after selection periods of 10, 20, 30 and 45 days on 5-FC, and PCR analyzed to determine whether the shoots carried integrated T-DNAs still harboring the transposon at its original resident position or whether they carried the transposon integrated into plant DNA (Table 2). The following primer sets were used for this purpose:

[0290] (1) indicative for the presence of the transposon: (NPTII)

[0291] (SEQ ID NO: 141): AGGAAGGAATTCCCCCGGATCAGC

[0292] (SEQ ID NO: 142): AGGAGCAAGGTGAGATGACAGG

[0293] (2) indicative for the presence of the T-DNA: (CodA)

[0294] (SEQ ID NO: 143): GAATCAGCTAATCAGGGAGTGTG

[0295] (SEQ ID NO: 144): GCCATGCGCGTTGTTTCACATCG

[0296] (3) indicative for the presence of a T-DNA carrying a non-excised transposon (the "full donor site"): 637 bp for F1-R1; 848 bp for F1-R2)

[0297] P1A (SEQ ID NO: 145): GCATGCTAAGTGATCCAGATG (F1)

[0298] P1B (SEQ ID NO: 146): CTGCAGTCATCCCGAATTAG (R1)

[0299] P1A and P1B amplify the upstream “full donor site”, representing the junction between T-DNA and 5'-transposon end, (651 bp) and

[0300] P2A (SEQ ID NO: 147): GGAATTCGCGTAGACTTATATGGC  
(F2)

[0301] P2B (SEQ ID NO: 148):TGATGACCAAAAATCTTGTCATCCTC  
(R2)

[0302] P2A and P2B amplify the downstream “full donor site”, representing the junction between 3'-transposon and T-DNA.

[0303] (4) indicative for the presence of a T-DNA that lost the transposon due to excision (the “empty donor site”, 656 bp):

[0304] P3A (SEQ ID NO: 149): GCATGCTAAGTGATCCAGATG (F1)

[0305] P3B (SEQ ID NO: 150): TGATGACCAAAAATCTTGTCATCCTC  
(R2)

[0306] Twenty-four plants contained both a full and empty donor site, indicating that the transposon in these plants excised from a stably integrated T-DNA. These plants were not considered for further studies.

[0307] In contrast, thirteen contained the transposon and lacked a full donor site. DNA gel blot analysis of these plants demonstrated that eleven of them contained the nptII gene and lacked the codA gene, indicating that they did not contain a stably integrated T-DNA. As shown in Table 2, most of these eleven plants were obtained from the 30-day 5-FC selection experiment.

[0308] Eight of eleven plants that lacked any T-DNA or backbone DNA sequences contained a single transposon insert. Because tobacco transformation results, on average, in the integration of two T-DNAs most of which still linked to backbone DNA, the frequency of single-copy and backbone-free transgenic plants is higher for TDTB transformation.

[0309] To confirm the integration of excised transposons into plant genomes, we determined the sequence of transposon-plant DNA junctions. Upstream junctions were isolated by (i) digesting DNA of the transgenic lines, (ii) circularizing this DNA using T4 DNA ligase, (iii) employing the resulting DNAs as template for a first PCR using the primer pair TR1 and TD1 (SEQ ID NOs: 151 and 152), and (iv) using the resulting template with the primer pair TR2 and TD2 for a second PCR (SEQ ID NOs: 153 and 154).

[0310] Similarly, the primer pair RTR1 and RTD1 (SEQ ID NOs: 155 and 156) was used for first round amplifications of the downstream junction, and the resulting template was used with RTR2 and RTD2 for second round amplifications (SEQ ID NOs: 157 and 158).

[0311] Sequence analysis of the junction fragments confirmed that the transposon had in each case excised from the non-integrating T-DNA and integrated into a unique position in plant DNA. As expected, the integrated transposons were flanked by eight-base pair direct repeats, created by duplication of the eight-base pair integration site.

[0312] Instead of T-DNAs, it is also possible to use plasmids that can be maintained in *Agrobacterium* and/or *Rhizobium* and contain at least one cleavage site. Instead of the transposon ends employed here, it is also possible to use the termini of other transposable elements that are functional in plants.

## **EXAMPLE 8**

### **Enhanced fidelity of DNA transfer with plasmids carrying the virC operon**

[0313] To study whether virC genes influence the frequency and fidelity of the T-DNA transfer, we isolated the entire virC operon (SEQ ID NO. 167) from *Agrobacterium* via PCR approach using virC operon specific primers 5' GTTTAAACAGCTTCCTCCATAGAAGACGG 3' (SEQ ID NO. 168) and 5' TTAATTAATCGTACGGGGGTGTGATGG 3' (SEQ ID NO. 169). The PCR

amplified virC operon was cloned into PmeI-PacI sites of the pSIM1008 plasmid DNA backbone that contains Le01 as initial cleavage site and the conventional Left Border of pTiC58 for secondary cleavage. Stably transgenic tobacco plants produced with the resulting plasmid pSIM1026 were analyzed, and the data were compared with those obtained with plasmid pSIM1008. Table 3 shows that the presence of the virC operon increased the frequency of backbone-free transformation more than two-fold.

## TABLES

**Table 1**

| Plasmid          | Backbone-free transformation with transfer DNA | Transformation with transfer DNA still linked to backbone |
|------------------|------------------------------------------------|-----------------------------------------------------------|
| Benchmark vector | 39%                                            | 61%                                                       |
| Control vector   | 26%                                            | 74%                                                       |
| pSIM794          | 55%                                            | 45%                                                       |
| pSIM795          | 44%                                            | 56%                                                       |

**Table 2.**

| Treatment       | Number of transformed plants | Only carrying the transposon in plant DNA | Carrying at least one T-DNA comprising the transposon at its original position |
|-----------------|------------------------------|-------------------------------------------|--------------------------------------------------------------------------------|
| 10 days on 5-FC | 39                           | 0 (0%)                                    | 4 (10%)                                                                        |
| 20 days on 5-FC | 51                           | 3 (6%)                                    | 12 (24%)                                                                       |
| 30 days on 5-FC | 35                           | 9 (26%)                                   | 5 (14%)                                                                        |
| 45 days on 5-FC | 16                           | 1 (6%)                                    | 3 (19%)                                                                        |
| Total           | 141                          | 13 (9%)                                   | 24 (17%)                                                                       |

**Table 3.** Genotypes of transgenic tobacco plants produced with pSIM1026 and pSIM1008.

| Plasmid  | Integration of sequences between Leo1 and Left Border only (1) | Integration of sequences comprising both the actual transfer DNA and plasmid backbone sequences (2) | Integration of plasmid backbone sequences only (3) |
|----------|----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------|
| pSIM1008 | 16.9 ± 1.7                                                     | 67.7 ± 5.3                                                                                          | 21.7 ± 3.7                                         |
| pSIM1026 | 39.5 ± 4.1                                                     | 51.5 ± 0.8                                                                                          | 9.2 ± 3.4                                          |

(1) Visualized using primers

5' TGCTCCTGCCGAGAAAGTAT 3' (SEQ ID NO: 170) and

5' AGCCAACGCTATGTCCTGAT 3' (SEQ ID NO: 171)

(2) Visualized using primers SEQ ID 170 and SEQ ID 171, SEQ ID 172 and SEQ ID 183

(3) Visualized using primers

5' GAATCAGCTAATCAGGGAG 3' (SEQ ID NO: 172) and

5' GCCATGCGCGTTGTTTCACATCG 3' (SEQ ID NO: 183).

## SEQUENCE TABLE

| SEQ ID NO: | NAME (if any)                 | SEQUENCE                                                                                                                |
|------------|-------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| 1          | Rb01                          | GTTTACCCGCCAATATATCCTGTCA                                                                                               |
| 2          | Rb02                          | AATTACAACGGTATATATCCTGCCA                                                                                               |
| 3          | Rb03                          | CATGACAGGAACATATATCCTGTCA                                                                                               |
| 4          | Rb04                          | AATTACAACGGTATATATCCTGTCA                                                                                               |
| 5          | Rb05                          | CCTGACCACAAGATATATCCTGTCA                                                                                               |
| 6          | Rb06                          | CTAGACAAGGGGATATATCCTGTCA                                                                                               |
| 7          | Rb07                          | CATTACTTTAGAATATATCCTGTCA                                                                                               |
| 8          | Sy01                          | CTTTACACAACAATATATCCTGTCA                                                                                               |
| 9          | Sy02                          | GTCTACACAACAATATATCCTGTCA                                                                                               |
| 10         | Sy03                          | GTTTAAACAACAATATATCCTGTCA                                                                                               |
| 11         | Sy04                          | GTTTACACAACAAGATATCCTGTCA                                                                                               |
| 12         | Sy05                          | GTTTACTCAACAATATATCCTGTCA                                                                                               |
| 13         | Sy06                          | GTTAACACAACAATATATCCTGTCA                                                                                               |
| 14         | Sy07                          | GTTTACACAACACTATATCCTGTCA                                                                                               |
| 15         | Sy08                          | GTTTACACAACAATATATCCTGGCA                                                                                               |
| 16         | Sy09                          | GTTTACACAACAATAAATCCTGTCA                                                                                               |
| 17         | Sy10                          | GTTTACACAACAATATATCCTGTCA                                                                                               |
| 18         | Sy11                          | GTTTACACAACAATATATGCTGTCA                                                                                               |
| 19         | Sy12                          | GTTTACACAACAATATATCGTGTCA                                                                                               |
| 20         | Sy13                          | GTTTACACAACAATATATCCAGTCA                                                                                               |
| 21         | extended UI region of pSIM551 | ACGAACGGATAAACCTTTTCACGCCCTTTTAAATATCCGTT<br>ATTCTAATAAACGCTCTTTTCTCTTAGAGATCTCAAACAAA<br>CATACACAGCGACTTATTCACAACACTAG |
| 22         | DI region of pSIM551          | GGGCCCCGGTACCCGGGGATCAATTCCCGATCTAGTAACATA<br>GATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTA<br>TATTTTGTTTTCTATCGCGTATTAAAT  |
| 23         | potato St01                   | GTTTACATCGGTATATATCCTGCCA                                                                                               |
| 24         | primer                        | YGR CAG GAT ATA TNN NNN KGT AAA C                                                                                       |
| 25         | anchor primer                 | GAC CAC ACC CGT CCT GTG                                                                                                 |
| 26         | Arabidopsis At01              | GTTGACATCACGATATATCCTGTCA                                                                                               |
| 27         | CON1                          | [A/C/G] [A/T] [A/T] [G/T] AC [A/C/T] N [C/G/T] [A/C/G] [A/C/G] [A/C/G] ATATATCCTG [C/T] CA                              |
| 28         | tomato Le01                   | CATTACCAACAAATATATCCTGGCC                                                                                               |
| 29         | tomato Le02                   | CTCTACCTCTGAATATATCCTGCGG                                                                                               |
| 30         | tomato Le03                   | GCATACCTCTGAATATATCCTGCGG                                                                                               |
| 31         | potato St03                   | GTTTACCTTAGCATATATCCTGCAT                                                                                               |
| 32         | alfalfa Ms01                  | GTATACCTCTGTATACATCCTGCCG                                                                                               |
| 33         | barley Hv01                   | ATATACCAAATGATACATCCTGCCC                                                                                               |
| 34         | rice Os01                     | ACTTACTCAAGGATATATCCTGGCT                                                                                               |
| 35         | rice Os0                      | CACTACAAAAAATATATCCTGCAT                                                                                                |
| 36         | rice Os03                     | ATGTACGTATATATATATCCTGTGT                                                                                               |
| 37         | wheat Ta01                    | ATATACGGAGCAATATATCCTGTCC                                                                                               |
| 38         | Soybean Gm01                  | AAAAACTGTTTTATATATCCTGTCA                                                                                               |

