



(86) Date de dépôt PCT/PCT Filing Date: 2003/01/31
(87) Date publication PCT/PCT Publication Date: 2003/08/14
(45) Date de délivrance/Issue Date: 2012/07/31
(85) Entrée phase nationale/National Entry: 2004/07/09
(86) N° demande PCT/PCT Application No.: US 2003/002958
(87) N° publication PCT/PCT Publication No.: 2003/066809
(30) Priorités/Priorities: 2002/02/01 (US60/353,722);
2002/03/14 (US10/098,155); 2002/08/01 (US10/211,079)

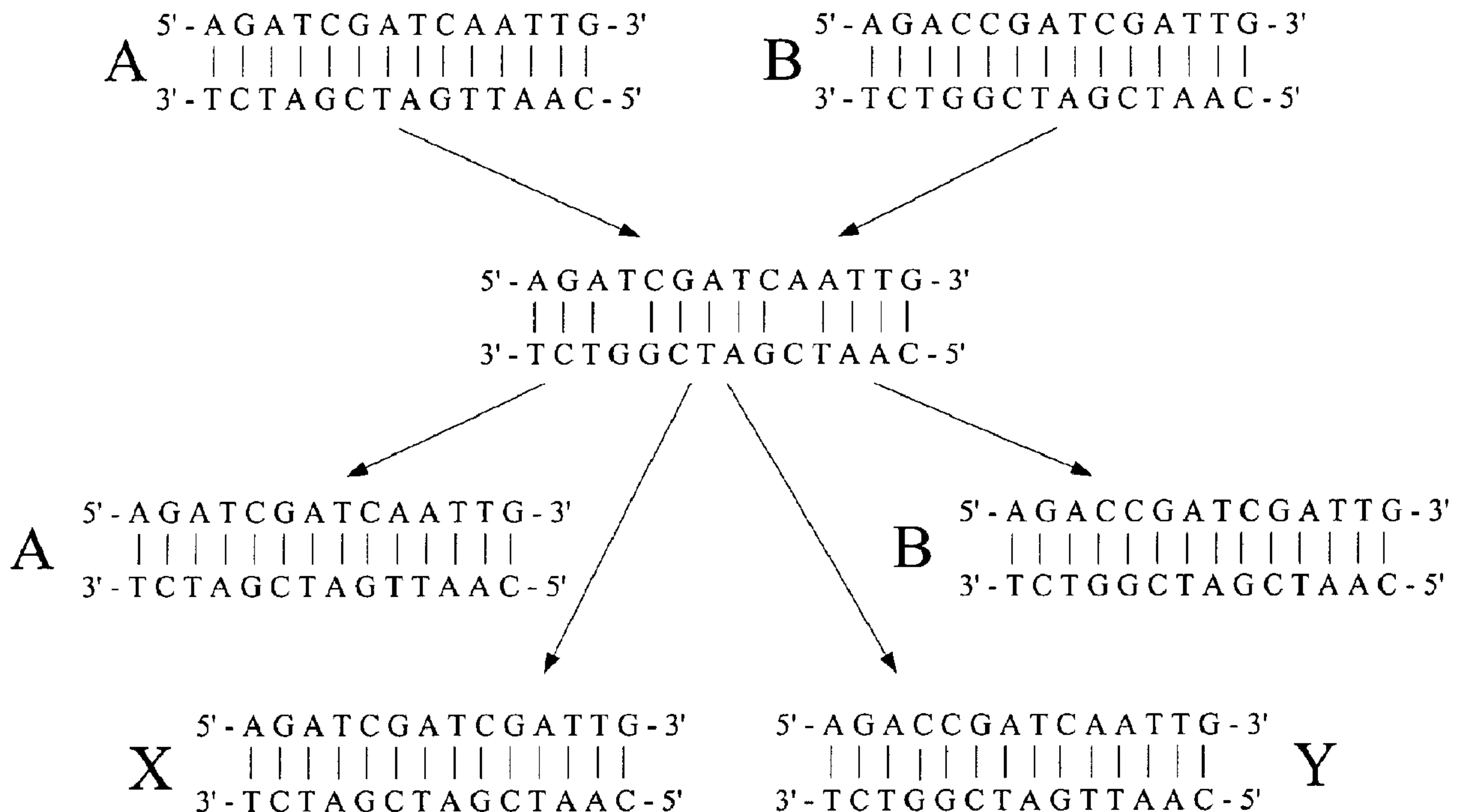
(51) Cl.Int./Int.Cl. *C12N 9/16* (2006.01),
C12N 15/10 (2006.01), *C12N 15/82* (2006.01),
C12N 9/22 (2006.01), *C12P 19/34* (2006.01),
C12P 21/00 (2006.01), *C12Q 1/44* (2006.01),
C12Q 1/68 (2006.01)

(72) Inventeurs/Inventors:
PADGETT, HAL S., US;
VAEWHONGS, ANDREW A., US;
VOJDANI, FAKHRIEH S., US;
SMITH, MARK L., US;
LINDBO, JOHN A., US;
FITZMAURICE, WAYNE P., US

(73) Propriétaire/Owner:

(54) Titre : ENDONUCLEASES NON APPARIEES ET LEURS PROCEDES D'UTILISATION

(54) Title: MISMATCH ENDONUCLEASES AND METHODS OF USE



(57) Abrégé/Abstract:

We describe here restriction endonucleases and their uses. Restriction endonucleases are useful in finding single nucleotide polymorphisms. They are also useful in an in vitro method of redistributing sequence variations between non-identical polynucleotide sequences.

(73) Propriétaires(suite)/Owners(continued):NOVICI BIOTECH LLC, US

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/066809 A2

(51) International Patent Classification⁷: **C12N**

(21) International Application Number: PCT/US03/02958

(22) International Filing Date: 31 January 2003 (31.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/353,722 1 February 2002 (01.02.2002) US
10/098,155 14 March 2002 (14.03.2002) US
10/211,079 1 August 2002 (01.08.2002) US

(71) Applicant (for all designated States except US): **LARGE SCALE BIOLOGY CORPORATION** [US]; 3333 Vaca Valley Parkway, Suite 1000, Vacaville, CA 95688 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PADGETT, Hal, S.** [US/US]; 638 Blossom Oak Court, Vacaville, CA 95687 (US). **VAEWHONGS, Andrew, A.** [US/US]; 999 Marshall No. 40, Vacaville, CA 95684 (US). **VOJDANI, Fakhrieh, S.** [US/US]; 1424 Pinnacles, Davis, CA 95616 (US). **SMITH, Mark, L.** [GB/US]; 909 Alvarado, Davis,

CA 95616 (US). **LINDBO, John, A.** [US/US]; 143 Sundance Drive, Vacaville, CA 95688 (US). **FITZMAURICE, Wayne, P.** [US/US]; 1218 Las Encinas Court, Vacaville, CA 95687 (US).

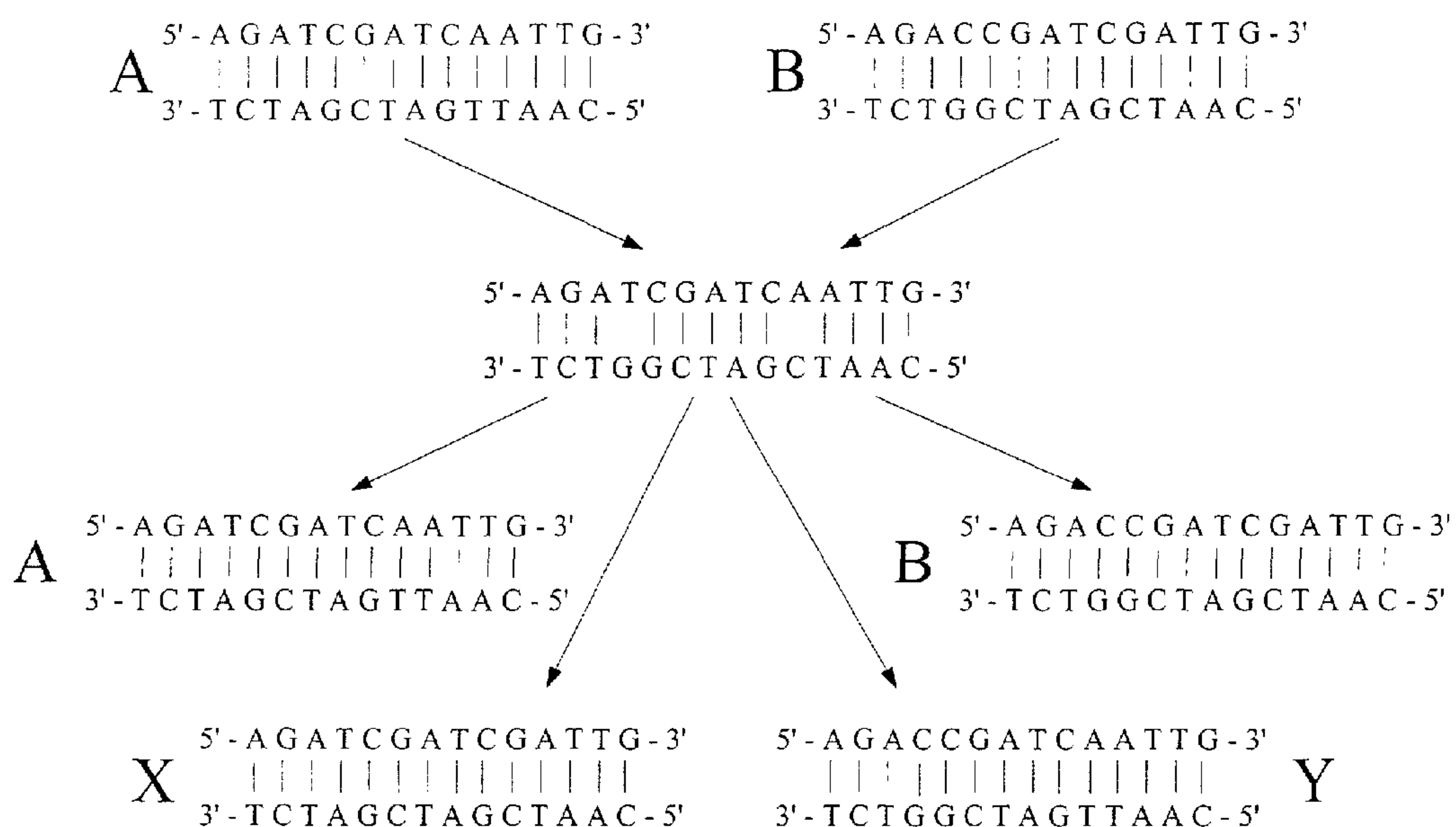
(74) Agent: **ROBBINS, John, C.**; Large Scale Biology Corporation, 3333 Vaca Valley Parkway, Vacaville, CA 95688 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

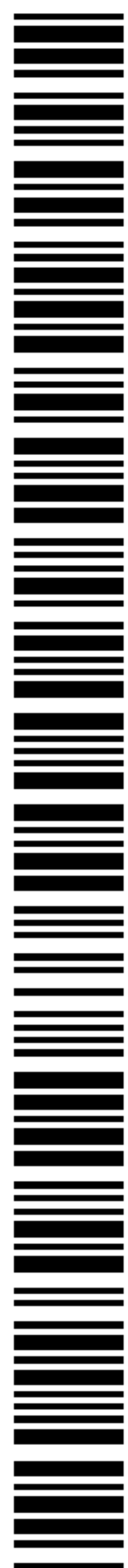
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF,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

[Continued on next page]

(54) Title: MISMATCH ENDONUCLEASES AND METHODS OF USE



(57) Abstract: We describe here restriction endonucleases and their uses. Restriction endonucleases are useful in finding single nucleotide polymorphisms. They are also useful in an *in vitro* method of redistributing sequence variations between non-identical polynucleotide sequences.