|    |                        |                            |
|----|------------------------|----------------------------|
| 39 | Soybean Gm02           | AATAACTCTGAAATATATCCTGTGT  |
| 40 | Potato St04,           | ACCTACCCCAAATATATCCTGCCT   |
| 41 | tomato Le04            | GGAAACTGTCTAATATATCCTGTGA  |
| 42 | tomato Le05            | ACCTACCCCAAATATATCCTGCCC   |
| 43 | tomato Le06            | GTTTAGACTTGTATATATCCTGCCC  |
| 44 | tomato Le07            | TCTTAGAACTCAATATATCCTGTAC  |
| 45 | tomato Le08            | CGTTAACACTGTATATATCCTGTAA  |
| 46 | tomato Le09            | GAATTATTTTGCATATATCCTGTAA  |
| 47 | tomato Le10            | TTGTTCCCTGGCCATATATCCTGCCA |
| 48 | tomato Le11            | GGTACCATGTAGATATATCCTGCTT  |
| 49 | M. truncatula<br>Mt01  | GTATACCTCTGTATACCTCCTGCCG  |
| 50 | maize Zm01,            | GCGTACGCATTTATATATCCTGTGG  |
| 51 | Zm01-derived<br>Zm01M1 | GCTTACGCATTTATATATCCTGTGG  |
| 52 | Brassica rapa<br>Br01  | CCCTACTGTATAATAAATCCTCTAG  |
| 53 | tomato Le10            | TTGTTCCCTGGCCATATATCCTGCCA |
| 54 | tomato Le1             | GGTACCATGTAGATATATCCTGCTT  |
| 55 | tomato Le12            | GTTCCGGTTGACATATATCCTGACA  |
| 56 | tomato Le13            | CACTACCGCCTCATAGTTCCTGCCA  |
| 57 | soybean Gm01           | TAAAGCAACACCATATATCCTGACA  |
| 58 | M. truncatula<br>Mt02  | GATTAGACAAATATTTATCCTGCCA  |
| 59 | rice Os04              | CTCTACTACCCGAGATGTCCTGGCA  |
| 60 | potato St05*           | GTTTGACACGACATATATACTGCAA  |
| 61 | potato St06*           | GTTTACCGTGGCACTTATGTGATGA  |
| 62 | potato St07            | CATTACCAACTATTATATCCTGGCC  |
| 63 | tomato Le14            | GTTTACTTGAAGATATCACCTATGT  |
| 64 | tomato Le15            | TTCCATACGAAGAGAAGTCCTGTCA  |
| 65 | tomato Le16            | TTCTAGCTGCAAATATATCCTGGCT  |
| 66 | tomato Le17            | GTTGACATGGATGAATATCCTGTCA  |
| 67 | tomato Le18            | G TTCAGCTTAGCATATATCCTGCAT |
| 68 | tomato Le19            | TTCCAGAAGTAGATATATCCTGTTG  |
| 69 | tomato Le20            | TGATTGCATCAAATATATCCTGCCA  |
| 70 | tomato Le21            | ATCCCCACCCATTTATATCCTGCCA  |
| 71 | tomato Le22            | CATCCCCACCATTTATATCCTGCCA  |
| 72 | tomato Le23            | GTCAGGAAGTGAATATATCCTGACA  |
| 73 | tomato Le24            | GTTTAAACCAATATATATCCTGATT  |
| 74 | tomato Le25            | AGTTATAAACTTATATATCCTGTTG  |
| 75 | tomato Le26            | CTAAAGTTGTACATAAATCCTGTCT  |
| 76 | tomato Le27            | TTCTACACAAAGACAAATCCTGGCG  |
| 77 | tomato Le28            | ATTAACAACGTTAGAAGTCCTGGCG  |
| 78 | M. truncatula<br>Mt03  | CATGACCCTGCAATATGTCCTGTGG  |
| 79 | maize Zm02             | AACTTAAAGATAAGAAGTCCTGCCA  |
| 80 | oat As01               | CTGTACAATAGGACAAATCCTGTCTG |
| 81 | potato St08*           | TTTTACCCGTGATATATCCCAGCC   |
| 82 | tomato Le29            | GATTGCATCAAATATATCCTGCCA   |

|     |                                    |                                                                                                                        |
|-----|------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| 83  | tomato Le30                        | AAGTACCGATGATATATCCTGCGT                                                                                               |
| 84  | CON2                               | [A/C/G] - [A/C/T] - [A/C/T] - [G/T] -A- [C/G] -<br>NNNNNN-A- [G/T] -A- [A/C/T] - [A/G] -TCCTG-<br>[C/G/T] - [A/C/G] -N |
| 85  | Ca01                               | CATTACCAACAAATATATCCTGGCC                                                                                              |
| 86  | St02                               | CATTACCAACAAATATATCCTGGCC                                                                                              |
| 87  | UI region                          | CTTAGAGATCTCAAACAAACATACACAGCGACTTATTCACA<br>ACTAGTC                                                                   |
| 88  | overdrive                          | CAAACAAACATACACAGCGACTTA                                                                                               |
| 89  | UI-derived                         | TTAGAGATCTCAAACAAACATACACAGCGACTTATTCACAA<br>CTAGTAC                                                                   |
| 90  | UI-derived                         | AGAGATCTCAAACAAACATACACAGCGACTTATTCACAAC<br>AGTCAAC                                                                    |
| 91  | UI-like from<br>Agrobacterium      | AGAAACAATCAAACAAACATACACAGCGACTTATTCACACG<br>AGCTCAA                                                                   |
| 92  | UI-like from<br>Agrobacterium      | GCCCTTTTAAATATCCGATTATTCTAATAAACGCTCTTTTC<br>TCTTAGG                                                                   |
| 93  | UI-like from<br>Agrobacterium      | TGACGAACTGACGAACTGACGAACTGACGAACTGACGAACT<br>GACGAAC                                                                   |
| 94  | UI-like from<br>Agrobacterium      | TAACAATTGAACAATTGAACAATTGAACAATTGAACAATTG<br>AACAAAC                                                                   |
| 95  | UI-like from<br>Agrobacterium      | TAGACATTGCACATCCAAAGGCAGGCACGTACAAACGAATT<br>TATTTAG                                                                   |
| 96  | UI-like from<br>Agrobacterium      | GAAGGCACGAAGGCACGAAGGCACGAAGGCACGAAGGCACG<br>AAGGCAC                                                                   |
| 97  | UI-like from<br>Agrobacterium      | TCATCACCGCCGTCCTAAACAAACATACCTCCACACAAATT<br>TATCTAC                                                                   |
| 98  | UI-like from<br>Agrobacterium      | AGATCTCAAACAAACATACACAGCGACTTATTCACAAC<br>TACCAAC                                                                      |
| 99  | UI region                          | TGACGAACTGACGAACTGACGAACTGACGAACTGACGAACT<br>ACCAAC                                                                    |
| 100 | UI-derived                         | CTGACGAACTGACGAACTGACGAACTGACGAACTGACGAACT<br>TACCAAC                                                                  |
| 101 | UI-like                            | TGTCTTTATCTCTTGTTGCCAAAAGTGCTCTCGAGTCGAGT<br>CACCAAC                                                                   |
| 102 | Downstream<br>from right<br>border | GTCAGCATCATCACACCAAAAGTTAGGCCCGAATAGTTTGA<br>AATTAGAAA                                                                 |
| 103 | Downstream<br>from right<br>border | AACACTGATAGTTTAAACCGAAGGCGGGAAACGACAATCTG<br>ATCATGAGCGG                                                               |
| 104 | Downstream<br>from right<br>border | AATAACAATCTCATGTTAGGTAATAATATCACCCAATCAAC<br>GCGGCCA                                                                   |
| 105 | Downstream<br>from right<br>border | GCACTAATATAAGAAATGTCCTGTCAGCACTAATATAAGAA<br>ATGTC                                                                     |
| 106 | Downstream<br>from right<br>border | AACCTATTTCGTTAATAGGGACGTCGTACCTACTTCCCTTCC<br>AGCGCAGCA                                                                |

|     |                           |                                                                                                                                              |
|-----|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| 107 | DR domain                 | [A/C/T] - [A/C] - [A/C/T] - [A/G/T] - [A/T] - T -<br>[A/C] - G - [G/T] - [G/T]                                                               |
| 108 | DI region<br>from potato  | GAGGTATAGAGGCATGACTGGCATGATCACTAAATTGATGC<br>CCACAGAGGAGACTTATAACCT                                                                          |
| 109 | DI region<br>from potato  | GGGCCCCGGTACCCGGGGATCAATTCCCGATCTAGTAACATA<br>GATGAC                                                                                         |
| 110 | DI-region                 | GGGCCCCGGTTCCCGGGGATCAATTGGGCCCCGGTACCCGGGG<br>ATCAATTCCCGATCTAGTAACATAGATGAC                                                                |
| 111 | DI-region                 | GGGCCCCGGTACCCGGAGGAGACTCCGATCTACGGCGCCAAA<br>TTCAAG                                                                                         |
| 112 | DI-region                 | CTGAGGACATTCAGAAGATTGGTTATATCCTCTTTCAAGAC<br>GCTAAGCAA                                                                                       |
| 113 | DI-region<br>from potato  | GAGGTATAGAGGCATGTCTGGCGTGATCACTAAATTGATGC<br>CCGCAGAGGGGACTTATAACAT                                                                          |
| 114 | DI-region<br>from alfalfa | GGGCCCCGGTACCCGTTAGGGCTAGCCCGAAAGGGCCGCGG<br>GCAGCCC                                                                                         |
| 115 | repeat                    | CCCG                                                                                                                                         |
| 116 | AF region                 | TCTCCATATTGACCATCATACTCATTGCTGATCCATGTAGA<br>TTTCCCGGACATGAAGCCATTTACAATTGAATATATCCTGC<br>CGCCGCTGCCGCTTTGCACCC                              |
| 117 | AF region                 | TGAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAG<br>TTGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGC<br>CACCAGCCAGCCAACAGCTCCCCGACC                        |
| 118 | AF region                 | ATCTGGTAATATAGCAAAAACGTGCTCAAAAATCGCTTCAA<br>AGCTCTTGTACTTAGCTCGTTTACACCACAATATATCCTGC<br>CACCCC                                             |
| 119 | AF region                 | TACATTTTATATTCGATAAAGCATGCGTTAAAACGACTTCG<br>CATGTCCATATCTAATCTGTTTACATCACAATATATCCTGC<br>CACCCAAGGAGCGACGCCTTCTGGCC                         |
| 120 | UL domain                 | A [C/T] T [C/G] A [A/T] T [G/T] [C/T]<br>[G/T] [C/G] A [C/T] [C/T] [A/T]                                                                     |
| 121 | left border<br>consensus  | [A/G] TTTACA [A/C/T] [A/C/T] [A/C/T]<br>[C/G] AATATATCCTGCC [A/G]                                                                            |
| 122 |                           | CCN(1-11)CCN(1-11)CCN(1-11)CC                                                                                                                |
| 123 | AF region                 | AAATCTGATTGATAAAGGATCGATCCTCTAGAGTCGACCTG<br>CAGTACTTACGTACAATTGTTTACACCACAATATATCCTGC<br>CACCGGATATATTGCCTAGGAGCCAGCCAACAGCTCCCCGA<br>CC    |
| 124 | AF region                 | AAATCTGATTGATAAAGGATCGATCCTCTAGAGTCGACCTG<br>CAGTACTTACGTACAATTGTTTACACCACAATATATCCTGC<br>CACCCCTAGGAGCCAGCCAACAGCTCCCCGACC                  |
| 125 | AF region                 | GTTTACACCACAATATATCCTGCCACCCCTAGGAGCCAGCC<br>AACAGCTCCCCGACC                                                                                 |
| 126 | AF region                 | GTAAAAAATAAAAGTGAAAATTCAATGAATTAACACAAATA<br>TAAATGTAATATAAAATTGTATACCTCTGTATACATCCTGC<br>CGCCAAGCTTCCAGCCACCTAGGAGCCAGCCAACAGCTCCC<br>CGACC |

|     |                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----|------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 127 | AF region                                                        | AATGGAGGTAAGTGTTTCTGCTCAGTGCTGATAGATGTAAA<br>TATCTCTGTTATGAAGCCGTATACCTCTGTATACATCCTGC<br>CGGGATGTATACCTTAGGCCAGCCAGCCAACAGCTCCCCGA<br>CC                                                                                                                                                                                                                                                         |
| 128 | AF region                                                        | TGTTGAAGGCTTGGATGTGATTAAGAAGGCCGAGGCTGTTG<br>GATCTAGTTCTTGAAGTTCATTACCAACAAATATATCCTGG<br>CCCCCTAGGAGCCAGCCAACAGCTCCCCGACC                                                                                                                                                                                                                                                                        |
| 129 | IHF site                                                         | [A/T] ATCAANNNN [A/G]                                                                                                                                                                                                                                                                                                                                                                             |
| 130 | search motif                                                     | CAGGATATATNNNNNNNGTA                                                                                                                                                                                                                                                                                                                                                                              |
| 131 | extended DNA<br>region of<br>pSIM843B for<br>initial<br>cleavage | GGCTGCACTGAACGTCAGAAGCCGACTGCACTATAGCAGCG<br>GAGGGGTTGGATCAAAGTACTTTGATCCCGAGGGGAACCT<br>GTGGTTGGCATGCACATACAAATGGACGAACGGATAAACCT<br>TTTCACGCCCTTTTAAATATCCGATTATTCTAATAAACGCT<br>CTTTTCTCTTAGAGATCTCAAACAAACATACACAGCGACTT<br>ATTCACAAC TAGTGTATACCTCTGTATACATCCTGCCGGGG<br>CCCGGTACCCGTTAGGGCTAGCCCGAAAGGGCCGCGGGCAG<br>CCCGTTAGCCCGCATAACTGCAGCCCGGG                                          |
| 132 | extended DNA<br>region of<br>pSIM843B for<br>final<br>cleavage   | CAGTACTTACGTACATAACAAAAAAATTCTATAAATTAT<br>ATATATTTTTCAAATAATTCTTTACACAGTTGATTATCAAA<br>GTAAAAAATAAAAGTGAAAATTCAATGAATTAACACAAATA<br>TAAATGTAATATAAAATTGTATACCTCTGTATACATCCTGC<br>CGCCAAGCTTCCAGCCACCTAGGAGCCAGCCAACAGCTCCC<br>CGACCGGCAGCTCGGCACAAAATCACC ACTCGATACAGGCA<br>GCCCATCAGTCCGGGACGGCGTCAGCGGGAGAGCCGTTGTA<br>AGGCGGCAGACTTTGCTCATGTTACCGATGCTATTTCGAAG<br>AACGGCAACTAAGCTGCCGGGTTTGA |
| 133 | alternative<br>final<br>cleavage site                            | CCCGAAAAACGGGACAGGATGTGCAATTGTAATACCGTCAC<br>ACGCGACGCTATTACAATTGCCATCTGGTCAGGGCTTCGCC<br>CCGACACCCC                                                                                                                                                                                                                                                                                              |
| 134 | alternative<br>final<br>cleavage site                            | CCCGAAAAACGGGACAGGATGTGCAATTGTAATACCGTCAC<br>ACGCGACGCTATTACAATTGCCA                                                                                                                                                                                                                                                                                                                              |
| 135 | alternative<br>final<br>cleavage site                            | CCCGAAAAACGGGACAGGATGTGCAATTGTAATACCGTCAC<br>ACGCGACGCTA                                                                                                                                                                                                                                                                                                                                          |
| 136 | alternative<br>final<br>cleavage site                            | AAAACGGGACAGGATGTGCAATTGTAATACCGTCACACGCG<br>ACGCTATTACAATTGCCATCTGGTCAGGGCTTCGCCCCGAC<br>ACCC                                                                                                                                                                                                                                                                                                    |
| 137 | alternative<br>final<br>cleavage site                            | ACCGAAAAACGGGACAGGATGTGCAATTGTAATACCGTCAC<br>ACGCGACGCTATTACAATTGCCATCTGGTCAGGGCTTCGCC<br>CCGACACCCC                                                                                                                                                                                                                                                                                              |

|     |                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 138 | Ac<br>transposase<br>gene | ATGACGCCTCCGGTTGGAAATAATCCTCCCTCAGGCTCAGC<br>CATAAGATTGGCCAAGTTGATGTCTACCACAAGAGCGCCTT<br>CTACTCGCAAAACAAATTCGGTATTCTCTGCATATGCTCAA<br>GGTATATATTAGAAAAACAGTAGCAATAGCATTAGCATTAC<br>TAATTGGTTGTAGATTGGGAAGCATCATATTGACTGTAGAA<br>TAATACGAAAAATCTGTTTATAACAGGGTTGAAAAGAAAAG<br>CTGAAGCCTCTTCTAGTCGGATTTCAGAATGTACGTGCACGT<br>GCGCGTGGGCATGGATGTGGCCGCACATCACCATCATCATC<br>AACAGCTGAGGCCGAGAGGCATTTTATTTCAGAGTGTAAGCA<br>GTAGTAATGCAAATGGTACAGCTACAGATCCGAGTCAAGAT<br>GATATGGCTATTGTTTCATGAACCACAACCACAACCACAACC<br>ACAACCAGAACCACAACCACAGCCACAACCTGAACCCGAAG<br>AGAAGCACCACAGAAGAGGGCAAAGAAGTGCACATCGGAT<br>GTATGGCAGCATTTTCACCAAGAAGGAAATTGAAGTGGAGGT<br>CGATGGAAAGAAATACGTTTCAGGTATGGGGACATTGCAACT<br>TTCCTAATTGCAAGGCTAAGTATAGGGCTGAGGGTTCATCAT<br>GGAACAAGCGGATTTTCGAAATCACTTGAGAACATCACATAG<br>TTTAGTTAAAGGTCAGTTGTGTCTAAAAAGTGAAAAGGATC<br>ATGGCAAAGACATAAATCTCATTGAGCCTTATAAGTACGAT<br>GAAGTGGTTAGCCTAAAGAAGCTTCATTGGAATAATCAT<br>GCATGAATATCCTTTCAATATTGTAGAACATGAGTACTTTG<br>TTGAGTTTGTAAAGTCTCTGCGCCCTCACTTTCCAATAAAG<br>TCCCGTGTCACTGCTAGAAAATATATCATGGATTTGTATTT<br>GGAAGAAAAAGAAAAGTTGTATGGAAAACATAAAGATGTTT<br>AGTCTCGCTTCAGTACAACCTATGGATATGTGGACATCTTG |
| 139 | 5' transposon<br>end      | CAGGGATGAAAGTAGGATGGGAAAATCCCGTACCGACCGTT<br>ATCGTATAACCGATTTTGTAGTTTATCCCGATCGATTTTC<br>GAACCCGAGGTAAAAAACGAAAACGGAACGGAAACGGGATA<br>TACAAAACGGTAAACGGAAACGGAAACGGTAGAGCTAGTTT<br>CCCGACCGTTTTCACCGGGATCCCGTTTTTAATCGGGATGAT<br>CCCGTTTCGTTACCGTATTTTCTAATTCGGGATGACTGCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 140 | 3' transposon<br>end      | GTAGACTTATATGGCTTCTTATGTTAGCCAAGAGCCCAAGA<br>CTTATCACTTATGTGCTACATTAAACTATGTGTGCTCCAGA<br>TTTATATGGATTTTATCTATGTTTAAATTAAGACTTGTGTTT<br>ACAATTTTTTATATTTGTTTTTAAGTTTTGAATATATGTTT<br>TCATGTGTGATTTTACCGAACAAAAATACCGGTTCCCGTCC<br>GATTTGACTTTAACCCGACCGGATCGTATCGGTTTTTCGAT<br>TACCGTATTTATCCCGTTCGTTTTTCGTTACCGGTATATCCC<br>GTTTTTCGTTTCCGTCCCGCAAGTTAAATATGAAAATGAAAA<br>CGGTAGAGGTATTTTACCGACCGTTACCGACCGTTTTTCATC<br>CCTA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 141 | NPTII primer              | AGGAAGGAATTCCCCCGGATCAGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 142 | NPTII primer              | AGGAGCAAGGTGAGATGACAGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 143 | codA primer               | GAATCAGCTAATCAGGGAGTGTG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 144 | codA primer               | GCCATGCGCGTTGTTTCACATCG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 145 | P1A primer                | GCATGCTAAGTGATCCAGATG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 146 | P1B primer                | CTGCAGTCATCCCGAATTAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 147 | P2A primer                | GGAATTCGCGTAGACTTATATGGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 148 | P2B primer                | TGATGACCAAAATCTTGTTCATCCTC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 149 | P3A primer                | GCATGCTAAGTGATCCAGATG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

|     |                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 150 | P3B primer                           | TGATGACCAAAATCTTGTCATCCTC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 151 | TR1 primer                           | ATCGGTTATACGATAACGGTCGGTACG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 152 | TD1 primer                           | ACGAAAACGGAACGGAAACGGGATATAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 153 | TR2 primer                           | GATTTTCCCATCCTACTTTCATCCCTG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 154 | TD2 primer                           | GTAGAGCTAGTTTCCCGACCGTTTCAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 155 | RTR1 primer                          | GCACATAAGTGATAAGTCTTGGGCTC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 156 | RTD1 primer                          | CGACCGGATCGTATCGGTTTTTCGATTAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 157 | RTR2 primer                          | CTAACATAAGAAGCCATATAAGTCTAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 158 | RTD2 primer                          | CGGTAGAGGTATTTTACCGACCGTTAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 159 | upstream<br>junction of<br>plant 1   | ATAGATAAGAGGAGTTTGTTACAAATTTCTACTCCACATTG<br>ATGAGAAATATACTAATGTTATCTCCCCTTCCCTCTATTAG<br>TAGATCTTACTCTATGTTAAACATGACAAGAAATAGAGAG<br>AGAACTCACACTTTCTTCCTCATCTGCTACTTCTGGTGCCG<br>AAGAAGTTTTACTCAAAGAGTCTAATTTAAGGCAACGAAGC<br>ATGTCCTTTTTGTCTCTTGCAAGTATTGCAAGAAGGCAGGAC<br>ACACTTTAAAGAAGTGTTATAAGTCATCCATTTTCCTCTGT<br>CTTCAATTTCTTAAAGACCAAAAGATCCAGTCTTTTGTGTC<br>CATGTTGATAATTTTACTCTAATACTCTTAGCTTCCA                                                                                                                                                                                                                                                                |
| 160 | downstream<br>junction of<br>plant 1 | AGCTTCCACATCCCAATTTGGTGATCATTCAGCACATAAAT<br>TTGCTCAGAAGCAATAGGAATATCTCATGTCTCTTCCTTCC<br>AAATAATCAATTCTCACCTAGGTTCAATAATGATGTTTCTT<br>TTAGAGAGATTTCTGACTATGATCATTTTGCAGGTTTAATT<br>AGTACATTTTTTTGTAGTTAATTATGTGTTTTTTCATGCATG<br>TTCATCATTTGCAATTAGGGGTAGATACTTGAATCTTTTACT<br>TGGGCCACTAGCCACATGACTCCATTTATGGTGTTTATAAG<br>CTATATCAGTGTATATCACATTGTATTTCCATATATCTCAG<br>GTGTACCATATATATCTGTGATTATGTGAAAGACCCCCCTA<br>ATTTGTGTCAAGACTGACAATGCTCTGTCAATCAGTGTAGC<br>AAAAATAAAAATAAAAATAAATCAAGGATTAGTACAACACC<br>ATCCAGGAACCTTTACTAGAAAATTAGTATAACCATATGAGT<br>CTTTTACAGTTTGGATCTATCATGGAGTAAAAGAATACATT<br>GCAGATTAGGATTATTCAAAATATGCCTTCTTGCAATCTAC<br>GTTGTGATCAACAGATATA |
| 161 | upstream<br>junction of<br>plant 2   | ATTCTCACCAAAAATTGAGCTGATTAGATAAAAAAAGATCA<br>ATTTGTTAAGACCAGCAGCAGCTCTTCAGTACCATTTTCATG<br>TCTTAACAGGACATATATATATATATATAGATATAGAGAGA<br>GAAAGTGGGCAAGACTTGATTTTATAGATCTAGAGAGAGA<br>AAAGGAGAGTTGGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