WO 03/066809 A2

WO 03/066809 A2



Published:

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

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.

MISMATCH ENDONUCLEASES AND METHODS OF USE

5

10

FIELD OF THE INVENTION

15 The invention relates generally to molecular biology and more specifically to mismatch endonuclease enzymes and their uses in methods of generating populations of related nucleic acid molecules or in detection of single nucleotide polymorphisms.

20

BACKGROUND INFORMATION

 DNA shuffling is a powerful tool for obtaining recombinants between two or more DNA sequences to evolve
25 them in an accelerated manner. The parental, or input, DNAs for the process of DNA shuffling are typically mutants or variants of a given gene that have some improved character over the wild-type. The products of DNA shuffling represent a pool of essentially random
30 reassortments of gene sequences from the parental nucleic acids that can then be analyzed for additive or synergistic effects resulting from new sequence combinations.

Recursive sequence reassortment is analogous to an evolutionary process where only variants with suitable properties are allowed to contribute their genetic material to the production of the next generation. Optimized variants are generated through DNA shuffling-mediated sequence reassortment followed by testing for incremental improvements in performance. Additional cycles of reassortment and testing lead to the generation of genes that contain new combinations of the genetic improvements identified in previous rounds of the process. Reassorting and combining beneficial genetic changes allows an optimized sequence to arise without having to individually generate and screen all possible sequence combinations.

Shuffling differs sharply from random mutagenesis, where subsequent improvements to an already improved sequence result largely from serendipity. For example, in order to obtain a protein that has a desired set of enhanced properties, it may be necessary to identify a mutant that contains a combination of various beneficial mutations. If no process is available for combining these beneficial genetic changes, further random mutagenesis will be required. However, random mutagenesis requires repeated cycles of generating and screening large numbers of mutants, resulting in a process that is tedious and highly labor intensive. Moreover, the rate at which sequences incur mutations with undesirable effects increases with the information content of a sequence. Hence, as the information content, library size, and mutagenesis rate increase, the ratio of deleterious mutations to beneficial mutations will increase, increasingly masking the selection of

further improvements. Lastly, some computer simulations have suggested that point mutagenesis alone may often be too gradual to allow the large-scale block changes that are required for continued and dramatic sequence
5 evolution.

Error-prone PCR uses low-fidelity polymerization conditions to introduce a low level of point mutations randomly over a sequence. A limitation to this method,
10 however, is that published error-prone PCR protocols suffer from a low processivity of the polymerase, making this approach inefficient at producing random mutagenesis in an average-sized gene.

15 In oligonucleotide-directed random mutagenesis, a short sequence is replaced with a synthetically mutagenized oligonucleotide. To generate combinations of distant mutations, different sites must be addressed simultaneously by different oligonucleotides. The
20 limited library size that is obtained in this manner, relative to the library size required to saturate all sites, requires that many rounds of selection are required for optimization. Mutagenesis with synthetic oligonucleotides requires sequencing of individual clones
25 after each selection round followed by grouping them into families, arbitrarily choosing a single family, and reducing it to a consensus motif. Such a motif is resynthesized and reinserted into a single gene followed by additional selection. This step creates a statistical
30 bottleneck, is labor intensive, and is not practical for many rounds of mutagenesis.

For these reasons, error-prone PCR and oligonucleotide-directed mutagenesis can be used for

mutagenesis protocols that require relatively few cycles of sequence alteration, such as for sequence fine-tuning, but are limited in their usefulness for procedures requiring numerous mutagenesis and selection cycles, especially on large gene sequences.

As discussed above, prior methods for producing improved gene products from randomly mutated genes are of limited utility. One recognized method for producing a randomly reassorted gene sequences uses enzymes to cleave a long nucleotide chain into shorter pieces. The cleaving agents are then separated from the genetic material, and the material is amplified in such a manner that the genetic material is allowed to reassemble as chains of polynucleotides, where their reassembly is either random or according to a specific order. The method requires several rounds of amplification to assemble variants of genes that were broken into random fragments. ((Stemmer, 1994a; Stemmer, 1994b), US Patent No. 5,605,793, US Patent No. 5,811,238, US Patent No. 5,830,721, US Patent No. 5,928,905, US Patent No. 6,096,548, US Patent No. 6,117,679, US Patent No. 6,165,793, US Patent No. 6,153,410). A variation of this method uses primers and limited polymerase extensions to generate the fragments prior to reassembly (US Patent No. 5,965,408, US Patent No. 6,159,687).

However, both methods have limitations. These methods suffer from being technically complex. This limits the applicability of these methods to facilities that have sufficiently experienced staffs. In addition there are complications that arise from the reassembly of molecules from fragments, including unintended mutagenesis and the increasing difficulty of the

reassembly of large target molecules of increasing size, which limits the utility of these methods for reassembling long polynucleotide strands.

5 Another limitation of these methods of fragmentation and reassembly-based gene shuffling is encountered when the parental template polynucleotides are increasingly heterogeneous. In the annealing step of those processes, the small polynucleotide fragments depend upon
10 stabilizing forces that result from base-pairing interactions to anneal properly. As the small regions of annealing have limited stabilizing forces due to their short length, annealing of highly complementary sequences is favored over more divergent sequences. In such
15 instances these methods have a strong tendency to regenerate the parental template polynucleotides due to annealing of complementary single-strands from a particular parental template. Therefore, the parental templates essentially reassemble themselves creating a
20 background of unchanged polynucleotides in the library that increases the difficulty of detecting recombinant molecules. This problem becomes increasingly severe as the parental templates become more heterogeneous, that is, as the percentage of sequence identity between the
25 parental templates decreases. This outcome was demonstrated by Kikuchi, et al., (Gene 243:133-137, 2000) who attempted to generate recombinants between *xylE* and *nahH* using the methods of family shuffling reported by Patten et al., 1997; Crameri et al., 1998; Harayama,
30 1998; Kumamaru et al., 1998; Chang et al., 1999; Hansson et al., 1999). Kikuchi, et al., found that essentially no recombinants (<1%) were generated. They also disclosed a method to improve the formation of chimeric genes by fragmentation and reassembly of single-stranded

DNAs. Using this method, they obtained chimeric genes at a rate of 14 percent, with the other 86 percent being parental sequences.

5 The characteristic of low-efficiency recovery of recombinants limits the utility of these methods for generating novel polynucleotides from parental templates with a lower percentage of sequence identity, that is, parental templates that are more diverse.

10

 Accordingly, there is a need for a method of generating gene sequences that addresses these needs. A method has been developed for reassorting mutations among related polynucleotides, *in vitro*, by forming
15 heteroduplex molecules and then addressing the mismatches such that sequence information at sites of mismatch is transferred from one strand to the other. The mismatches are addressed by incubating the heteroduplex molecules in a reaction containing a) an enzyme that recognizes and
20 nicks a sequence strand at a mismatch site, b) a polymerase with a proofreading activity in the presence of dNTPs, and c) a ligase. These respective activities act in concert such that, at a given site of mismatch, the heteroduplex is nicked, unpaired bases are excised
25 from one of the strands, then replaced using the opposite strand as a template, and nicks are sealed. Output polynucleotides may be amplified before cloning, or cloned directly and tested for improved properties. Additional cycles of mismatch resolution reassortment and
30 testing may lead to further improvement.

 This method utilizes a mismatch endonuclease that is capable of recognizing and nicking at the site of a

mismatch between a base or a sequence of bases along opposite strands of a nucleic acid sequence.

To address the need for enzymes that will recognize
5 a mismatch, we have cloned the gene for the CEL I enzyme
and a novel enzyme we refer to as RES I. Both of these
enzymes are mismatch endonucleases and both are
particularly suited to recognizing a base pair mismatch
along a nucleic acid sequence, such as a chromosome, a
10 plasmid, a gene, a portion of a gene or any artificial
sequence of nucleic acids.

SUMMARY OF THE INVENTION

15 The present invention provides enzymes CEL I and RES
I that are capable of detecting a base mismatch in a
nucleotide sequence, and nicking the sequence at the site
of the base mismatch. These enzymes can be obtained from
the native plants in which they occur. Methods are
20 provided herein to make them as recombinant enzymes in a
host organism by cloning into the host organism.

CEL I and RES I have utility in gene shuffling
technology for developing new genes, and single
25 nucleotide polymorphism (SNP) detection for e.g.
detecting evidence of cancer susceptibility.

Also provided is a method of cloning *CellI* and *ResI*
genes into a plasmid or viral vector for transfer to a
30 host organism that will amplify the number of copies of
the gene or make the enzyme encoded in the plasmid or
viral vector.

Also provided is a method making CEL I and RES I enzymes a plant host.

The *CelI* and *ResI* nucleic acid sequences have been
5 deposited at the ATCC under the Budapest Treaty.

Also provided is a complete description of a method by which CEL I and RES I are used for reassorting mutations among related polynucleotides, *in vitro*, by
10 forming heteroduplex molecules and then addressing the mismatches such that sequence information at sites of mismatch is transferred from one strand to the other. The mismatches are addressed by incubating the heteroduplex molecules in a reaction containing CEL I or
15 RES I mismatch endonuclease enzyme that recognizes a mismatch and nicks one of the strands at the mismatch site, a polymerase with a proofreading activity in the presence of dNTPs, and a ligase. These respective activities act in concert such that, at a given site of
20 mismatch, the heteroduplex is nicked, unpaired bases are excised from one of the strands, then replaced using the opposite strand as a template, and nicks are sealed. Output polynucleotides may be amplified before cloning, or cloned directly and tested for improved properties.
25 Additional cycles of mismatch resolution reassortment and testing may lead to further improvement.

The method is also described as a method of increasing the number of complementary base pairs in a
30 heteroduplex polynucleotide sequence where said heteroduplex polynucleotide sequence has at least two non-complementary nucleotide base pairs, the method includes mixing the heteroduplex polynucleotide sequence with an effective amount of mismatch-directed strand

cleavage activity provided by CEL I or RES I enzyme,
proofreading activity, and ligase activity; and allowing
sufficient time for a number of non-complementary
nucleotide base pairs to be converted to complementary
5 base pairs, wherein the homogeneity between the strands
is increased by at least one complementary base pair.

The method is also described as an *in vitro* method
of making a population of sequence variants from a
10 heteroduplex polynucleotide sequence wherein said
heteroduplex polynucleotide sequence has at least two
non-complementary nucleotide base pairs, said method
includes mixing copies of the heteroduplex polynucleotide
sequence with an effective amount of mismatch-directed
15 strand cleavage activity provided by CEL I or RES I
enzyme, proofreading activity, and ligase activity; and
allowing sufficient time for a number of non-
complementary nucleotide base pairs to be converted to
complementary base pairs, wherein a diverse population of
20 polynucleotide sequences results.

The method is also described as an *in vitro* method
of obtaining a polynucleotide sequence encoding a desired
functional property, includes preparing at least one
25 heteroduplex polynucleotide sequence; mixing copies of
the heteroduplex polynucleotide sequence with an
effective amount of mismatch-directed strand cleavage
activity provided by CEL I or RES I enzyme, proofreading
activity, and ligase activity; and allowing sufficient
30 time for the percentage of complementarity between
strands of the heteroduplex polynucleotide sequence to
increase, wherein sequence diversity in the population is
increased; and screening or selecting a population of
variants for the desired functional property.