|     |                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-----|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 162 | downstream<br>junction of<br>plant 2 | GAGTTGGGGAGAAAAAGAAGGGATTTTACAGATCTAGAGA<br>GAGAAAGACTTGATTCTTCCCTATTTTCTCTTCACCATTTCC<br>TATGTTTTCTCTCCCTCTCTTTTCTCTTTCTTGATTTCTCT<br>ATAAATTTTCACTCATTAGTATATTCATCACTCTCAATTTA<br>CCTTTTATATAAAAAATAAAAAACAATAAAAAATTACTAAATAC<br>ATTTAATTTTAATTATAAATAGAAATTATTACACTATTGAT<br>TTTTTATTTGACTTATTTATTTATTTTAGTCTATTTCGAAAA<br>ATATGTCTTTTTTCGTTTTCTAATAACTCTTTCATTTTAGTC<br>TTTTCCATTTAATATTACAAAATTTAAAAAAATGCATTTTG<br>GTACCTTTTTTAAGATTACAAAATTTGAATATATTATTACT<br>TTATTAAATTACGCATTTAATCAAAACAAAACAATCAAAAT<br>GAAAGCATTTTGGTACCTTCTAGAATACGTATATTTAATTT<br>GAAATTACAAAATTTGAATATATTCTTTATTTTGTAAATT<br>ACGTATTTAGTCAAAACAGGACAATAAAAAAAAACGAAAGG<br>AGTAATTACTAATAACAATAACATTTTGACTAAAATTAAAT<br>TAAAGAAAAAAGGATTTTGGTACCTTCTAG                                                                                                                                                                                                                                                                                                  |
| 163 | upstream<br>junction of<br>plant 3   | TGTGATTTAGGAAGGTAAGATGACTTTGCAAGGATTGTCTT<br>CAAATGGCATAAATCTAACATTCAAATTAAGTCTATTTTT<br>AAACAATAAAAAATACATGAGATTTGCAATTTATAAGTCAAC<br>GTTGTCATATAACCCATTAGTTCCGGTTTTAAGGATATGAAT<br>AGAGGTTTGAAACGTGTTGCAAATGCTCTCAACTATGGACA<br>TAACCCAGTACCCATGTCAGCACTAAGGACCACCGGGAAC<br>ACCCCCCGGAACCATCGGAACCACCAGATACCACTAGCTAC<br>ATGATGGAGGACCCAGAATCGAATCAGAGCTTTAAGGATAT<br>TCTCCTGAACAAAAATAAGGAGATAAATCAACTACACCACC<br>CTACCGGAAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 164 | downstream<br>junction of<br>plant 4 | ACCGGAACTGGAACAGCAGGATCATAAGAGGACCTTGACA<br>TGGACTCCATCCAACCTATCGACAGAGGACAAGCAACAAATT<br>TACCAACCGTGGAACCTCTCTGTGATAGTAAAGGTATTTGG<br>AAAAAATCGCCACGCATACTTGAAAAACAAGTTGGTTGA<br>TCTATGGAAGCGATCAGAACCTCTAACACTGATAGATTTTG<br>GCTGTGAATACTTTATATTGCAAAATTCAATAATCCAACCA<br>GCCTACATAAGTCCCTCCATGAGGGTCCGTGGTTTCATCGCA<br>GGAAACTTCCTGTTAGTAAAAAAATGGGAGCCAACTTTG<br>TGCCAGACACATCAACACTCACCATAACAACGATATGGGCA<br>AGGCTGCTGCAACTCCCAGCGGAGTTCTATGACAGGCAAAT<br>ACTAGAAAAGGTAGGGGGAAAGCTCGGGTCCCTCCTAAAAA<br>TTGATACCTGCACCTCTGCTGCACTAAGAGGACGTTATGCA<br>CGCATACAGGTTTCACTAGAGAATCCAGTCAAGACGACGGT<br>CAAAATTGGAAACCATGTTCAAAAAGTGGTATACGAGGGGG<br>ACAAAATCCTTTGCACAGAATGTGGGAGACTCGGGAACACC<br>TTATTGACCTCATCCAGGATTTTGAGATGATGGGTACACGA<br>TTATAAAAAAGTTGATCTATGATTTAAATTTGATCGGTTTA<br>ATATTTAAATTTTACTACTAAAAACCGTTAAATTTTAAA<br>ATTATAGGTCTAAAATTAATTCCTTATATATATATATACA<br>CACACCAATTACCACTTAGAGAAGTGTTATCTAATTTTAGA<br>AAGAAAAATAAAACAAGATAAATATAAATTTCAAATTTCTA<br>ACCTCGTGGAGAGAGGTGCACCCAGTCATAATCGCATTATG<br>TGATACTTCAAGTG |

|     |                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 165 | upstream<br>junction of<br>plant 4   | AGATCGAGTGAGAAGTAGCTGGAAACATCATGAGTGGCAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 166 | downstream<br>junction of<br>plant 4 | AGTGGCAGAAGTGGGAAGAGATAAACTCATGATGATTGTAA<br>TGAGGGTGGTGGACAAGATGAATCTGGTGCCCCAAAACA<br>AAAATACTAATGCCAACAAAAGATCAGGACCAACGGTGCCA<br>CCTAAAAGGGGAAGCATAGCAAAACAGATAGTACGAGATTT<br>AAAGGATACATCAAGCTCTCTGAGTACTGTATTACATTGT<br>TTTTCTTTAACTTCCTTCTCATGGCGATTATATCGACAAAT<br>TATGAGAACAAAATATAGGAAGTTTACAACATTGAGGAAAG<br>CAAGTAACCAGTAGTAATAATCTAAATGACCATTGTTAATA<br>TTACTTGACAACCAGCTAACTCCACCTCCATATGAAGTAAC<br>ACTATCCACAACATTCCTAAAACACTCCCCAAAAAGCCAG<br>CTACAGACATTCCTAAGTGTGAAAATAGCCACAACAAAATA<br>GACATGCTTTTTTGGGAAGTTCAGAGTAAAGGAACTCTACCAA<br>TCCGATTGCATTGAAAGCATCAGCTAGTCCAAGAAGCACGT<br>ACTGTGGCACGAACCACATAGCCGACATGTTTATATTTAGA<br>CTGTCTTGTGGATCTTCTGATCAATTGCTATGCCCCGCCT<br>TATGCCTTCTGTTATCGCTGAAAGTACCATCG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| 167 | virC2 region                         | TTAACTCCGCTCGATATCGATGAAGCATTGTCGACCTACCG<br>CTATGTCATTGAACTGCTGCTGAGCGAGAACTTGGCAATTC<br>CGACAGCCGTATTGCGCCAACGCGTGCCGGTTGGTCGATTG<br>ACCACATCGCAGCGCGCGATGTCGGACATGCTCGCAAGCCT<br>TCCAGTTGTACAGTCTCCCATGCACGAGAGAGACGCATTTG<br>CCGCGATGAAGGAACGTGGCATGTTGCATCTCACATTGCTG<br>AATATGAGAACCGATCCGACAATGCGCCTCCTCGAGCGGAA<br>TCTCAGAATCGCCATGGAGGAACTCGTCACTATCTCCAAAT<br>TGGTTAGCGAAGCCTTGGAGGGGTGAAGATGGGAATTCGCA<br>AACCCGCTTTGTCTGTCTGCGGGAGGCCAGGCGGCTTGCCGCC<br>GCTCGACCCGAAATCGTCCATCCTTCTTTGCCTGTTGCCAC<br>CCAAAACCTCGACCTGCGCGAGCCGCTGAAAATCTCGACG<br>AGGAAGATCGACGACCTGCCCCAGCCACCGCCAAGCGTTGT<br>CACAGCTCTGATCAGCAATCGATGCTGACCGTGGATGCTTT<br>GAGTTCGACGACAGCGCCAGAAAGGATCCAGGTCTTCCTTT<br>CAGCGCGCCCGCCCGCGCCTGAAGTATCGAAGATATATGAC<br>AACCTGATCCTGCAATACAGTCCTTCCAAGTCGCTACAAAT<br>GATCTTGCGCCGTGCGCTTGGCGATTTTGAAAACATGCTGG<br>CGGATGGATCGTTTTCGTGCGGCTCCGAAGAGTTATCCGATC<br>CCTCACACAGCTTTCGAAAAATCAATCATCGTTTACAGCCTC<br>CCGCATGTTCCCGGTCTCGCTAATAGAAGCCGCTCGCAATC<br>ACTTTGATCCATTGGGATTGGAGACCGCCCGGGCTTTCGGC<br>CACAAGCTGGCTACCGCAGCGCTTGCATGTTTCTTTGCTCG<br>GGAGAAGGCAACGAACAGCTGATCTCTCAAAGATAGGACC<br>CATCCAATCACTCCGCAGTGTGAGTTTTTCGGATAGTACC<br>GAGGAAAGGCAGCTTTGCCAAGCCGCATAGCAATCTGCTCA<br>CGTTGGGAACAGATTGCTAAAGGCGAAATGCACCTCTACCT<br>CAGGCCGCCATCACACCCCCGTACGA |
| 168 | virC primer                          | GTTTAAACAGCTTCCTCCATAGAAGACGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 169 | virC primer                          | TTAATTAATCGTACGGGGGTGTGATGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 170 | primer                               | TGCTCCTGCCGAGAAAGTAT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 171 | primer                               | AGCCAACGCTATGTCCTGAT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 172 | primer                               | GAATCAGCTAATCAGGGAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 173 | human<br>AC027708                    | TGGCAGGATATATACATATGTACAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

|     |                                                                                                            |                                                                      |
|-----|------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| 174 | human<br>AC024192                                                                                          | CTGCAGGATATATTTCTCAGTAAAC                                            |
| 175 | human<br>AC003685                                                                                          | TGCCAGGATATATACATGGCTAATG                                            |
| 176 | human<br>AL390883                                                                                          | GGCCAGGATATATTACCCAGTAATT                                            |
| 177 | human<br>AC022858                                                                                          | AGGCAGGACTTCTGTGTATGTTAAC                                            |
| 178 | mouse<br>AC110541                                                                                          | AGGCAGGACTTAATGTGGTGTAAC                                             |
| 179 | mouse<br>AC132685                                                                                          | TGGCAGGATATATATCTTGGTAAAT                                            |
| 180 | rat AC096051                                                                                               | TGGCAGGATATATGGCATTGTCATT                                            |
| 181 | Neurospora<br>BX897673                                                                                     | ATACAGGATATATAGGTAGGTAAAG                                            |
| 182 | Saccharomyces<br>AJ316068                                                                                  | AGACAGGATATATTGGAAGGTATTC                                            |
| 183 | primer                                                                                                     | GCCATGCGCGTTGTTTCACATCG                                              |
| 184 | UF region of<br>pSIM108                                                                                    | TCCTTCATAGCTACACTTTCTAAAGGTACGATAGATTTTGG<br>ATCAACCACACACACTTC      |
| 185 | UF region of<br>pSIM843B                                                                                   | GTAAAAAATAAAAGTGAAAATTCAATGAATTAACACAAATA<br>TAAATGTAATATAAAATT      |
| 186 | UF region<br>ofpSIM781                                                                                     | TGTTGAAGGCTTGATGTGATTAAGAAGGCCGAGGCTGTTG<br>GATCTAGTTCTTGAAGTT       |
| 187 | C-clusters of<br>pSIM831                                                                                   | CCACAATATATCCTGCCACCGGATATATTGCCTAGGAGCCA<br>GCCAACAGCTCCCCGACC      |
| 188 | C-clusters of<br>pSIM843                                                                                   | CCTCTGTATACATCCTGCCGCCAAGCTTCCAGCCACCTAGG<br>AGCCAGCCAACAGCTCCCCGACC |
| 189 | modified<br>Brassica rapa<br>cleavage site<br>Br01M1                                                       | CCCTACTGTATAATAAATCCTGTAG                                            |
| 190 | modified<br>Brassica rapa<br>cleavage site<br>BrM2                                                         | CTCTACTGTATAATAAATCCTGTGG                                            |
| 191 | approximate<br>part of<br>alternative<br>final<br>cleavage site<br>that is<br>transferred<br>to plant cell | CCCGAAAAACGGGA                                                       |
| 192 | Arabidopsis<br>sequence<br>resembling<br>SEQ ID 191                                                        | ACCGAAAAACGGGA                                                       |
| 193 | Maize Zm03                                                                                                 | GCGTACGCATTTATATATCCTGTGG                                            |
| 194 | Zm03-modified<br>cleavage site                                                                             | GCTTACGCATTTATATATCCTGTGG                                            |