The method is also described as an *in vitro* method of obtaining a polynucleotide encoding a desired functional property, includes preparing at least one
5 heteroduplex polynucleotide, mixing copies of the heteroduplex polynucleotide sequence with an effective amount of mismatch-directed strand cleavage activity provided by CEL I or RES I enzyme, proofreading activity, and ligase activity, allowing sufficient time for some or
10 all of the mismatched nucleotide base pairs in the heteroduplex polynucleotide sequence to be converted to complementary bases, wherein a diverse population of polynucleotide sequences results, screening or selecting for a population of variants having a desired functional
15 property, denaturing said population of variants to obtain a population of single stranded polynucleotide sequences, annealing said population of single stranded polynucleotide sequences to form a diverse population heteroduplex polynucleotide sequences, mixing the
20 heteroduplex polynucleotide sequence with an effective amount of mismatch-directed strand cleavage activity, proofreading activity, and ligase activity, allowing sufficient time for some or all of the mismatched nucleotide base pairs in the heteroduplex polynucleotide
25 sequence to be converted to matched base pairs, wherein a diverse population of polynucleotide sequences results, and screening or selecting for a population of variants having a desired functional property. DNA can be converted to RNA prior to screening by transcription of
30 the DNA. A ligase activity can be added to seal the strands after proofreading.

One of the advantages of this method is that the sequence is either circular or linear. This allows for

shuffling of nearly unlimited sequence length. The variant polynucleotide sequences have different amounts of complementarity. We report increasing the complementarity in a polynucleotide heteroduplex between
5 two polynucleotides with sequence homology as low as 47%.

This process can occur simultaneously at many sites and on either strand of a given heteroduplex DNA molecule. The result is a randomization of sequence
10 differences among input strands to give a population of sequence variants that is more diverse than the population of starting sequences.

In another embodiment, a method of identifying a
15 reassorted DNA molecule encoding a protein with a desired functional property, includes providing at least one single-stranded uracil-containing DNA molecule, which single-stranded uracil-containing DNA molecule, or a complementary strand thereto, encodes a protein;
20 providing one or a plurality of non-identical single-stranded DNA molecules capable of hybridizing to the single-stranded uracil-containing DNA molecule, wherein said DNA molecules encode at least one additional variant of the protein; contacting the single-stranded uracil-
25 containing DNA molecule with at least one single-stranded DNA molecule of step (b), thereby producing an annealed DNA molecule; incubating the annealed DNA molecule with a mismatch endonuclease, proofreading polymerase and a ligase, thereby producing a sequence-reassorted DNA
30 strand annealed to the uracil-containing DNA molecule; amplifying the reassorted DNA strand under conditions wherein the uracil-containing DNA molecule is not amplified, thereby producing a population of reassorted DNA molecules; and, screening or selecting the population

of reassorted DNA molecules to identify those that encode a polypeptide having the desired functional property, thereby identifying one or more DNA molecules(s) that encode a polypeptide with the desired functional
5 property. This process can also occur using an RNA molecule as a template.

In one embodiment, *CellI* is a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO:01, or
10 SEQ ID NO:02, or SEQ ID NO:03, or SEQ ID NO:04.

In another embodiment, a recombinant virus comprising a nucleic acid sequence of *CellI* as represented by SEQ ID NO:01, or SEQ ID NO:02, or SEQ ID NO:03, or SEQ
15 ID NO:04 is presented.

In another embodiment, a recombinant plasmid comprising a nucleic acid sequence of *CellI* as represented by SEQ ID NO:01, or SEQ ID NO:02, or SEQ ID NO:03, or SEQ
20 ID NO:04 is presented.

In another embodiment, a plant or plant cell comprising a recombinant virus encoding a nucleic acid sequence of *CellI* as represented by SEQ ID NO:01, or SEQ
25 ID NO:02, or SEQ ID NO:03, or SEQ ID NO:04 is presented.

In another embodiment, a plant or plant cell comprising a recombinant plasmid encoding a nucleic acid sequence of *CellI* as represented by SEQ ID NO:01, or SEQ
30 ID NO:02, or SEQ ID NO:03, or SEQ ID NO:04 is presented.

In another embodiment, the recombinant virus can be a plant virus, an animal virus, a fungus virus, or a bacterial virus.

In another embodiment method of expressing CEL I endonuclease using a recombinant virus is presented.

- 5 In another embodiment, a method of using CEL I in an *in vitro* method of making sequence variants from at least one heteroduplex polynucleotide where said heteroduplex has at least two non-complementary nucleotide base pairs, said method comprising:
- 10 a. preparing at least one heteroduplex polynucleotide;
- b. combining said heteroduplex polynucleotide with an effective amount of CEL I, T4 DNA polymerase, and T4 DNA ligase; and
- 15 allowing sufficient time for the percentage of complementarity to increase, wherein one or more variants are made.

In one embodiment, *ResI* is a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO:16. The RES I amino acid sequence of SEQ ID NO:34 is presented.

In another embodiment, a recombinant virus comprising a nucleic acid sequence of *ResI* as represented by SEQ ID NO:16.

In another embodiment, a recombinant plasmid comprising a nucleic acid sequence of *ResI* as represented by SEQ ID NO:16.

30

In another embodiment, a plant or plant cell comprising a recombinant virus encoding a nucleic acid sequence of *ResI* as represented by SEQ ID NO:16.

In another embodiment, a plant or plant cell comprising a recombinant plasmid encoding a nucleic acid sequence of *ResI* as represented by SEQ ID NO:16.

5

In another embodiment, the recombinant virus can be a plant virus, an animal virus, a fungus virus, or a bacterial virus.

10 In another embodiment method of expressing RES I endonuclease using a recombinant virus is presented.

In another embodiment, a method of using RES I in an *in vitro* method of making sequence variants from at least one heteroduplex polynucleotide where said heteroduplex has at least two non-complementary nucleotide base pairs, said method comprising:

- a. preparing at least one heteroduplex polynucleotide;
- 20 b. combining said heteroduplex polynucleotide with an effective amount of RES I, T4 DNA polymerase, and T4 DNA ligase; and
- c. allowing sufficient time for the percentage of complementarity to increase,
- 25 wherein one or more variants are made.

BRIEF DESCRIPTION OF THE FIGURES

30 Figure 1 depicts the process of Genetic Reassortment by Mismatch Resolution (GRAMMR). Reassortment is contemplated between two hypothetical polynucleotides differing at least two nucleotide positions. Annealing between the top strand of A and the bottom strand of B is

shown which results in mismatches at the two positions.
After the process of reassortment mismatch resolution,
four distinct product polynucleotides are seen, the
parental types A and B, and the reassorted products X and
5 Y.

Figure 2 depicts an exemplary partially
complementary nucleic acid population of two molecules.
Figure 2A shows the sequence of two nucleic acid
10 molecules "X" and "Y" having completely complementary
top/bottom strands 1+/2- and 3+/4-, respectively. The
positions of differing nucleotides between the nucleic
acids X and Y are indicated (*). Figure 2B shows
possible combinations of single strands derived from
15 nucleic acids X and Y after denaturing and annealing and
indicates which of those combinations would comprise a
partially complementary nucleic acid population of two.

Figure 3 shows nucleic acid sequence encoding RES I
20 endonuclease (SEQ ID NO: 16) as taught in EXAMPLE 8.

Figure 4 shows the corresponding amino acid sequence
for RES I (SEQ ID NO: 34).

25 Figure 5 shows the nucleic acid sequence for plasmid
pBSC3BFP (SEQ ID NO: 32) as taught in EXAMPLE 9.

Figure 6 shows the nucleic acid sequence for the
tobamovirus movement protein open reading frame of TMV-Cg
30 (SEQ ID NO: 18) as taught in Example 15.

Figure 7 shows the nucleic acid sequence for the
tobamovirus movement protein open reading frame of TMV-Ob
(SEQ ID NO: 19) as taught in Example 15.

Figure 8 shows the nucleic acid sequence for the tobamovirus movement protein open reading frame of TMV-U2 (SEQ ID NO: 20) as taught in Example 15.

5

Figure 9 shows a resultant clone from TMV-Cg and ToMV GRAMMR reaction (SEQ ID NO: 21) as taught in Example 15.

10 Figure 10 shows a second resultant clone from a TMV-Cg and ToMV GRAMMR reaction (SEQ ID NO: 22) as taught in Example 15.

15 Figure 11 shows a resultant clone from a TMV-Ob and ToMV GRAMMR reaction (SEQ ID NO: 23) as taught in Example 15.

20 Figure 12 shows a second resultant clone from a TMV-Ob and ToMV GRAMMR reaction (SEQ ID NO: 24) as taught in Example 15.

25 Figure 13 shows a resultant clone from a TMV-U2 and ToMV GRAMMR reaction (SEQ ID NO: 25) as taught in Example 15.

25

Figure 14 shows a second resultant clone from a TMV-U2 and ToMV GRAMMR reaction (SEQ ID NO: 26) as taught in Example 15.

30 Figure 15 shows a resultant clone from a TMV-U1 and ToMV GRAMMR reaction (SEQ ID NO: 27) as taught in Example 15.

Figure 16 shows a second resultant clone from a TMV-U1 and ToMV GRAMMR reaction (SEQ ID NO: 28) as taught in Example 15.

5 Figure [17] shows the nucleic acid sequence for the tobamovirus movement protein open reading frame of TMV (SEQ ID NO: 9) as taught in Example 15.

10 Figure [18] shows the nucleic acid sequence for the tobamovirus movement protein open reading frame of ToMV (SEQ ID NO: 10) as taught in Example 15.

Definitions

15 In order to provide a clear and consistent understanding of the specification and the claims, including the scope given herein to such terms, the following definitions are provided:

20 As used herein, the term "amplification" refers to a process where the number of copies of a polynucleotide is increased.

25 As used herein, the term "annealing" refers to the formation of at least partially double stranded nucleic acid by hybridization of at least partially complementary nucleotide sequences. A partially double stranded nucleic acid can be due to the hybridization of a smaller nucleic acid strand to a longer nucleic acid strand,
30 where the smaller nucleic acid is 100% identical to a portion of the larger nucleic acid. A partially double stranded nucleic acid can also be due to the hybridization of two nucleic acid strands that do not share 100% identity but have sufficient homology to

hybridize under a particular set of hybridization conditions.

As used herein, the term "clamp" refers to a unique
5 nucleotide sequence added to one end of a polynucleotide,
such as by incorporation of the clamp sequence into a PCR
primer. The clamp sequences are intended to allow
amplification only of polynucleotides that arise from
hybridization of strands from different parents (*i.e.*,
10 heteroduplex molecules) thereby ensuring the production
of full-length hybrid products as described previously
(Skarfstad, J. Bact, vol 182, No 11, P. 3008-3016).

As used herein, the term "cleaving" means digesting
15 the polynucleotide with enzymes or otherwise breaking
phosphodiester bonds within the polynucleotide. As used
herein, the term "strand cleavage activity" or "cleavage"
refers to the breaking of a phosphodiester bond in the
backbone of the polynucleotide strand, as in forming a
20 nick. Strand cleavage activity can be provided by an
enzymatic agent. Such agents include CEL I, RES I and
their variants.

As used herein, the term "complementary base pair"
25 refers to the correspondence of DNA (or RNA) bases in the
double helix such that adenine in one strand is opposite
thymine (or uracil) in the other strand and cytosine in
one strand is opposite guanine in the other.

30 As used herein, the term "complementary to" is used
herein to mean that the complementary sequence is
identical to the reverse-complement of all or a portion
of a reference polynucleotide sequence or that each
nucleotide in one strand is able to form a base-pair with

a nucleotide, or analog thereof in the opposite strand. For illustration, the nucleotide sequence "TATAC" is complementary to a reference sequence "GTATA".

5 As used herein, the term "denaturing" or "denatured," when used in reference to nucleic acids, refers to the conversion of a double stranded nucleic acid to a single stranded nucleic acid. Methods of denaturing double stranded nucleic acids are well known
10 to those skilled in the art, and include, for example, addition of agents that destabilize base-pairing, increasing temperature, decreasing salt, or combinations thereof. These factors are applied according to the complementarity of the strands, that is, whether the
15 strands are 100% complementary or have one or more non-complementary nucleotides.