|     |                                    |                                                                   |
|-----|------------------------------------|-------------------------------------------------------------------|
|     | Zm03M1                             |                                                                   |
| 195 | Gm01-derived<br>Gm01M1             | AAATACTGTTTTATATATCCTGTCA                                         |
| 196 | Gm02-derived<br>Gm02M1             | AATTACTCTGAAATATATCCTGTGT                                         |
| 197 | Brassica rapa<br>Br02              | TGGAAGTGTCTATATGTCCTGTCA                                          |
| 198 | Br02-derived<br>Br02M1             | AGGAAGTGTCTATATGTCCTGTCA                                          |
| 199 | UI region of<br>SEQ ID: 87         | CTTAGAGATCTCAAACAAACATACACAGCGACTTATTCACA<br>ACTAGTC              |
| 200 | UI-like<br>region of SEQ<br>ID: 91 | AGAAACAATCAAACAAACATACACAGCGACTTATTCACACG<br>AGCTCAA              |
| 201 | UI-like<br>region of SEQ<br>ID: 92 | GCCCTTTTAAATATCCGATTATTCTAATAAACGCTCTTTTC<br>TCTTAGG              |
| 202 | UI-like<br>region of SEQ<br>ID: 93 | TGACGAAGTACGAAGTACGAAGTACGAAGTACGAAGTACGAAGT<br>GACGAAC           |
| 203 | UI-like<br>region of SEQ<br>ID: 94 | TAACAATTGAACAATTGAACAATTGAACAATTGAACAATTG<br>AACAAAC              |
| 204 | UI-like<br>region of SEQ<br>ID: 95 | TAGACATTGCACATCCAAAGGCAGGCACGTACAAACGAATT<br>TATTTAG              |
| 205 | UI-like<br>region of SEQ<br>ID: 96 | GAAGGCACGAAGGCACGAAGGCACGAAGGCACGAAGGCACG<br>AAGGCAC              |
| 206 | UI-like<br>region of SEQ<br>ID: 97 | TCATCACCGCCGTCCTAAACAAACATACCTCCACACAAATT<br>TATCTAC              |
| 207 | UI region of<br>SEQ ID: 98         | AGATCTCAAACAAACATACACAGCGACTTATTCACAACACTAG<br>TACCAAC            |
| 208 | UI region of<br>SEQ ID: 99         | TGACGAAGTACGAAGTACGAAGTACGAAGTACGAAGTACGAAGT<br>ACCAAC            |
| 209 | UI region of<br>SEQ ID: 100        | CTGACGAAGTACGAAGTACGAAGTACGAAGTACGAAGTACGAAGT<br>TACCAAC          |
| 210 | UI region of<br>SEQ ID: 101        | TGTCTTTATCTCTTGTTGCCAAAACGTCTCTCGAGTCGAGT<br>CACCAAC              |
| 211 | UF-like region of<br>SEQ ID 116    | TCTCCATATTGACCATCATACTCATTTGCTGATCCATGTAGA<br>TTTCCCGGACATGAAGCC  |
| 212 | UF-like region of<br>SEQ ID 117    | TGAATTGAGTACATTAAAAACGTCCGCAATGTGTTATTAAG<br>TTGTCTAAGCGTCAATTT   |
| 213 | UF-like region of<br>SEQ ID 118    | ATCTGGTAATATAGCAAAAACGTGCTCAAAAATCGCTTCAA<br>AGCTCTTGTAAGTCTAGCTC |
| 214 | UF-like region of<br>SEQ ID 119    | TACATTTTATATTCGATAAAGCATGCGTTAAAACGACTTCG<br>CATGTCCATATCTAATCT   |

|     |                                 |                                  |
|-----|---------------------------------|----------------------------------|
| 215 | AF-like region of<br>SEQ ID 116 | CCTGCCGCCGCTGCCGCTTTGCACCC       |
| 216 | AF-like region of<br>SEQ ID 116 | CCTGCCACCAGCCAGCCAACAGCTCCCCGACC |
| 217 | AF-like region of<br>SEQ ID 116 | CCACAATATATCCTGCCACCCC           |
| 218 | AF-like region of<br>SEQ ID 116 | CCTGCCACCCAAGGAGCGACGCCTTCTGGCC  |

**WHAT IS CLAIMED IS:**

1. A vector, comprising: (1) a first polynucleotide, comprising the consensus sequence [A/C/G]-[A/C/T]-[A/C/T]-[G/T]-A-[C/G]-NNNNNN-A-[G/T]-A-[A/C/T]-[A/G]-TCCTG-[C/G/T]-[A/C/G]-N that is consistent with SEQ ID NO:84, which is (i) nicked when exposed to an enzyme involved in bacterial-mediated plant transformation and (ii) not identical to a bacterial border sequence; (2) a second polynucleotide, which is a bacterial T-DNA border; and (3) a desired polynucleotide positioned between the first and second polynucleotides.
2. The vector of claim 1, wherein the sequence of the first polynucleotide is native to a plant genome.
3. The vector of claim 1, wherein the first polynucleotide is targeted and nicked by a vir gene-encoded protein.
4. The vector of claim 2, wherein the plant sequence is endogenous to (1) a monocotyledonous plant, wherein the monocotyledonous plant is wheat, turf grass, maize, rice, oat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, or palm; or (2) a dicotyledonous plant, wherein the dicotyledonous plant is potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia, legumes, alfalfa, soybean, pea, bean, cucumber, grape, brassica, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, cactus, cucumber, melon, canola, apple, or pine.
5. A method for transforming a plant cell, comprising introducing the vector of claim 1 into a plant cell.
6. The method of claim 5, wherein the plant cell is located in either (1) a monocotyledonous plant or explant thereof, wherein the monocotyledonous plant or explant thereof is wheat, turf grass, maize, rice, oat, barley, sorghum, orchid, iris, lily, onion, banana, sugarcane, or palm; or (2) a dicotyledonous plant or explant thereof, wherein the dicotyledonous plant or explant thereof is potato, tobacco, tomato, avocado, pepper, sugarbeet, broccoli, cassava, sweet potato, cotton, poinsettia,

legumes, alfalfa, soybean, pea, bean, cucumber, grape, brassica, carrot, strawberry, lettuce, oak, maple, walnut, rose, mint, squash, daisy, or cactus.

7. The method of claim 5, wherein the vector is a plasmid which is maintained in a bacterial strain, wherein the bacterial strain is *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, or *MesoRhizobium loti*.

8. A method for transforming a plant with a desired polynucleotide, comprising infecting a plant with a bacterial strain, wherein the bacterial strain is *Agrobacterium tumefaciens*, *Rhizobium trifolii*, *Rhizobium leguminosarum*, *Phyllobacterium myrsinacearum*, *SinoRhizobium meliloti*, or *MesoRhizobium loti*, wherein said bacterial strain comprises the vector of claim 1, and wherein said vector is a plasmid.

9. The vector of claim 1, wherein the sequence of the consensus sequence of the first polynucleotide is from a potato plant and is SEQ ID NO: 31, SEQ ID NO: 40, or SEQ ID NO: 86.

10. The vector of claim 1, wherein the sequence of the consensus sequence of the first polynucleotide is from a tomato plant and is SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 42 or SEQ ID NO: 43.

PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

U.S. Appl. No. 11/229,498 filed 09/07/2005

Docket No. 05895T-0229

Atty: Richard C. Peet, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 1 of 8

PCT/US05/31537

**A** Rb01/SEQ ID 1 GTTTACCCGCCAATATATCTCTGCA (AJ237588)  
Rb02/SEQ ID 2 AATTACAACGGTATATATCTCTGCA (X00493)  
Rb03/SEQ ID 3 CATGACAGGACATATATCTCTGCA (X68345)  
Rb04/SEQ ID 4 AATTACAACGGTATATATCTCTGCA (U83987)  
Rb05/SEQ ID 5 CCTGACCACAAGATATATCTCTGCA (X51338)  
Rb06/SEQ ID 6 CTAGACAAGGGATATATCTCTGCA (X68345)  
Rb07/SEQ ID 7 CATTACTTTAGAATATATCTCTGCA (AF002086)

CON1/SEQ ID 27  
CTTT C GCC C  
G T TGG

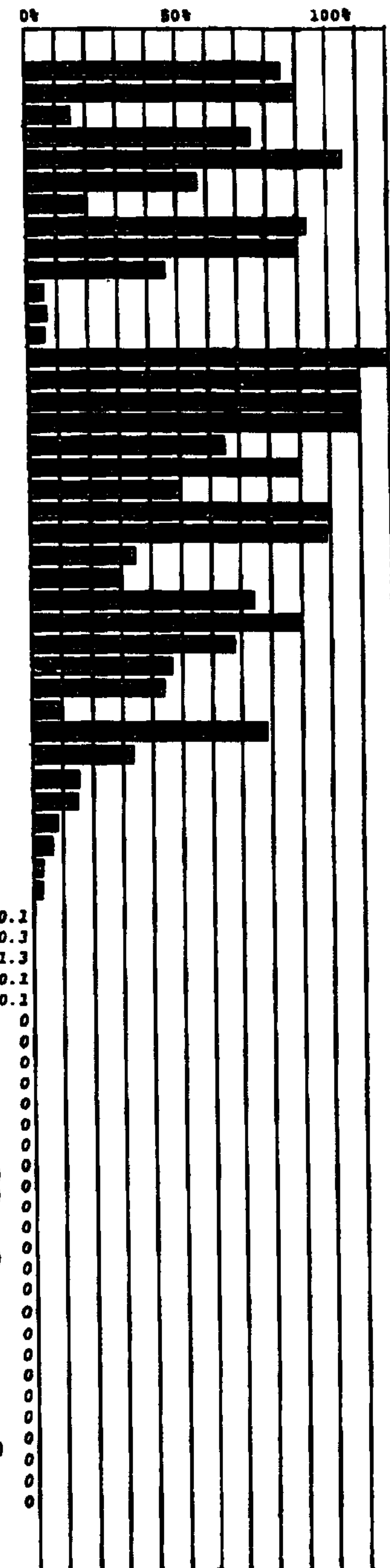
**B** Sy01/SEQ ID 8 CTTTACACAACAATATATCTCTGCA  
Sy02/SEQ ID 9 GTTTACACAACAATATATCTCTGCA  
Sy03/SEQ ID 10 GTTTACACAACAATATATCTCTGCA  
Sy04/SEQ ID 11 GTTTACACAACAATATATCTCTGCA  
Sy05/SEQ ID 12 GTTTACTCAACAATATATCTCTGCA  
Sy06/SEQ ID 13 GTTTACACAACAATATATCTCTGCA  
Sy07/SEQ ID 14 GTTTACACAACAATATATCTCTGCA  
Sy08/SEQ ID 15 GTTTACACAACAATATATCTCTGCA  
Sy09/SEQ ID 16 GTTTACACAACAATATATCTCTGCA  
Sy10/SEQ ID 17 GTTTACACAACAATATATCTCTGCA  
Sy11/SEQ ID 18 GTTTACACAACAATATATCTCTGCA  
Sy12/SEQ ID 19 GTTTACACAACAATATATCTCTGCA  
Sy13/SEQ ID 20 GTTTACACAACAATATATCTCTGCA

**C** St01/SEQ ID 219 GTTTACAGTACCATATATCTCTGCA (AY566555)  
Le01/SEQ ID 28 CATTACCAACAATATATCTCTGCC (AN398191)  
St02/SEQ ID 86 CATTACCAACAATATATCTCTGCC  
Ca01/SEQ ID 85 CATTACCAACAATATATCTCTGCC  
At01/SEQ ID 26 GTTGACATCAGATATATCTCTGCA (AL138652)  
St03/SEQ ID 31 GTTTACCTTAGCATATATCTCTGCC (SGN-U179068)  
St04/SEQ ID 40 AGCTACCCCAAAATATATCTCTGCC (SGN-U158307)  
Le02/SEQ ID 29 CTCTACCTCTGAATATATCTCTGCC  
Le03/SEQ ID 30 GCATACCTCTGAATATATCTCTGCC  
Le04/SEQ ID 41 GGAAGCTGTCAATATATCTCTGCA (SGN-U162952)  
Le05/SEQ ID 42 ACCTACCCCAAAATATATCTCTGCC (SGN-U153074)  
Ma01/SEQ ID 32 GTATACCTCTGTATATATCTCTGCC  
Hv01/SEQ ID 33 ATATACCAAGGATATATCTCTGCC (TC137112)  
Os01/SEQ ID 34 ACTTACTCAAGGATATATCTCTGCC (AL772425)  
Os02/SEQ ID 35 CACTACAAAAAATATATCTCTGCC (AP004161)  
Os03/SEQ ID 36 ATGTACTATATATATATCTCTGTT (AP005498)  
Sm01/SEQ ID 50 GGGTACGCAATATATCTCTGTT (CF052264)  
Ta01/SEQ ID 37 ATATACGGAGCAATATATCTCTGCC (CA500064)  
Mt01/SEQ ID 49 GTATACCTCTGTATATATCTCTGCC (CA919279)  
Le06/SEQ ID 43 GTTTAGCTGTATATATCTCTGCC (SGN-U146678)  
Le07/SEQ ID 44 TCTTAGAATCAATATATCTCTGAC (AN093946)  
Le08/SEQ ID 45 CGTTACACGTATATATCTCTGTA (SGN-U156238)  
Le09/SEQ ID 46 GAATTAATTTGCATATATCTCTGTA (SGN-U160800)  
Le10/SEQ ID 53 TTGTCTCTGCCATATATCTCTGCC (SGN-U149760)  
Le11/SEQ ID 54 GGTACCATGTAGATATATCTCTGCT (SGN-U150174)  
Le12/SEQ ID 55 GTTGGGTTGACATATATCTCTGCA (AN618843)  
Le13/SEQ ID 56 CACTACCGGCTCATAGTTCTCTGCC  
Gm01/SEQ ID 57 TAAAGCAACCATATATCTCTGCA (p8IM593)  
Mt02/SEQ ID 58 GATTAGACAAATATATCTCTGCC 0.2  
Os04/SEQ ID 59 CTCTACTACCGGATATCTCTGCC (AL713937) 0.3  
St05/SEQ ID 60 GTTTGACAGACATATATCTCTGCA (p8IM606) 1.3  
St06/SEQ ID 61 GTTTACCGTGGCACTATATCTCTGCA (p8IM607) 0.2  
St07/SEQ ID 62 CATTACCAACTATATATCTCTGCC (p8IM561) 0.1  
Le14/SEQ ID 63 GTTTACTGAAGATATATCTCTGCA (BI421743) 0  
Le15/SEQ ID 64 TTCTATACGAAGATATATCTCTGCA (SGN-U155057) 0  
Le16/SEQ ID 65 TTCTAGCTGCAATATATCTCTGCC (SGN-U152635) 0  
Le17/SEQ ID 66 GTTGACATGGATATATCTCTGCA (SGN-U160333) 0  
Le18/SEQ ID 67 GTTCACTTAGCATATATCTCTGAT (OB683) 0  
Le19/SEQ ID 68 TTCTAGAGTAGATATATCTCTGTT (BG131232) 0  
Le20/SEQ ID 69 TGATTCATCAATATATCTCTGCC (AY013258) 0  
Le21/SEQ ID 70 ATCCGCCACCATTTATATCTCTGCC (SGN-U149980) 0  
Le22/SEQ ID 71 CATCCGCCACATTTATATCTCTGCC (SGN-U149979) 0  
Le23/SEQ ID 72 GTCAAGGAAGTGAATATATCTCTGACA (AN031162) 0  
Le24/SEQ ID 73 GTTTAACCAATATATATCTCTGAT (BI205800) 0  
Le25/SEQ ID 74 AGTTAATACCTATATATCTCTGTT (SGN-U111362) 0  
Le26/SEQ ID 75 CTAAAGTTGTACATAATCTCTGCT (AC135462) 0  
Le27/SEQ ID 76 TTCTACACAAGACAAATCTCTGCC (AC121237) 0  
Le28/SEQ ID 77 ATTAACAAGTTAGAACTCTGCC (AC144619) 0  
Mt03/SEQ ID 78 CATGACCCCTGCAATATATCTCTGTT 0  
Zm02/SEQ ID 79 AACTTAAGATAAGAACTCTGCC (AY107199) 0  
As01/SEQ ID 80 CTGTACAATAGCAAAATCTCTGTT (X75777) 0  
St08/SEQ ID 81 TTTTACCCCTG-ATATATCCCAAGC (p8IM604) 0  
Le29/SEQ ID 82 GATTGCA-TCAAATATATCTCTGCC (AY013255) 0  
Le30/SEQ ID 83 AAGTACC-GATGATATATCTCTGCT (SGN-U155745) 0

**D** CON2/SEQ ID 84  
AAAGACNNNNNATAAATCTCTGCAN  
CCCT G CG GC  
GTT T TG

Figure 1

Transformation frequencies



WASH\_1508321.1



PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

U.S. Appl. No. 11/220,408 filed 09/07/2005

Docket No.: 058951-0229

Atty: Richard C. Peet, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 3 of 8

Figure 3

| A                                                                                                  | Transformation-% |      |      |
|----------------------------------------------------------------------------------------------------|------------------|------|------|
|                                                                                                    | T                | TB   | B    |
| pT115955 (SEQ ID 116)<br>TCTCCATATTGACCATCATACTCATTTGCTGATCCATGTAGATTTCCCGACATGAAGCC<br>[REDACTED] | 26±8             | 55±9 | 19±6 |
| pT1C58 (SEQ ID 117)<br>TGAATTCAGTACATTAAAAACGTCCCAATGTGTTATTAAGTTGTCTAAGCGTCAATTT<br>[REDACTED]    | ND               |      |      |
| pR12659 (SEQ ID 118)<br>ATCTGGTAATATAGCAAAAACGTGCTCAAAAATCGCTTCAAAGCTCTTGTACTTAGCTC<br>[REDACTED]  | ND               |      |      |
| pR1A4 (SEQ ID 119)<br>TACATTTTATATTGATAAAGCATGCGTTAAAAACGACTTCGCATGTCCATATCTAATCT<br>[REDACTED]    | ND               |      |      |
| B                                                                                                  |                  |      |      |
| >831 (SEQ ID 123)<br>AAATCTGATTGATAAAGGATCGATCCTCTAGAGTCGACCTGCAGTACTTACGTACAATT<br>[REDACTED]     | 41±3             | 52±5 | 7±2  |
| >829 (SEQ ID 124)<br>AAATCTGATTGATAAAGGATCGATCCTCTAGAGTCGACCTGCAGTACTTACGTACAATT<br>[REDACTED]     | 17±2             | 68±5 | 15±4 |
| C                                                                                                  |                  |      |      |
| pSIM108 (SEQ ID 125)<br>TCCTTCATAGCTACACTTCTTAAAGGTACGATAGATTTTGGATCAACCACACACACTTC<br>[REDACTED]  | 41±5             | 48±6 | 11±3 |
| pSIM843B (SEQ ID 126)<br>GTAAAAATAAAGTGAATAATGAATTAACACAAATATAAATGTAATATAAATT<br>[REDACTED]        | 34±4             | 53±6 | 13±2 |
| pSIM849 (SEQ ID 127)<br>AATGGAGGTAAGTGTCTCTGCTCAGTGTGATAGATGTAAATATCTCTGTTATGAAGCC<br>[REDACTED]   | 10±2             | 83±7 | 7±2  |
| pSIM781 (SEQ ID 128)<br>TGTTGAAGGCTTGGATGTGATTAAGAAGGCGAGGCTGTTGGATCTAGTTCTTGAAGTT<br>[REDACTED]   | 11±1             | 58±3 | 31±2 |

WASH\_1508321.1

PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

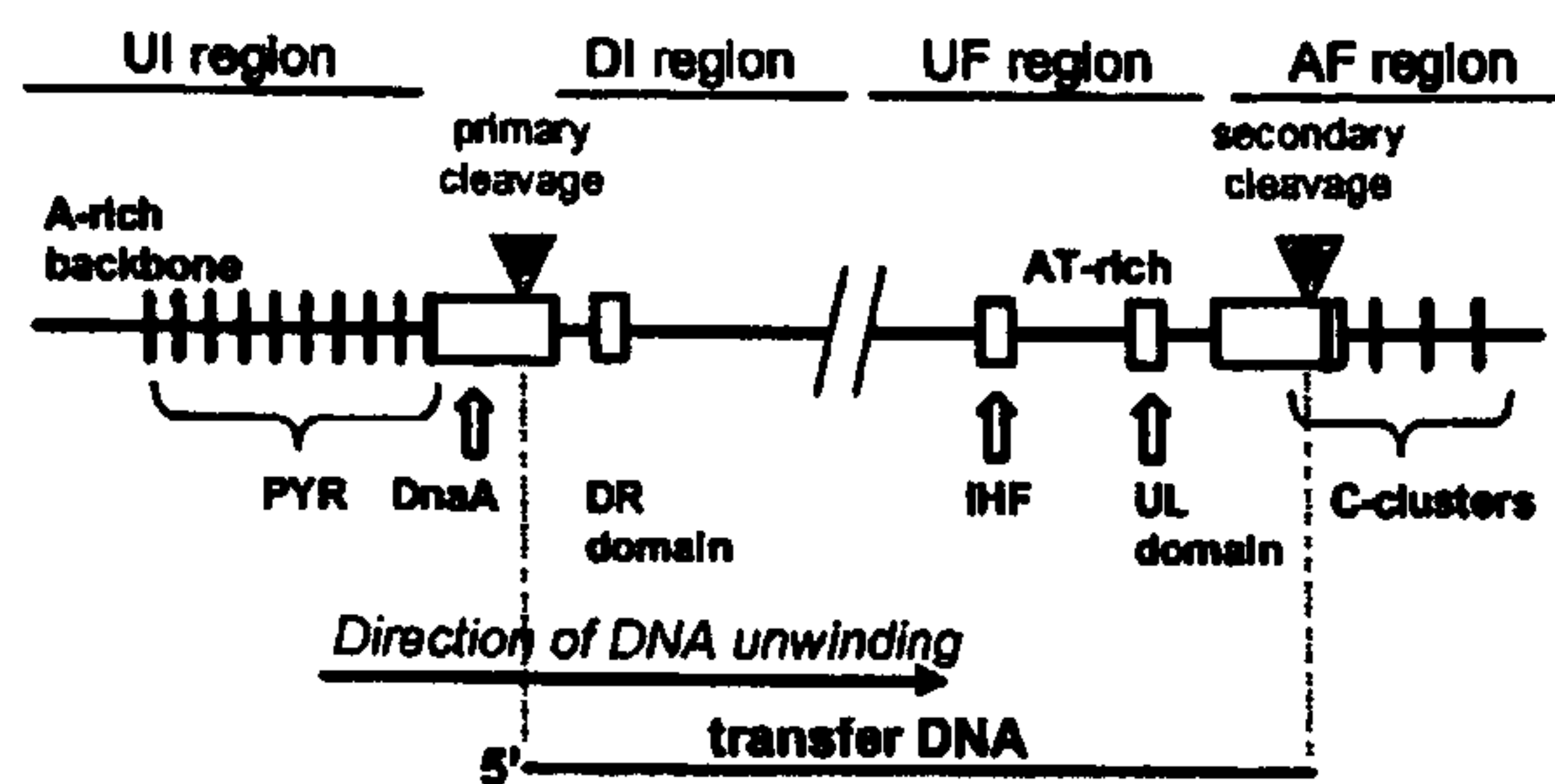
U.S. Appl. No. 11/220,408 filed 09/07/2005

Docket No.: 058951-0229

Atty: Richard C. Peet, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 4 of 8