 As used herein, the term "desired functional property" means a phenotypic property, which include but
20 are not limited to, encoding a polypeptide, promoting transcription of linked polynucleotides, binding a protein, improving the function of a viral vector, and the like, which can be selected or screened for. Polynucleotides with such desired functional properties,
25 can be used in a number of ways, which include but are not limited to expression from a suitable plant, animal, fungal, yeast, or bacterial expression vector, integration to form a transgenic plant, animal or microorganism, function of a ribozyme, and the like.

30

 As used herein, the term "DNA shuffling" is used herein to indicate reassortment of sequence information between substantially homologous but non-identical sequences.

As used herein, the term "effective amount" refers to the amount of an agent necessary for the agent to provide its desired activity. For the present invention, this determination is well within the knowledge of those of ordinary skill in the art.

As used herein, the term "host" refers to a cell, tissue or organism capable of replicating a vector or plant viral nucleic acid and which is capable of being infected by a virus containing the viral vector or plant viral nucleic acid. This term is intended to include prokaryotic and eukaryotic cells, organs, tissues or organisms, where appropriate.

As used herein, the term "infection" refers to the ability of a virus to transfer its nucleic acid to a host or introduce viral nucleic acid into a host, wherein the viral nucleic acid is replicated, viral proteins are synthesized, and new viral particles assembled. In this context, the terms "transmissible" and "infective" are used interchangeably herein.

As used herein, the term "non-native" refers to any RNA sequence that promotes production of subgenomic mRNA including, but not limited to, 1) plant viral promoters such as ORSV and brome mosaic virus, 2) viral promoters from other organisms such as human Sindbis viral promoter, and 3) synthetic promoters.

As used herein, the term "phenotypic trait" refers to an observable property resulting from the expression of a gene.

As used herein, the term "plant cell" refers to the structural and physiological unit of plants, consisting of a protoplast and the cell wall.

5

As used herein, the term "plant organ" refers to a distinct and visibly differentiated part of a plant, such as root, stem, leaf or embryo.

10

As used herein, the term "plant tissue" refers to any tissue of a plant *in planta* or in culture. This term is intended to include a whole plant, plant cell, plant organ, protoplast, cell culture, or any group of plant cells organized into a structural and functional unit.

15

As used herein, the term "production cell" refers to a cell, tissue or organism capable of replicating a vector or a viral vector, but which is not necessarily a host to the virus. This term is intended to include prokaryotic and eukaryotic cells, organs, tissues or organisms, such as bacteria, yeast, fungus and plant tissue.

20

As used herein, the term "promoter" refers to the 5'-flanking, non-coding sequence adjacent a coding sequence which is involved in the initiation of transcription of the coding sequence.

25

As used herein, the term "protoplast" refers to an isolated plant cell without cell walls, having the potency for regeneration into cell culture or a whole plant.

30

As used herein, the term "recombinant plant viral nucleic acid" refers to plant viral nucleic acid, which has been modified to contain non-native nucleic acid sequences.

5

As used herein, the term "recombinant plant virus" refers to a plant virus containing the recombinant plant viral nucleic acid.

10

As used herein, the term "subgenomic promoter" refers to a promoter of a subgenomic mRNA of a viral nucleic acid.

As used herein, the term "substantial sequence
15 homology" refers to nucleotide sequences that are substantially functionally equivalent to one another. Nucleotide differences between such sequences having substantial sequence homology will be de minimus in affecting function of the gene products or an RNA coded
20 for by such sequence.

25

As used herein, the term "transcription" refers to production of an RNA molecule by RNA polymerase as a complementary copy of a DNA sequence.

As used herein, the term "vector" refers to a self-replicating DNA molecule which transfers a DNA segment between cells.

30

As used herein, the term "virus" refers to an infectious agent composed of a nucleic acid encapsidated in a protein. A virus may be a mono-, di-, tri- or multi-partite virus, as described above.

As used herein, the term "Genetic Reassortment by Mismatch Resolution (GRAMMR)" refers to a method for reassorting sequence variations among related polynucleotides by an *in vitro* method of redistributing sequence variations between non-identical polynucleotide sequences, by making a heteroduplex polynucleotide from two non-identical polynucleotides; introducing a nick in one strand at or near a base pair mismatch site; removing mismatched base(s) from the mismatch site where the nick occurred; and using the opposite strand as template to replace the removed base(s) with bases that complement base(s) in the first strand. By this method, information is transferred from one strand to the other at sites of mismatch.

Multiple sites in a partially complementary molecule can be addressed independently and simultaneously in this process. The result is an increase in the percentage of complementary base pairs in the polynucleotide sequence.

One or more base pair mismatches between two strands of the heteroduplex polynucleotide sequence are resolved by an *in vitro* method of mixing the heteroduplex polynucleotide sequence with an effective amount of mismatch-directed strand cleavage activity provided by CEL I or RES I enzyme, proofreading activity, and ligase activity to resolve one or more of the mismatches. By this method, information is transferred from one strand to the other at sites of a mismatch.

A mismatch can be the result of two non-complementary bases occurring opposite each other. A

mismatch site can consist of a cluster of any number of unpaired nucleotides, including nucleotide base-pairs that are made unstable by neighboring mismatches. A mismatch can also be the result of one or more bases occurring on one strand that do not have a numerical opposite on the opposite strand. For example, at the site of a mismatch there might be 1 unpaired base on one strand and no unpaired bases on the other strand. This would result in a site of sequence length heterogeneity in which a single unpaired nucleotide is contained in one strand at that site. Depending on the strand that is initially nicked at this site of mismatch, the process of this invention would result in either the insertion of a single base relative to the shorter strand, or in the deletion of a single base relative to the strand that originally had the extra unpaired nucleotide. This principle of transfer of sequence length information from one strand to the other can apply to any site of mismatch where the number of mismatched bases on the two strands do not equal one-another.

Usually many copies of the heteroduplex polynucleotide are present in the reaction. In this situation, sequence information at a mismatch site might be templated from the top strand on one copy of the polynucleotide and from the bottom strand in another copy. Assuming a sufficient number of copies are available, if a single mismatch is present, then two output variants are possible. If two mismatch sites are present then 2 times 2 variants can result. If n mismatch sites are present, then at least 2 to the n power or 2^n genetic reassortments are possible by mismatch resolution. The possible result is at least 2^n variant polynucleotides. We say at least, because the exact

mechanism is not fully understood. It can be speculated that for a mismatch site that is 2 or more bases in length, an individual event might template 1, 2 or more of the mismatched bases. If that is the case, then the
5 result would be an increase in the probable number of variants.

As used herein, the term "GENEWARE" or "GENEWARE®" refers to a viral vector derived at least in part from a
10 Tobamovirus and modified to contain an additional (usually heterologous) subgenomic promoter. A Tobamovirus found in nature, typically contains subgenomic promoters for the movement protein and the coat protein. GENEWARE® is a registered trademark of
15 Large Scale Biology corporation.

As used herein, the term "granularity" refers to the amount of a nucleic acid's sequence information from a given parental polynucleotide sequence that occurs as a
20 contiguous sequence in a given progeny polynucleotide.

As used herein, the term "template sequence" refers to a first single stranded polynucleotide sequence that is partially complementary to a second polynucleotide
25 sequence such that treatment by GRAMMR results in transfer of genetic information from the template strand to the second strand.

The larger the units of sequence information transferred from a template strand, the higher the
30 granularity. The smaller the blocks of sequence information transferred from the template strand, the lower or finer the granularity. Lower granularity indicates that a DNA shuffling or reassortment method is

able to transfer smaller discrete blocks of genetic information from the template strand to the second strand. The advantage of a DNA shuffling or reassortment method with lower granularity is that it is able to
5 resolve smaller nucleic acid sequences from others, and to transfer the sequence information. DNA shuffling or reassortment methods that return primarily high granularity are not readily able to resolve smaller nucleic acid sequences from others.

10

As used herein, the term "heteroduplex polynucleotide" refers to a double stranded polynucleotide formed by annealing single strands, typically separate strands, where the strands are non-
15 identical. A heteroduplex polynucleotide may have unpaired regions existing as single strand loops or bubbles. A heteroduplex polynucleotide region can also be formed by one single-strand polynucleotide wherein partial self-complementarity allows the formation of a
20 stem-loop structure where the annealing portion of the strand is non-identical.

As used herein, the term "heteroduplex DNA" refers to a double-stranded DNA formed by annealing single
25 strands, typically separate strands), where the strands are non-identical. A heteroduplex DNA may have unpaired regions existing as single strand loops or bubbles. A heteroduplex DNA region can also be formed by one single-strand polynucleotide wherein partial self-
30 complementarity allows the formation of a stem-loop structure where the annealing portion of the strand is non-identical.

As used herein, the term "homologous" means that one single-stranded nucleic acid sequence may hybridize to an at least partially complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on
5 a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentrations as discussed later.

10 Nucleic acids are "homologous" when they are derived, naturally or artificially, from a common ancestor sequence. During natural evolution, this occurs when two or more descendent sequences diverge from a parent sequence over time, *i.e.*, due to mutation and
15 natural selection. Under artificial conditions, divergence occurs, *e.g.*, in one of two basic ways. First, a given sequence can be artificially recombined with another sequence, as occurs, *e.g.*, during typical cloning, to produce a descendent nucleic acid, or a given
20 sequence can be chemically modified, or otherwise manipulated to modify the resulting molecule. Alternatively, a nucleic acid can be synthesized *de novo*, by synthesizing a nucleic acid that varies in sequence from a selected parental nucleic acid sequence. When
25 there is no explicit knowledge about the ancestry of two nucleic acids, homology is typically inferred by sequence comparison between two sequences. Where two nucleic acid sequences show sequence similarity over a significant portion of each of the nucleic acids, it is inferred that
30 the two nucleic acids share a common ancestor. The precise level of sequence similarity that establishes homology varies in the art depending on a variety of factors.

For purposes of this disclosure, two nucleic acids are considered homologous where they share sufficient sequence identity to allow GRAMMR-mediated information transfer to occur between the two nucleic acid molecules.

5

As used herein, the term "identical" or "identity" means that two nucleic acid sequences have the same sequence or a complementary sequence. Thus, "areas of identity" means that regions or areas of a polynucleotide or the overall polynucleotide are identical or complementary to areas of another polynucleotide.

As used herein, the term "increase in percent complementarity" means that the percentage of complementary base-pairs in a heteroduplex molecule is made larger.

As used herein, the term "ligase" refers to an enzyme that establishes a phosphodiester bond between adjacent nucleotides in a nucleic acid.

As used herein, the term "mismatch" refers to a base-pair that is unable to form normal base-pairing interactions (*i.e.*, other than "A" with "T" (or "U"), or "G" with "C").

As used herein, the term "mismatch resolution" refers to the conversion of a mismatched base-pair into a complementary base-pair.

30

As used herein, the term "mutations" means changes in the sequence of a wild-type or reference nucleic acid sequence or changes in the sequence of a polypeptide. Such mutations can be point mutations such as transitions

or transversions. The mutations can be deletions, insertions or duplications.

As used herein, the term "nucleic acid" or "nucleic acid molecule" means a polynucleotide such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) and encompasses single-stranded and double-stranded nucleic acid as well as an oligonucleotide. Nucleic acids useful in the invention include genomic DNA, cDNA, mRNA, plasmids, cosmids, PCR products, and synthetic oligonucleotides, and can represent the sense strand, the anti-sense strand, or both. A nucleic acid generally incorporates the four naturally occurring nucleotides adenine, guanine, cytosine, and thymidine/uridine. An invention nucleic acid can also incorporate other naturally occurring or non-naturally occurring nucleotides, including derivatives thereof, so long as the nucleotide derivatives can be incorporated into a polynucleotide by a polymerase at an efficiency sufficient to generate a desired polynucleotide product.

As used herein, the term a "parental nucleic acid" refers to a double stranded nucleic acid having a sequence that is 100% identical to an original single stranded nucleic acid in a starting population of partially complementary nucleic acids. Parental nucleic acids would include, for example in the illustration of Figure 2, nucleic acids X and Y if partially complementary nucleic acid combinations 1+/4- or 2-/3+ were used as a starting population in an invention method.

As used herein, the term "partially complementary" refers to a nucleic acid having a substantially

complementary sequence to another nucleic acid but that differs from the other nucleic acid by at least two or more nucleotides.

5 As used herein, the term "partially complementary nucleic acid population" refers to a population of nucleic acids comprising individual groups of nucleic acids having substantially complementary sequences but no nucleic acids belonging to a particular group having an
10 exact complementary sequence for any other group of sequences in the population.

 As used herein, any member of a partially complementary nucleic acid population differs from
15 another nucleic acid of the population, or the complement thereto, by two or more nucleotides. As such, a partially complementary nucleic acid specifically excludes a population containing sequences that are exactly complementary, that is, a complementary sequence
20 that has 100% complementarity. Therefore, each member of such a partially complementary nucleic acid population differs from other members of the population by two or more nucleotides, including both strands. One strand is designated the top strand, and its complement is
25 designated the bottom strand.

 As used herein, the term "top" strand refers to a polynucleotide read in the 5' to 3' direction and the "bottom" its complement. It is understood that, while a
30 sequence is referred to as bottom or top strand, such a designation is intended to distinguish complementary strands since, in solution, there is no orientation that fixes a strand as a top or bottom strand.

For example, a population containing two nucleic acid members can be derived from two double stranded nucleic acids, with a potential of using any of the four strands to generate a single stranded partially complementary nucleic acid population. An example of potential combinations of strands of two nucleic acids that can be used to obtain a partially complementary nucleic acid population of the invention is shown in Figure 2. The two nucleic acid sequences that are potential members of a partially complementary nucleic acid population are designated "X" (AGATCAATTG) and "Y" (AGACCGATTG) (Figure 2A). The nucleic acid sequences differ at two positions (positions 4 and 6 indicated by "*"). The "top" strand of nucleic acids X and Y are designated "1+" and "3+," respectively, and the "bottom" strand of nucleic acids X and Y are designated "2-" and "4-," respectively.

Figure 2B shows the possible combinations of the four nucleic acid strands. Of the six possible strand combinations, only the combination of 1+/2-, 1+/4-, 2-/3+, or 3+/4- comprise the required top and bottom strand of a partially complementary nucleic acid population. Of these top/bottom sequence combinations, only 1+/4- or 2-/3+ comprise an example of a partially complementary nucleic acid population of two different molecules because only these combinations have complementary sequences that differ by at least one nucleotide. The remaining combinations, 1+/2- and 2+/4-, contain exactly complementary sequences and therefore do not comprise a partially complementary nucleic acid population of the invention.

In the above described example of a population of two different molecules, a partially complementary population of nucleic acid molecules excluded combinations of strands that differ by one or more nucleotides but which are the same sense, for example, 1+/3+ or 2-/4-. However, it is understood that such a combination of same stranded nucleic acids can be included in a larger population, so long as the population contains at least one bottom strand and at least one top strand. For example, if a third nucleic acid "Z," with strands 5+ and 6- is included, the combinations 1+/3+/6- or 2-/4-/5+ would comprise a partially complementary nucleic acid population. Similarly, any number of nucleic acids and their corresponding top and bottom strands can be combined to generate a partially complementary nucleic acid population of the invention so long as the population contains at least one top strand and at least one bottom strand and so long as the population contains no members that are the exact complement.

The populations of nucleic acids of the invention can be about 3 or more, about 4 or more, about 5 or more, about 6 or more, about 7 or more, about 8 or more, about 9 or more, about 10 or more, about 12 or more, about 15 or more, about 20 or more, about 25 or more about 30 or more, about 40 or more, about 50 or more, about 75 or more, about 100 or more, about 150 or more, about 200 or more, about 250 or more, about 300 or more, about 350 or more, about 400 or more, about 450 or more, about 500 or more, or even about 1000 or more different nucleic acid molecules. A population can also contain about 2000 or more, about 5000 or more, about 1×10^4 or more, about 1×10^5 or more, about 1×10^6 or more, about 1×10^7 or more, or even

about 1×10^8 or more different nucleic acids. One skilled in the art can readily determine a desirable population to include in invention methods depending on the nature of the desired reassortment experiment outcome and the
5 available screening methods, as disclosed herein.

As used herein, the term a "polymerase" refers to an enzyme that catalyzes the formation of polymers of nucleotides, that is, polynucleotides in a template-
10 directed fashion. A polymerase useful in the invention can be derived from any organism or source, including animal, plant, bacterial and viral polymerases. A polymerase can be a DNA polymerase, RNA polymerase, or a reverse transcriptase capable of transcribing RNA into
15 DNA.

As used herein, the term "proofreading" describes the property of an enzyme where a nucleotide, such as, a mismatched nucleotide, can be removed in a 3'-to-5' fashion and replaced by, typically, a base-paired
20 nucleotide. In the case of addressing a loop caused by insertion or deletion, proofreading may involve only removal of the mismatched nucleotide(s) or only addition of base-paired nucleotide(s).

25

As used herein, the term a "recombinant" polynucleotide refers to a polynucleotide that comprises sequence information from at least two different polynucleotides.

30

As used herein, the term "related polynucleotides" means that regions or areas of the polynucleotides are identical and regions or areas of the polynucleotides are non-identical.

As used herein, the term DNA "reassortment" is used herein to indicate a redistribution of sequence variations between non-identical sequences.

5

As used herein, the term "replicon" refers to a genetic unit of replication including a length of polynucleotide and its site for initiation of replication.

10

As used herein, the term "sequence diversity" refers to the abundance of non-identical polynucleotides. The term "increasing sequence diversity in a population" means to increase the relative abundance of non-identical polynucleotides in a population.

15

As used herein, the term "sequence variant" refers to a molecule (DNA, RNA polypeptide, and the like) with one or more sequence differences compared to a reference molecule. For example, the sum of the separate independent mismatch resolution events that occur throughout the heteroduplex molecule during the GRAMMR process results in reassortment of sequence information throughout that molecule. The sequence information will reassort in a variety of combinations to generate a complex library of "sequence variants".

20

25

As used herein, the term "Mismatch-directed strand cleavage" means strand cleavage activity by an agent that recognizes a site of a mismatched base pair, group of mismatched base pairs, or extrahelical base or bases on a heteroduplex polynucleotide sequence and cleaves one strand at the site of the mismatch.

30

As used herein, the term "sufficient time" refers to the period of time necessary for a reaction or process to render a desired product. For the present invention, the determination of sufficient time is well within the knowledge of those of ordinary skill in the art. It is noted that "sufficient time" can vary widely, depending on the desires of the practitioner, without impacting on the functionality of the reaction, or the quality of the desired product.

10

As used herein, the term "wild-type" means that a nucleic acid fragment does not contain any mutations. A "wild-type" protein means that the protein will be active at a level of activity found in nature and typically will be the amino acid sequence found in nature. In an aspect, the term "wild type" or "parental sequence" can indicate a starting or reference sequence prior to a manipulation of the invention.

In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right-hand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified otherwise, the left-hand end of single-stranded polynucleotide sequences is the 5' end; the left-hand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction.

30

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides nucleic acid molecules comprising a nucleic acid sequence which

include SEQ ID NO:01, SEQ ID NO:02, SEQ ID NO:03, or SEQ ID NO:04, useful as vectors or plasmids for the expression of CEL I endonuclease. The nucleic acid molecules of SEQ ID NO:03, and SEQ ID NO:04 are CEL I
5 open reading frames contained within SEQ ID NO:01 and SEQ ID NO:02, respectively. The preparation and use of the nucleic acid molecules of SEQ ID NO:01, SEQ ID NO:02, SEQ ID NO:03 and SEQ ID NO:04, are further taught in EXAMPLE 2 herein. The present invention also provides nucleic
10 acid molecules comprising the nucleic acid sequence of Figure 3 (SEQ ID NO:16), useful as vectors or plasmids for the expression of RES I endonuclease.

Mismatch endonucleases from plants have the ability to detect mismatches between hybridized nucleic acid
15 strands, detect polynucleotide loops and insertions between such hybridized strands, detect polymorphisms between such hybridized strands, detect sequence differences in polynucleotide strands, and detect such mutations in a target polynucleotide sequence without
20 substantial adverse effects of flanking DNA sequences (Yeung, Anthony, T. and Patrick J. Hagan International Patent Application No. WO 97/46701). These endonucleases are found in a wide variety of plants and are probably found throughout the plant kingdom. They have been found
25 in Alfalfa sprouts, Asparagus, mung bean shoots, celery stalks, Cha ha, iceberg lettuce, parsley, celery-cabbage, broccoli tops, cabbage, cauliflower tops, and tomatoes (Oleykowski, C.A. et al. Nucleic Acids Research, 1998, Vol. 26, No. 20 4597-4702).

30

Endonucleases from plants can now be used in a process to evolve sets of nucleotide sequence variants from hybridized sequences. In the process, the mismatch endonuclease detects the mismatched bases, and nicks at

least one of the hybridized strands. A polymerase enzyme then replaces one or more mismatched bases with one or more bases that complement the bases on the opposite strand. The nick is then sealed by a ligase enzyme. If
5 two or more mismatch sites are present on the hybrid, then a population of variant sequences will result, depending on the number and position of the mismatches at which replacements are made, and the strand on which base replacement was made. In the evolution of sequence
10 variants it is not necessarily desirable for all mismatches on all copies of the sequence to be resolved. From a set of sequence variants, one or more variants that have preferred activity can be selected. RES I and CEL I exemplify mismatch endonucleases useful in this
15 process. Other plant derived mismatch endonucleases are also useful in a GRAMMR reaction.

The present invention also provides a process for expressing CEL I or RES I endonuclease using a
20 recombinant plant viral nucleic acid comprising of a nucleic acid sequence selected from the group consisting of SEQ ID NO:01, SEQ ID NO:02, SEQ ID NO:03, SEQ ID NO:04, or Figure 3 (SEQ ID NO:16).

25 The present invention also provides a recombinant plant viral nucleic acid comprising of at least one sub-genomic promoter capable of transcribing or expressing CEL I or RES I endonuclease in a plant cell, wherein the plant cell is a host cell, or production cell.

30

In another embodiment, a plant viral nucleic acid is provided in which the native coat protein coding sequence has been deleted from a viral nucleic acid, a non-native plant viral coat protein coding sequence and a non-native

promoter, preferably the subgenomic promoter of the non-native coat protein coding sequence, capable of expression in the plant host, packaging of the recombinant plant viral nucleic acid, and ensuring a systemic infection of the host by the recombinant plant viral nucleic acid, has been inserted. Alternatively, the coat protein gene may be inactivated by insertion of the non-native nucleic acid sequence within it, such that a fusion protein is produced. The recombinant plant viral nucleic acid may contain one or more additional non-native subgenomic promoters. Each non-native subgenomic promoter is capable of transcribing or expressing adjacent genes or nucleic acid sequences in the plant host and incapable of recombination with each other and with native subgenomic promoters. Non-native (foreign) nucleic acid sequences may be inserted adjacent the native plant viral subgenomic promoter or the native and a non-native plant viral subgenomic promoters if more than one nucleic acid sequence is included. The non-native nucleic acid sequences are transcribed or expressed in the host plant under control of the subgenomic promoter to produce the desired products.