Figure 4



WASH\_1508321.1

AMENDED SHEET - IPEA/US

PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

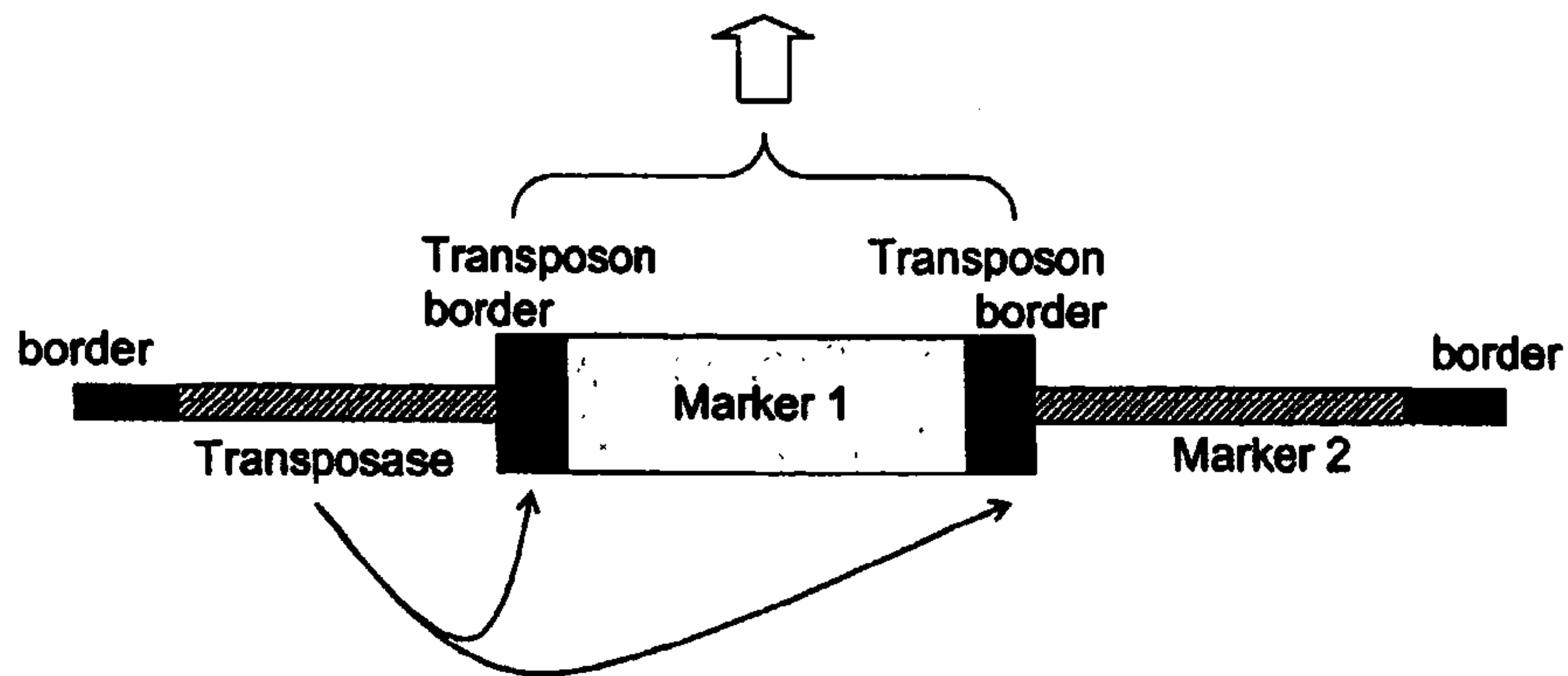
U.S. Appl.: 11/220,408 filed 09/07/2005

Docket No.: 058951-0229

Atty: Richard C. Peet, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 5 of 8

FIGURE 5: "Transposon" jumps from non-integrating T-DNA into plant genome

**Transformation process:**

- (1) Infect explants
- (2) T-DNA is transferred from Agro to plant cell nucleus
- (3) Select for plant cells containing Marker 1 (transposon)
- (4) Select against plant cells containing marker 2 (T-DNA)
- (5) Regenerate plants only containing the transposon

WASH\_1508321.1

PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

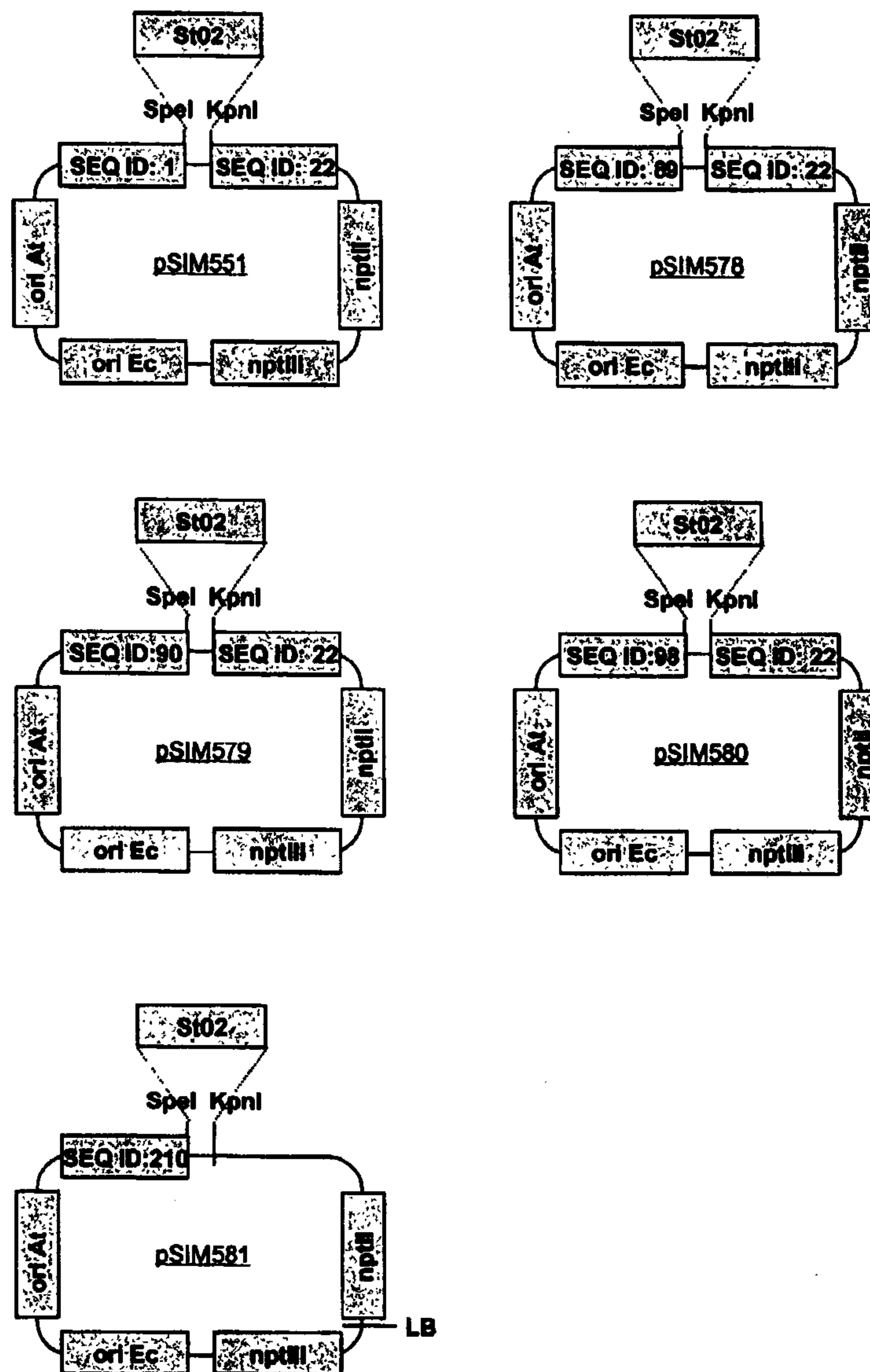
U.S. Appl.: 11/220,408 filed 09/07/2005

Docket No.: 058951-0229

Atty: Richard C. Peet, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 6 of 8

Figure 6A



PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

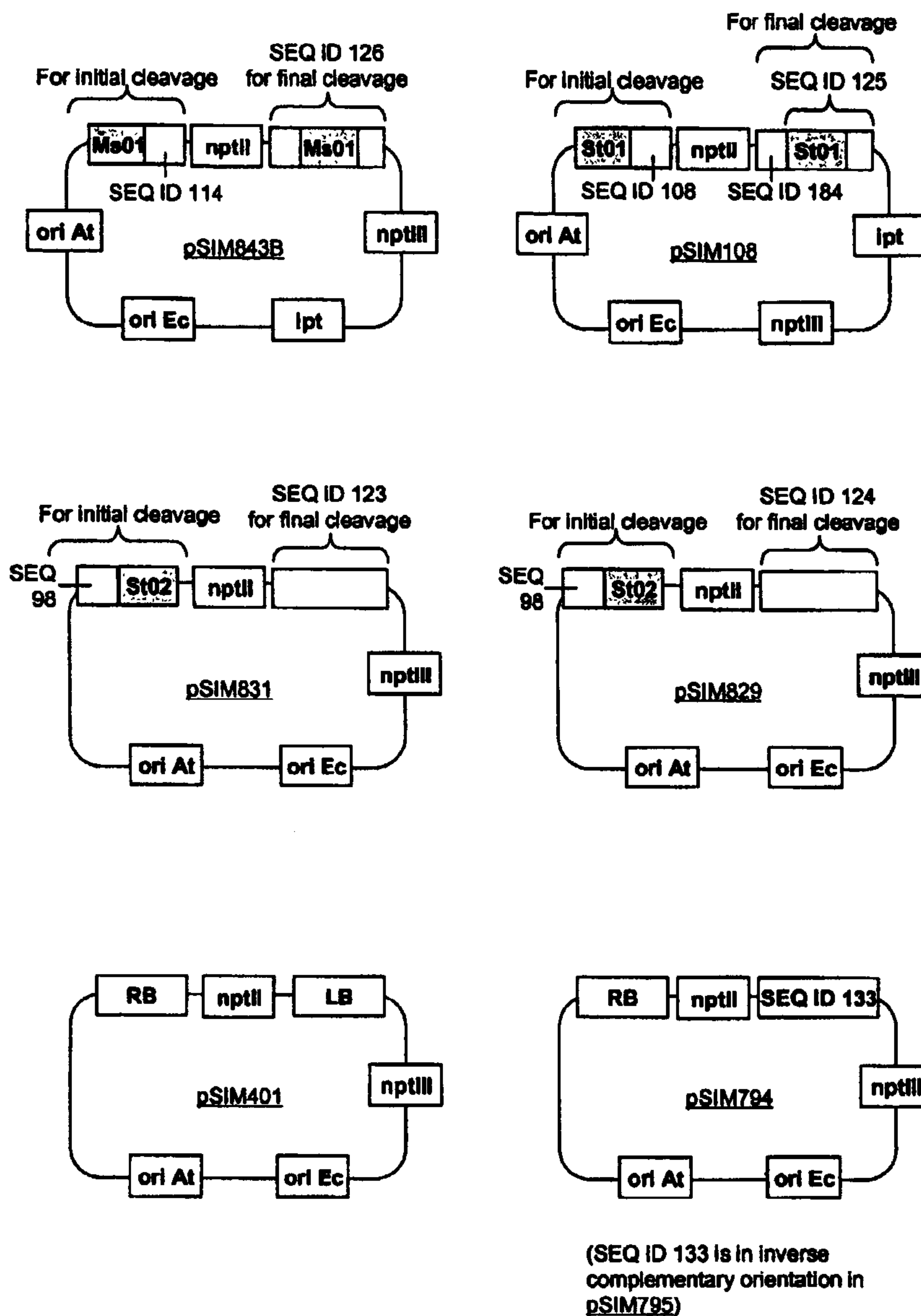
U.S. Appl. No. 14/229,408 filed 09/07/2004

Docket No. 058951-0229

Atty: Richard C. Pect, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 7 of 8

Figure 6B



PCT/US05/31537 17-04-2006

Title: Plant-Specific Genetic Elements and Transfer Cassettes for Plant Transformation

Inventors: Caius ROMMENS, et al.