In another embodiment, a recombinant plant viral nucleic acid is provided as in the first embodiment except that the native coat protein coding sequence is placed adjacent one of the non-native coat protein subgenomic promoters instead of a non-native coat protein coding sequence.

30

In yet another embodiment, a recombinant plant viral nucleic acid is provided in which the native coat protein gene is adjacent its subgenomic promoter and one or more non-native subgenomic promoters have been inserted into

the viral nucleic acid. The inserted non-native subgenomic promoters are capable of transcribing or expressing adjacent genes in a plant host and are incapable of recombination with each other and with native subgenomic promoters. Non-native nucleic acid sequences may be inserted adjacent the non-native subgenomic plant viral promoters such that said sequences are transcribed or expressed in the host plant under control of the subgenomic promoters to produce the desired product.

In another embodiment, a recombinant plant viral nucleic acid is provided as in the third embodiment except that the native coat protein coding sequence is replaced by a non-native coat protein coding sequence.

The viral vectors are encapsidated by the coat proteins encoded by the recombinant plant viral nucleic acid to produce a recombinant plant virus. The recombinant plant viral nucleic acid or recombinant plant virus is used to infect appropriate host plants. The recombinant plant viral nucleic acid is capable of replication in the host, systemic spread in the host, and transcription or expression of foreign gene(s) in the host to produce the desired product.

Another embodiment to the present invention is directed to recombinant plant viral nucleic acids and recombinant viruses which are stable for maintenance and transcription or expression of non-native (foreign) nucleic acid sequences and which are capable of systemically transcribing or expressing such foreign sequences in the host plant. More specifically, recombinant plant viral nucleic acids according to the

present invention comprise a native plant viral subgenomic promoter, at least one non-native plant viral subgenomic promoter, a plant viral coat protein coding sequence, and optionally, at least one non-native,
5 nucleic acid sequence.

The present invention provides for the infection of a plant host by a recombinant plant virus containing recombinant plant viral nucleic acid or by the
10 recombinant plant viral nucleic acid which contains one or more non-native nucleic acid sequences which are transcribed or expressed in the infected tissues of the plant host. The product of the coding sequences may be recovered from the plant or cause a phenotypic trait in
15 the plant.

The first step in achieving any of the features of the invention is to modify the nucleotide sequences of the plant viral nucleotide sequence by known conventional
20 techniques such that one or more non-native subgenomic promoters are inserted into the plant viral nucleic acid without destroying the biological function of the plant viral nucleic acid. The subgenomic promoters are capable of transcribing or expressing adjacent nucleic acid
25 sequences in a plant host infected by the recombinant plant viral nucleic acid or recombinant plant virus. The native coat protein coding sequence may be deleted in two embodiments, placed under the control of a non-native subgenomic promoter in a second embodiment, or retained
30 in a further embodiment. If it is deleted or otherwise inactivated, a non-native coat protein gene is inserted under control of one of the non-native subgenomic promoters, or optionally under control of the native coat protein gene subgenomic promoter. The non-native coat

protein is capable of encapsidating the recombinant plant viral nucleic acid to produce a recombinant plant virus. Thus, the recombinant plant viral nucleic acid contains a coat protein coding sequence, which may be native or a
5 nonnative coat protein coding sequence, under control of one of the native or non-native subgenomic promoters. The coat protein is involved in the systemic infection of the plant host.

10 Some of the viruses which meet this requirement, and are therefore suitable, include viruses from the tobacco mosaic virus group such as Tobacco Mosaic virus (TMV), Cowpea Mosaic virus (CMV), Alfalfa Mosaic virus (AMV), Cucumber Green Mottle Mosaic virus watermelon strain
15 (CGMMV-W) and Oat Mosaic virus (OMV) and viruses from the brome mosaic virus group such as Brome Mosaic virus (BMV), broad bean mottle virus and cowpea chlorotic mottle virus. Additional suitable viruses include Rice Necrosis virus (RNV), and geminiviruses such as tomato
20 golden mosaic virus (TGMV), Cassava latent virus (CLV) and maize streak virus (MSV).

Another embodiment of the present invention is a recombinant plant viral nucleic acid, which further
25 comprises one or more non-native nucleic acid sequences capable of being transcribed in the plant host. The non-native nucleic acid sequence is placed adjacent one or the non-native viral subgenomic promoters and/or the native coat protein gene promoter depending on the
30 particular embodiment used. The non-native nucleic acid is inserted by conventional techniques, or the non-native nucleic acid sequence can be inserted into or adjacent the native coat protein coding sequence such that a fusion protein is produced. The non-native nucleic acid

sequence, which is transcribed, may be transcribed as an RNA capable of regulating the expression of a phenotypic trait by an anti-sense mechanism. Alternatively, the non-native nucleic acid sequence in the recombinant plant viral nucleic acid may be transcribed and translated in the plant host, to produce a phenotypic trait. The non-native nucleic acid sequence(s) may also code for the expression of more than one phenotypic trait. The recombinant plant viral nucleic acid containing the non-native nucleic acid sequence is constructed using conventional techniques such that non-native nucleic acid sequence(s) are in proper orientation to whichever viral subgenomic promoter is utilized.

The present invention further provides a plant cell comprising a vector or plasmid comprising of a nucleic acid sequence selected from the group consisting of SEQ ID NO:01, SEQ ID NO:02, SEQ ID NO:03, SEQ ID NO:04, or Figure 3 (SEQ ID NO:16) where the plant cell is a host cell, or production cell.

Methods of making CEL I and RES I enzyme are not limited to plant viral vectors. These genes or their variants can be introduced into any known host organism by any of several methods known in the art today.

Methods for transforming various host cells are disclosed in Klein et al. "Transformation of microbes, plants and animals by particle bombardment", Bio/Technol. New York, N.Y., Nature Publishing Company, March 1992, v. 10 (3) pp. 286-291. Techniques for transforming a wide variety of higher plant species are well known and described in the technical, scientific, and patent

literature. See, for example, Weising et al., Ann. Rev. Genet. 22: 421-477 (1988).

For example, the DNA construct may be introduced
5 directly into the genomic DNA of the plant cell using
techniques such as *electroporation*, PEG-mediated
transfection, particle bombardment, silicon fiber
delivery, or microinjection of plant cell protoplasts or
embryogenic callus. See, e.g., Tomes, et al., Direct DNA
10 Transfer into Intact Plant Cells Via Microprojectile
Bombardment. pp.197-213 in Plant Cell, Tissue and Organ
Culture, Fundamental Methods. eds. O. L. Gamborg and G.
C. Phillips. Springer-Verlag Berlin Heidelberg New York,
1995. The introduction of DNA constructs using
15 polyethylene glycol precipitation is described in
Paszkowski et al., Embo J. 3: 2717-2722 (1984).
Electroporation techniques are described in Fromm et al.,
Proc. Natl. Acad. Sci. 82: 5824 (1985). Ballistic
transformation techniques are described in Klein et al.,
20 Nature 327: 70-73 (1987).

Alternatively, the DNA constructs may be combined
with suitable T-DNA flanking regions and introduced into
a conventional *Agrobacterium tumefaciens* host vector. The
25 virulence functions of the *Agrobacterium tumefaciens* host
will direct the insertion of the construct and adjacent
marker into the plant cell DNA when the cell is infected
by the bacteria. *Agrobacterium tumefaciens*-mediated
transformation techniques are well described in the
30 scientific literature. See, for example Horsch et al.,
Science 233: 496-498 (1984), and Fraley et al., Proc.
Natl. Acad. Sci. 80: 4803 (1983). For instance,
Agrobacterium transformation of maize is described in
U.S. Pat. No. 5,981,840. *Agrobacterium* transformation of

monocot is found in U.S. Pat. No. 5,591,616.

Agrobacterium transformation of soybeans is described in U.S. Pat. No. 5,563,055.

5 Other methods of transformation include (1)
Agrobacterium rhizogenes-mediated transformation (see,
e.g., Lichtenstein and Fuller In: Genetic Engineering,
vol. 6, PWJ Rigby, Ed., London, Academic Press, 1987; and
Lichtenstein, C. P., and Draper, J., In: DNA Cloning,
10 Vol. II, D. M. Glover, Ed., Oxford, IRI Press, 1985),
Application PCT/US87/02512 (WO 88/02405 published Apr. 7,
1988) describes the use of A. rhizogenes strain A4 and
its Ri plasmid along with A. tumefaciens vectors pARC8 or
pARC16 (2) liposome-mediated DNA uptake (see, e.g.,
15 Freeman et al., Plant Cell Physiol. 25: 1353, 1984), (3)
the vortexing method (see, e.g., Kindle, Proc. Natl.
Acad. Sci., USA 87: 1228, (1990).

DNA can also be introduced into plants by direct DNA
20 transfer into pollen as described by Zhou et al., Methods
in Enzymology, 101:433 (1983); D. Hess, Intern Rev.
Cytol., 107:367 (1987); Luo et al., Plant Mol. Biol.
Reporter, 6:165 (1988). Expression of polypeptide coding
nucleic acids can be obtained by injection of the DNA
25 into reproductive organs of a plant as described by Pena
et al., Nature, 325:274 (1987). DNA can also be injected
directly into the cells of immature embryos and the
rehydration of desiccated embryos as described by Neuhaus
et al., Theor. Appl. Genet., 75:30 (1987); and Benbrook
30 et al., in Proceedings Bio Expo 1986, Butterworth,
Stoneham, Mass., pp.27-54 (1986).

Transformed plant cells which are derived by any of
the above transformation techniques can be cultured to

regenerate a whole plant which possesses the transformed genotype. Such regeneration techniques often rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with a polynucleotide of the present invention. For transformation and regeneration of maize see, Gordon-Kamm et al., *The Plant Cell*, 2:603-618 (1990).

CEL I and RES I and their variants can also be made by other host or host/vector systems. A polynucleotide sequences may be inserted into a recombinant plasmid, virus or other vehicle known in the art that has been manipulated by insertion or incorporation of the genetic sequences. Such expression vectors contain a promoter a sequence that facilitates the efficient transcription of the inserted genetic sequence of the host. The expression vector typically contains an origin of replication, a promoter. Vectors suitable for use in the present invention include, but are not limited to the T7-based expression vector for expression in bacteria (Rosenberg, et al., *Gene* 56:125, 1987), the pMSXND expression vector for expression in mammalian cells (Lee and Nathans, *J Biol. Chem.*, 263:3521, 1988) and baculovirus-derived vectors for expression in insect cells. The DNA segment can be present in the vector operably linked to regulatory elements, for example, a promoter (e.g., T7, metallothionein I, or polyhedrin promoters).

Polynucleotide sequences can be expressed in either prokaryotes or eukaryotes. Hosts can include microbial, yeast, insect and mammalian organisms. Methods of expressing DNA sequences having eukaryotic or viral sequences in prokaryotes are well known in the art.

Biologically functional viral and plasmid DNA vectors capable of expression and replication in a host are known in the art. Such vectors are used to incorporate DNA sequences of the invention.

5

Methods which are well known to those skilled in the art can be used to construct expression vectors containing a CEL I or RES I coding sequence and appropriate transcriptional/translational control signals. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo recombination/genetic techniques. (See, for example, the techniques described in Maniatis et al., 1989 Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y.)

A variety of host-expression vector systems may be utilized to express a CEL I or RES I coding sequence. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing a CEL I or RES I coding sequence; yeast transformed with recombinant yeast expression vectors containing a CEL I or RES I coding sequence; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing a CEL I or RES I coding sequence; or animal cell systems infected with recombinant virus expression vectors (e.g., retroviruses, adenovirus, vaccinia virus) containing a CEL I or RES I coding sequence, or transformed animal cell systems engineered for stable expression.

Depending on the host/vector system utilized, any of a number of suitable transcription and translation

elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al., 1987, Methods in Enzymology 5 153:516-544). For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage lambda, plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used. When cloning in mammalian cell systems, promoters derived from the genome of mammalian 10 cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the retrovirus long terminal repeat; the adenovirus late promoter; the vaccinia virus 7.5K promoter) may be used. Promoters produced by recombinant DNA or synthetic techniques may also be used to provide 15 for transcription of the inserted RES I OR CEL I coding sequence.

In bacterial systems a number of expression vectors may be advantageously selected depending upon the use 20 intended for the expressed protein. For example, when large quantities of RES I OR CEL I are to be produced, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Those that are engineered to contain a 25 cleavage site to aid in recovering are preferred. Such vectors include but are not limited to the E. coli expression vector pUR278 (Ruther et al., EMBO J. 2:1791, 1983), in which a CEL I or RES I coding sequence may be ligated into the vector in frame with the lac Z coding 30 region so that a hybrid -lac Z protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like.

In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review see, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant et al., 1987, Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Eds. Wu & Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp. 516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684; and The Molecular Biology of the Yeast *Saccharomyces*, 1982, Eds. Strathem et al., Cold Spring Harbor Press, Vols. I and II. A constitutive yeast promoter such as ADH or LEU2 or an inducible promoter such as GAL may be used (Cloning in Yeast, Ch. 3, R. Rothstein In: DNA Cloning Vol. 11, A Practical Approach, Ed. D M Glover, 1986, IRL Press, Wash., D.C.). Alternatively, vectors may be used which promote integration of foreign DNA sequences into the yeast chromosome.

Eukaryotic systems, and preferably mammalian expression systems, allow for proper post-translational modifications of expressed mammalian proteins to occur. Eukaryotic cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, phosphorylation, and advantageously, secretion of the gene product may be used as host cells for the expression of RES I OR CEL I. Mammalian cell lines may be preferable. Such host cell lines may include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, -293, and WI38.

Mammalian cell systems that utilize recombinant viruses or viral elements to direct expression may be engineered. For example, when using adenovirus expression vectors, a CEL I or RES I coding sequence may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by in vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing the protein in infected hosts (e.g., see Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81: 3655-3659). Alternatively, the vaccinia virus 7.5K promoter can be used. (e.g., see, Mackett et al., 1982, Proc. Natl. Acad. Sci. USA 79: 7415-7419; Mackett et al., 1984, J. Virol. 49: 857-864; Panicali et al., 1982, Proc. Natl. Acad. Sci. USA 79: 4927-4931). Vectors based on bovine papilloma virus which have the ability to replicate as extrachromosomal elements (Sarver, et al., 1981, Mol. Cell. Biol. 1: 486). Shortly after entry of this DNA into mouse cells, the plasmid replicates to about 100 to 200 copies per cell. Transcription of the inserted cDNA does not require integration of the plasmid into the host's chromosome, thereby yielding a high level of expression. These vectors can be used for stable expression by including a selectable marker in the plasmid, such as, for example, the neo gene. Alternatively, the retroviral genome can be modified for use as a vector capable of introducing and directing the expression of a CEL I or RES I gene in host cells (Cone & Mulligan, 1984, Proc. Natl. Acad. Sci. USA 81:6349-6353). High level expression may also be achieved using inducible promoters, including, but not

limited to, the metallothionine IIA promoter and heat shock promoters.

Host cells can be transformed with a CEL I or RES I
5 cDNA controlled by appropriate expression control
elements (e.g., promoter, enhancer, sequences,
transcription terminators, poly-adenylation sites, etc.),
and a selectable marker. The selectable marker in the
recombinant plasmid confers resistance to the selection
10 and allows cells to stably integrate the plasmid into
their chromosomes and grow to form foci, which in turn
can be cloned and expanded into cell lines. For example,
following the introduction of foreign DNA, engineered
cells may be allowed to grow for 1-2 days in an enriched
15 media, and then are switched to a selective media. A
number of selection systems may be used, including but
not limited to the herpes simplex virus thymidine kinase
(Wigler, et al., 1977, Cell 11: 223), hypoxanthine-
guanine phosphoribosyltransferase (Szybalska & Szybalski,
20 1962, Proc. Natl. Acad. Sci. USA 48: 2026), and adenine
phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:
817) genes can be employed in tk⁻, hgprt⁻ or aprt⁻ cells
respectively. Also, antimetabolite resistance can be used
as the basis of selection for dhfr, which confers
25 resistance to methotrexate (Wigler, et al., 1980, Natl.
Acad. Sci. USA 77: 3567; O'Hare, et al., 1981, Proc.
Natl. Acad. Sci. USA 78: 1527); gpt, which confers
resistance to mycophenolic acid (Mulligan & Berg, 1981,
Proc. Natl. Acad. Sci. USA 78: 2072; neo, which confers
30 resistance to the aminoglycoside G-418 (Colberre-Garapin,
et al., 1981, J. Mol. Biol. 150: 1); and hygro, which
confers resistance to hygromycin (Santerre, et al., 1984,
Gene 30: 147) genes. Recently, additional selectable
genes have been described, namely trpB, which allows

cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, 1988, Proc. Natl. Acad. Sci. USA 85: 8047); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue L., 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory ed.).

Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as E. coli, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the CaCl_2 method using procedures well known in the art. Alternatively, MgCl_2 or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell if desired.

Animal and lower eukaryotic (e.g., yeast) host cells are competent or rendered competent for transformation by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, biolistics, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be cotransformed with DNA sequences encoding a CEL I or RES I of the invention, and a second foreign DNA molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40), Sindbis virus or bovine papilloma virus, to transiently

infect or transform eukaryotic cells and express the protein. (see for example, Eukaryotic Viral Vectors, Cold Spring Harbor Laboratory, Gluzman ed., 1982). The transfected cells are cultured by means well known in the art. Kuchler, R. J., Biochemical Methods in Cell Culture and Virology, Dowden, Hutchinson and Ross, Inc. (1977).

The present invention provides an *in vitro* method of making sequence variants from at least one heteroduplex polynucleotide wherein the heteroduplex has at least two non-complementary nucleotide base pairs, the method comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with strand cleavage activity provided by CEL I or RES I enzyme, proofreading activity, and ligase activity; and allowing sufficient time for the percentage of complementarity to increase, wherein at least one or more variants are made.

20

Another aspect of the present invention is where the heteroduplex polynucleotides are circular, linear or a replicon.

Another aspect of the present invention is where the desired variants have different amounts of complementarity.

Another aspect of the present invention is where the strand cleavage activity, proofreading activity, and ligase activity is added sequentially, or concurrently.

30

Another aspect of the present invention provides the addition of ligase activity, provided by agents such as, T4 DNA ligase, *E. coli* DNA ligase, or Taq DNA ligase.

5 In another aspect of the present invention, the strand cleavage activity is provided by an enzyme, such as, CEL I, RES I, T4 endonuclease VII, or T7 endonuclease I.

10 In another aspect of the present invention, polymerase activity is provided by Pol beta.

 In another aspect of the present invention, proofreading activity is provided T4 DNA polymerase or T7
15 DNA polymerase.

 In another aspect of the present invention, the effective amount of strand cleavage activity, and proofreading activity and ligase activity are provided by
20 CEL I or RES I enzyme, T4 DNA polymerase, and *E. coli* DNA ligase.

 In another aspect of the present invention, the effective amount of strand cleavage activity, and
25 proofreading activity and ligase activity are provided by CEL I or RES I enzyme, T7 DNA polymerase, and T4 DNA ligase.

 Another embodiment of the present invention provides
30 an in vitro method of increasing diversity in a population of sequences, comprising, preparing at least one heteroduplex polynucleotide; combining the heteroduplex polynucleotide with an effective amount of an agent or agents with proofreading activity, ligase

activity, and strand cleavage activity provided by CEL I or RES I enzyme; and allowing sufficient time for the percentage of complementarity to increase, wherein diversity in the population is increased.

5

Another embodiment of the present invention provides a method of obtaining a polynucleotide encoding a desired functional property, comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with proofreading activity, ligase activity, and strand cleavage activity provided by CEL I or RES I enzyme; allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, wherein diversity in the population is increased; and screening or selecting a population of variants for the desired functional property.

Another embodiment of the present invention provides a method of obtaining a polynucleotide encoding a desired functional property, comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with proofreading activity, ligase activity, and strand cleavage activity provided by CEL I or RES I enzyme; allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, wherein diversity in the population is increased; converting DNA to RNA; and screening or selecting a population of ribonucleic acid variants for the desired functional property.

Yet another embodiment of the present invention provides a method of obtaining a polypeptide having a desired functional property, comprising: preparing at least one heteroduplex polynucleotide; combining said
5 heteroduplex polynucleotide with an effective amount of an agent or agents with proofreading activity, ligase activity, and provided by CEL I or RES I enzyme; allowing sufficient time for the percentage of complementarity between strands of said heteroduplex
10 polynucleotide to increase, converting said heteroduplex polynucleotide to RNA, and said RNA to a polypeptide; and screening or selecting a population of polypeptide variants for said desired functional property.

15 Still another embodiment of the present invention provides a method of obtaining a polynucleotide encoding a desired functional property, comprising: preparing at least one heteroduplex polynucleotide, where the heteroduplex is optionally, about 95%, 90%, 85%, 80%,
20 75%, 62%, 58% or 47% identical, and about 100 base pairs, 1000 base-pairs, 10,000 base-pairs, or 100,000 base-pairs or more in size; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with proofreading activity, ligase activity, and
25 strand cleavage activity provided by CEL I or RES I enzyme; allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, screening or selecting for a population of variants having a desired functional
30 property; denaturing said population of variants to obtain single strand polynucleotides; annealing said single strand polynucleotides to form at least one second heteroduplex polynucleotide; combining said second heteroduplex polynucleotide with an effective amount of

an agent or agents with proofreading activity, ligase activity, and strand cleavage activity provided by CEL I or RES I enzyme; and allowing sufficient time for the percentage of complementarity between strands of the
5 heteroduplex polynucleotide to increase.

The present invention is directed to a method for generating an improved polynucleotide sequence or a population of improved polynucleotide sequences,
10 typically in the form of amplified and/or cloned polynucleotides, whereby the improved polynucleotide sequence(s) possess at least one desired phenotypic characteristic (e.g., encodes a polypeptide, promotes transcription of linked polynucleotides, binds a protein,
15 improves the function of a viral vector, and the like) which can be selected or screened for. Such desired polynucleotides can be used in a number of ways such as expression from a suitable plant, animal, fungal, yeast, or bacterial expression vector, integration to form a
20 transgenic plant, animal or microorganism, expression of a ribozyme, and the like.

GRAMMR provides for resolution of mismatched base pairs on heteroduplex DNA strands in an *in vitro*
25 reaction. This reaction begins with cleavage of one strand or the other at or near a mismatch followed by excision of mismatched bases from the cleaved strand and polymerization to fill in the resulting gap with nucleotides that are templated to the sequence of the
30 other strand. The resulting nick can be sealed by ligation to rejoin the backbone. The sum of the separate independent mismatch resolution events that occur throughout the heteroduplex molecule will result in reassortment of sequence information throughout that

molecule. The sequence information will reassort in a variety of combinations to generate a complex library of sequence variants.