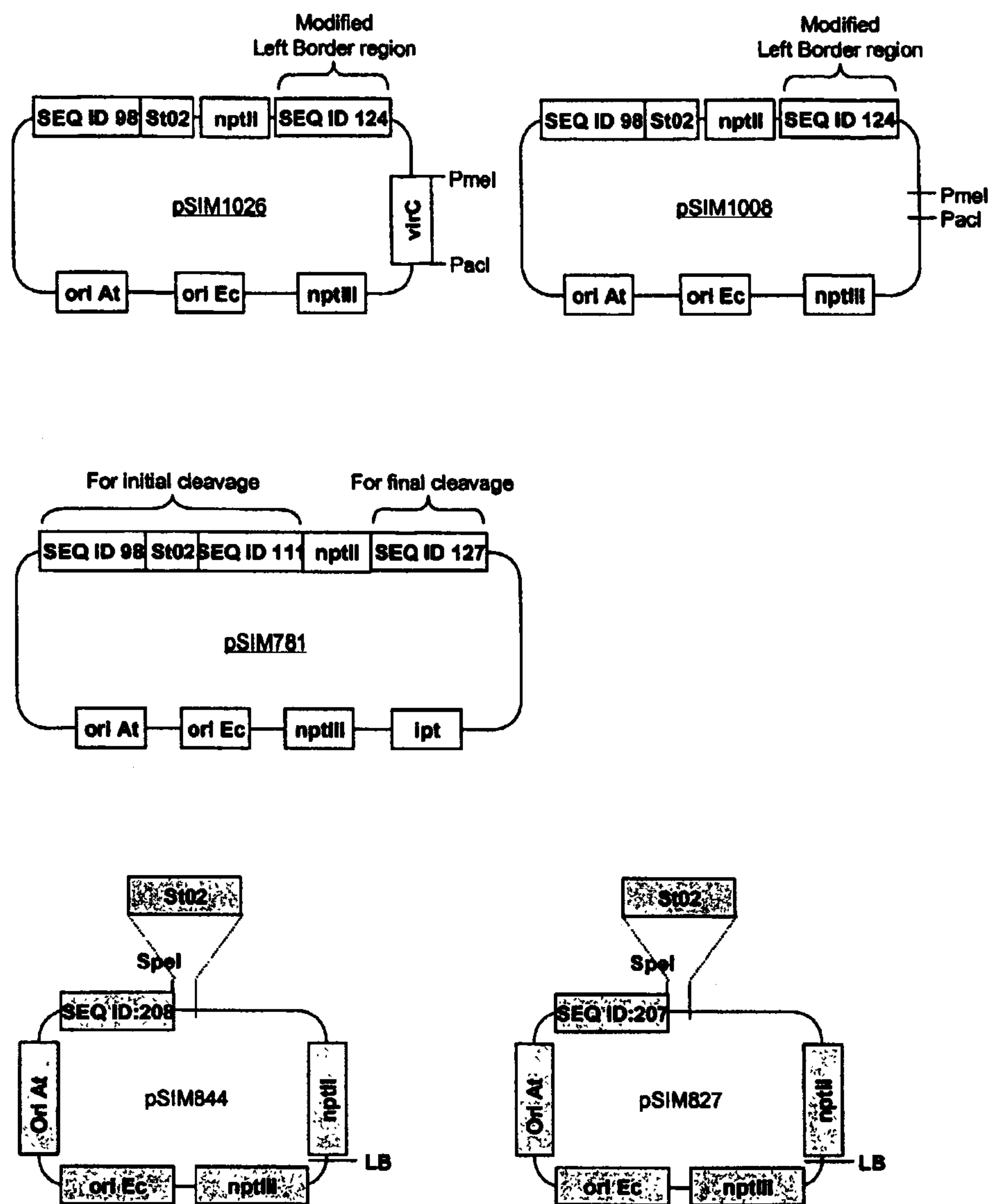
U.S. Appl. No. 11/220,408 filed 09/07/2005

Docket No.: 058951-0229

Atty: Richard C. Peet, Foley &amp; Lardner LLP

Replacement, Formal Drawing Sheet 8 of 8

Figure 6C



A

|               |                            |            |
|---------------|----------------------------|------------|
| Rb01/SEQ ID 1 | GTTTACCCGCCAATATATCTCTGTCA | (AJ237588) |
| Rb02/SEQ ID 2 | AATTACAACGGTATATATCTCTGCA  | (X00493)   |
| Rb03/SEQ ID 3 | CATGACAGGACATATATCTCTGTCA  | (X68345)   |
| Rb04/SEQ ID 4 | AATTACAACGGTATATATCTCTGTCA | (U03987)   |
| Rb05/SEQ ID 5 | CCGACCCACAAGATATATCTCTGTCA | (X51336)   |
| Rb06/SEQ ID 6 | CTAGACAAGGGATATATCTCTGTCA  | (X68345)   |
| Rb07/SEQ ID 7 | CATTACTTTAGATATATCTCTGTCA  | (AF002086) |

CON1/SEQ ID 21

|     |   |     |   |
|-----|---|-----|---|
| CTT | C | GCC | C |
| G   | T | TGG |   |

B

|                |                             |
|----------------|-----------------------------|
| Sy01/SEQ ID 8  | GTTTACACACAAATATATCTCTGTCA  |
| Sy02/SEQ ID 9  | GTTTACACACAAATATATCTCTGTCA  |
| Sy03/SEQ ID 10 | GTTTACACACAAATATATCTCTGTCA  |
| Sy04/SEQ ID 11 | GTTTACACACAAATATATCTCTGTCA  |
| Sy05/SEQ ID 12 | GTTTACTCAACAAATATATCTCTGTCA |
| Sy06/SEQ ID 13 | GTTTACACACAAATATATCTCTGTCA  |
| Sy07/SEQ ID 14 | GTTTACACACAAATATATCTCTGTCA  |
| Sy08/SEQ ID 15 | GTTTACACACAAATATATCTCTGTCA  |
| Sy09/SEQ ID 16 | GTTTACACACAAATATATCTCTGTCA  |
| Sy10/SEQ ID 17 | GTTTACACACAAATATATCTCTGTCA  |
| Sy11/SEQ ID 18 | GTTTACACACAAATATATCTCTGTCA  |
| Sy12/SEQ ID 19 | GTTTACACACAAATATATCTCTGTCA  |
| Sy13/SEQ ID 20 | GTTTACACACAAATATATCTCTGTCA  |

C

|                 |                             |               |
|-----------------|-----------------------------|---------------|
| St01/SEQ ID 219 | GTTTACACACAAATATATCTCTGTCA  | (AT566565)    |
| La01/SEQ ID 28  | CATTACCAACAAATATATCTCTGTCA  | (AM398191)    |
| St02/SEQ ID 86  | CATTACCAACAAATATATCTCTGTCA  |               |
| Ca01/SEQ ID 85  | CATTACCAACAAATATATCTCTGTCA  |               |
| At01/SEQ ID 26  | GTTGACATCAGATATATCTCTGTCA   | (AL138652)    |
| St03*/SEQ ID 31 | GTTTACCTTAGCATATATCTCTGTCA  |               |
| St04/SEQ ID 40  | AGTTACCCCAAAATATATCTCTGTCA  | (SGN-U179068) |
| La02/SEQ ID 39  | CTCTACCTCTGAATATATCTCTGTCA  | (SGN-U158107) |
| La03*/SEQ ID 30 | GCATACCTCTGAATATATCTCTGTCA  |               |
| La04/SEQ ID 41  | GGAACTGTCTAATATATCTCTGTCA   | (SGN-U162952) |
| La05/SEQ ID 42  | ACCTACCCCAAAATATATCTCTGTCA  | (SGN-U153074) |
| Ma01*/SEQ ID 33 | GTATACCTCTGTATATATCTCTGTCA  |               |
| Hv01/SEQ ID 33  | ATATACCAAGGATATATCTCTGTCA   | (TC137112)    |
| On01/SEQ ID 34  | ACTTACTCAAGATATATCTCTGTCA   | (AL772425)    |
| On02/SEQ ID 35  | CATTACCAAAATATATCTCTGTCA    | (AP004161)    |
| On03/SEQ ID 36  | ATGTACTATATATATATCTCTGTCA   | (AP005498)    |
| Sm01/SEQ ID 50  | GCTTACCTCTGAATATATCTCTGTCA  | (CF052364)    |
| Ta01/SEQ ID 37  | ATATACCAAGGATATATCTCTGTCA   | (CA500064)    |
| Me01/SEQ ID 49  | GTATACCTCTGTATATATCTCTGTCA  | (CA919279)    |
| La06/SEQ ID 43  | GTTTACCTCTGTATATATCTCTGTCA  | (SGN-U146874) |
| La07/SEQ ID 44  | TCTTACCACTCAATATATCTCTGTCA  | (AM092946)    |
| La08/SEQ ID 48  | CCTTACCACTCAATATATCTCTGTCA  | (SGN-U156238) |
| La09/SEQ ID 46  | GAATATTTTGCATATATCTCTGTCA   | (SGN-U160800) |
| La10/SEQ ID 53  | TTTCTCTGTGCAATATATCTCTGTCA  | (SGN-U149760) |
| La11/SEQ ID 54  | GCTTACCTCTGTATATATCTCTGTCA  | (SGN-U150174) |
| La12/SEQ ID 55  | GTTTACCTCTGTATATATCTCTGTCA  | (AM618641)    |
| La13/SEQ ID 56  | CATTACCTCTGTATATATCTCTGTCA  |               |
| On01/SEQ ID 57  | TAAAGCAACAGATATATCTCTGTCA   | (pSIN593)     |
| Me02/SEQ ID 58  | GATTACCAACAAATATATCTCTGTCA  |               |
| On04/SEQ ID 59  | CTCTACTACCCGAGATATCTCTGTCA  | (AL713937)    |
| St05*/SEQ ID 60 | GTTTACCAACAAATATATCTCTGTCA  | (pSIN506)     |
| St06*/SEQ ID 61 | GTTTACCTCTGTATATATCTCTGTCA  | (pSIN507)     |
| St07/SEQ ID 62  | CATTACCAACAAATATATCTCTGTCA  | (pSIN561)     |
| La14/SEQ ID 63  | GTTTACTTCAAGATATATCTCTGTCA  | (BI421743)    |
| La15/SEQ ID 64  | TCTTACCAACAAATATATCTCTGTCA  | (SGN-U155057) |
| La16/SEQ ID 65  | TCTTACCTCTGAATATATCTCTGTCA  | (SGN-U152635) |
| La17/SEQ ID 66  | GTTGACATCAGATATATCTCTGTCA   | (SGN-U160333) |
| La18/SEQ ID 67  | GTTTACCTTACGATATATCTCTGTCA  | (OB683)       |
| La19/SEQ ID 68  | TCTTACCAACAAATATATCTCTGTCA  | (BI131238)    |
| La20/SEQ ID 69  | TGATTCATCAAAATATATCTCTGTCA  | (AY013288)    |
| La21/SEQ ID 70  | ATGCTCACCAATATATATCTCTGTCA  | (SGN-U149980) |
| La22/SEQ ID 71  | CATGCTCACCAATATATATCTCTGTCA | (SGN-U149979) |
| La23/SEQ ID 72  | GTCAGCAAGTCAATATATCTCTGTCA  | (AM031162)    |
| La24/SEQ ID 73  | GTTTACCAACAAATATATCTCTGTCA  | (BI205800)    |
| La25/SEQ ID 74  | AGTTTAAACCTATATATCTCTGTCA   | (SGN-U111362) |
| La26/SEQ ID 75  | CTAAAGTTGACATATATCTCTGTCA   | (AC135462)    |
| La27/SEQ ID 76  | TCTTACCAACAAATATATCTCTGTCA  | (AC121237)    |
| La28/SEQ ID 77  | ATTAACAACCTTAAATATCTCTGTCA  | (AC144619)    |
| Me03/SEQ ID 78  | CATGACCTCTCAATATATCTCTGTCA  |               |
| Em02/SEQ ID 79  | AAGTTAAAGATAAAGATCTCTGTCA   | (AY107199)    |
| Am01/SEQ ID 80  | CTGTACCAATACCAAAATCTCTGTCA  | (X75777)      |
| St08*/SEQ ID 81 | TTTACCCCTG-ATATATCCCAAGCC   | (pSIN504)     |
| La29/SEQ ID 82  | GATTGCA-TCAAATATATCTCTGTCA  | (AY013255)    |
| La30/SEQ ID 83  | AAGTACC-GATGATATATCTCTGTCA  | (SGN-U155748) |

D

CON2/SEQ ID 44

|      |   |    |    |
|------|---|----|----|
| CCCT | G | CG | GC |
| GTT  | T | TG |    |

Transformation frequencies