5 In one embodiment of GRAMMR, a library of mutants is generated by any method known in the art such as mutagenic PCR, chemical mutagenesis, etc. followed by screening or selection for mutants with a desired property. The mutant DNAs are mixed, denatured to
10 single strands, and allowed to anneal. Partially complementary strands that hybridize will have non-base-paired nucleotides at the sites of the mismatches. Treatment with CEL I (Oleykowski et al., 1998; Yang et al., 2000), or a similar mismatch-directed activity, such
15 as RES I, will cause nicking of one or the other polynucleotide strands at or near mismatches. In addition, CEL I or RES I can nick at or near an insertion/deletion. The presence of a polymerase containing a proofreading activity (e.g., T4 DNA Pol)
20 will allow excision of the mismatch, and subsequent polymerase activity will fill in the gap using the other strand as a template. A polymerase that lacks 5' to 3' exonuclease activity and strand-displacement activity will fill in the gap and will cease to polymerize when it
25 reaches the 5' end of DNA located at the original CEL I cleavage site, thus re-synthesizing only short patches of sequence. DNA ligase (e.g., T4 DNA ligase or *E. coli* DNA ligase) can then seal the nick by restoring the phosphate backbone. This process can occur simultaneously at many
30 sites and on either strand of a given heteroduplex DNA molecule. The result is a randomization of sequence differences among input strands to give a population of sequence variants that is more diverse than the population of starting sequences. These output

polynucleotides can be cloned directly into a suitable vector, or they can be amplified by PCR before cloning. Alternatively, the reaction can be carried out on heteroduplex regions within the context of a double-stranded circular plasmid molecule or other suitable replicon that can be directly introduced into the appropriate host following the GRAMMR reaction. In another alternative, the output polynucleotides can be transcribed into RNA polynucleotides and used directly, for example, by inoculation of a plant viral vector onto a plant, such as in the instance of a viral vector transcription plasmid. The resulting clones are subjected to a selection or a screen for improvements in a desired property. The overall process can then be repeated one or more times with the selected clones in an attempt to obtain additional improvements.

If the output polynucleotides are cloned directly, there is the possibility of incompletely resolved molecules persisting that, upon replication in the cloning host, could lead to two different plasmids in the same cell. These plasmids could potentially give rise to mixed-plasmid colonies. If it is desired to avoid such a possibility, the output polynucleotide molecules can be grown in the host to allow replication/resolution, the polynucleotides isolated and retransformed into new host cells.

In another embodiment, when sequence input from more than two parents per molecule is desired, the above procedure is performed in a cyclic manner before any cloning of output polynucleotides. After the GRAMMR reaction, the double stranded polynucleotides are denatured, allowed to anneal, and the mismatch resolution

process is repeated. After a desired number of such cycles, the output polynucleotides can be cloned directly, introduced into a suitable vector, or they can be amplified by PCR before cloning. The resulting clones
5 are subjected to a selection or a screen for improvements in a desired property.

In another embodiment, a "molecular backcross" is performed to help eliminate the background of deleterious
10 mutations from the desired mutations. A pool of desired mutant DNAs can be hybridized to wild-type DNA to perform the method. Clones can be selected for improvement, pooled, and crossed back to wild-type again until there is no further significant change.

15

The efficiency of the process is improved by various methods of enriching the starting population for heteroduplex molecules, thus reducing the number of unaltered parental-type output molecules. The mismatched
20 hybrids can be affinity purified using aptamers, dyes, or other agents that bind to mismatched DNA. A preferred embodiment is the use of MutS protein affinity matrix (*Wagner et al., Nucleic Acids Res. 23(19):3944-3948 (1995); Su et al., Proc. Natl. Acad. Sci. (U.S.A.),*
25 *83:5057-5061(1986)*) or mismatch-binding but non-cleaving mutants of phage T4 endonuclease VII (*Golz and Kemper, Nucleic Acids Research, 1999; 27: e7*).

In one embodiment, the procedure is modified so that
30 the input polynucleotides consist of a single strand of each sequence variant. For example, single-stranded DNAs of opposite strandedness are produced from the different parent sequences by asymmetric PCR to generate partially complementary single-stranded molecules. Annealing of

the strands with one-another to make heteroduplex is performed as described in EXAMPLE 2. Alternatively, single-stranded DNAs can be generated by preferentially digesting one strand of each parental double-stranded DNA with Lambda exonuclease followed by annealing the remaining strands to one-another. In this embodiment, the annealing strands have no 100% complementary strand present with which to re-anneal. Hence, there is a lower background of unmodified polynucleotides, that is, "parental polynucleotides" among the output polynucleotides leading to a higher efficiency of reassorting sequence variations. This increased efficiency will be particularly valuable in situations where a screen rather than a selection is employed to test for the desired polynucleotides.

Another method for heteroduplex formation is to mix the double-stranded parent DNAs, denature to dissociate the strands, and allow the single-stranded DNAs to anneal to one-another to generate a population of heteroduplexes and parental homoduplexes. The heteroduplexes can then be selectively enriched by a heteroduplex capture method such as those described above using MutS or a non-cleaving T4 endonuclease VII mutant. Alternatively, the parental homoduplex molecules in the population may be cleaved by restriction enzymes that overlap with sites of mismatch such that they are not cleaved in the heteroduplex but are cleaved in the parental homoduplex molecules. Uncleaved heteroduplex DNA can then be isolated by size fractionation in an agarose gel as was performed to generate full-length plasmid on full-length plasmid heteroduplex DNA molecules as describe in EXAMPLE 9. Nick-sealing in those full-length heteroduplexed

plasmid molecules was then brought about by incubation with DNA ligase.

In another embodiment, the parental, or input,
5 double-stranded polynucleotides are modified by the addition of "clamp" sequences. One input polynucleotide or pool of polynucleotides is amplified by PCR with the addition of a unique sequence in the 5' primer. The other input polynucleotide or pool is amplified by PCR
10 with the addition of a unique sequence in the 3' primer. The clamp sequences can be designed to contain a unique restriction enzyme site for the 5' end of the gene of interest and another for the 3' end such that, at the step of cloning the products of the GRAMMR reaction, only
15 products with the 5' clamp from the first polynucleotide (or pool) and the 3' end from the second polynucleotide (or pool) will have appropriate ends for cloning. Alternatively, the products of the GRAMMR reaction can be PCR amplified using the unique sequences of the 5' and 3'
20 clamps to achieve a similar result. Hence, there is a lower background of unmodified polynucleotides, that is, "parental polynucleotides" among the output polynucleotide clones leading to a higher efficiency of reassorting sequence variations. This increased
25 efficiency will be particularly valuable in situations where a screen rather than a selection is employed to test for the desired polynucleotides. Optionally, oligonucleotide primers can be added to the GRAMMR reaction that are complementary to the clamp primer
30 sequences such that either parent can serve as the top strand, thus permitting both reciprocal heteroduplexes to participate in the mismatch-resolution reaction.

Another method for generating cyclic heteroduplexed polynucleotides is performed where parental double-stranded DNAs have terminal clamp sequences as described above where the single-stranded clamp sequences extending
5 from one end of the heteroduplex are complementary to single-stranded clamp sequences extending from the other end of the heteroduplex. These complementary, single-stranded clamps are allowed to anneal, thereby circularizing the heteroduplexed DNA molecule. Parental
10 homoduplexes that result from re-annealing of identical sequences have only one clamp sequence and therefore, no complementary single-stranded sequences at their termini with which circularization can occur. Additionally, a DNA polymerase and a DNA ligase can be used to fill-in
15 any gaps in the circular molecules and to seal the nicks in the backbone, respectively, to result in the formation of a population of covalently closed circular heteroduplex molecules. As the covalently-closed circular heteroduplex molecules will not dissociate into
20 their component strands if subjected to further denaturing conditions, the process of denaturation, circularization, and ligation can be repeated to convert more of the linear double-stranded parental duplexes into closed into closed circular heteroduplexes.

25

In another embodiment, a region of a single-stranded circular phagemid DNA can be hybridized to a related, but non-identical linear DNA, which can then be extended with a polymerase such as T7 DNA polymerase or T4 DNA
30 polymerase plus T4 gene 32 protein, then ligated at the resulting nick to obtain a circular, double-stranded molecule with heteroduplexed regions at the sites of differences between the DNAs. GRAMMR can then be carried

out on this molecule to obtain a library of sequence-reassorted molecules.

Alternately, two single-stranded circular phagemid
5 DNAs of opposite strand polarity relative to the plasmid backbone, and parent gene sequences that are the target of the reassortment are annealed to one and other. A region of extensive mismatch will occur where the phage fl origin sequences reside. Upon GRAMMR treatment,
10 however, this region of extensive mismatch can revert to either parental type sequence restoring a functional fl origin. These double-stranded molecules will also contain mismatch regions at the sites of differences between the strands encoding the parent genes of
15 interest. GRAMMR can then be carried out on this molecule to obtain a library of sequence re-assorted molecule.

As discussed in the preceding paragraphs, the
20 starting DNA or input DNA can be of any number of forms. For example, input DNA can be full-length, single stranded and of opposite sense, as is taught in EXAMPLE 2. Alternatively, the input DNA can also be a fragment of the full-length strand. The input DNAs can be double-
25 stranded, either one or both, or modified, such as by, methylation, phosphorothiolate linkages, peptide-nucleic acid, incorporation of uracil into the DNA, substitution of RNA in one or both strands, or the like. Either strand of a duplex can be continuous along both strands,
30 discontinuous but contiguous, discontinuous-with overlaps, or discontinuous with gaps.

GRAMMR can also be applied to DNA fragmentation and reassembly-based DNA shuffling schemes. For instance, in

methods where gene fragments are taken through cycles of denaturation, annealing, and extension in the course of gene reassembly, GRAMMR can be employed as an intermediate step.

5

In one such embodiment, the DNA from a gene, or pool of mutant genes is fragmented by enzymatic, mechanical or chemical means, and optionally a size range of said fragments is isolated by a means such as separation on an agarose gel. The starting polynucleotide, such as a wild-type, or a desired variant, or a pool thereof, is added to the fragments and the mixture is denatured and then allowed to anneal. The annealed polynucleotides are treated with a polymerase to fill in the single stranded gaps using the intact strand as a template. The resulting partially complementary double strands will have non-base-paired nucleotides at the sites of the mismatches. Treatment with CEL I (Oleykowski et al., 1998; Yang et al., 2000), or an agent with similar activity, such as RES I, will cause nicking of one or the other polynucleotide strands at or near mismatches. Addition of a polymerase containing a proofreading activity, such as T4 DNA Polymerase, will allow excision of the mismatch, and subsequent polymerase activity will fill in the gap using the other strand as a template. A DNA ligase, such as, T4 DNA Ligase, can then seal the nick by restoring the phosphate backbone. The result is a randomization of sequence variation among input strands to give output strands with potentially improved properties. These output polynucleotides can be cloned directly into a suitable vector, or they can be amplified by PCR before cloning. The resulting clones are subjected to a selection or a screen for improvements in a desired property.

10

15

20

25

30

In one such embodiment, the DNA from a pool of mutant genes is fragmented by enzymatic, mechanical or chemical means, or fragments are generated by limited extension of random oligonucleotides annealed to parental templates (United States Patent 5,965,408), and optionally a size range of said fragments is isolated by a means such as separation on an agarose gel. The mixture is denatured and then allowed to anneal. The annealed polynucleotides are optionally treated with a polymerase to fill in the single stranded gaps. The resulting partially complementary double-strand fragments will have non-base paired nucleotides at the sites of the mismatches. Treatment with CEL I (Oleykowski *et al.*, 1998; Yang *et al.*, 2000), or an agent with similar activity, such as RES I, will cause nicking of one or the other polynucleotide strand 3' of each mismatch. The activity of a polymerase containing a proofreading activity, such as T4 DNA Polymerase, will allow excision of the mismatch, and subsequent polymerase activity will fill in the gap using the other strand as a template. Optionally, DNA ligase, such as, T4 DNA Ligase, can then seal the nick by restoring the phosphate backbone. The result is a randomization of sequence variation among input strands to give output strands with potentially improved properties. Subsequent rounds of denaturing, annealing, and GRAMMR allows gene reassembly. PCR can be used to amplify the desired portion of the reassembled gene. These PCR output polynucleotides can be cloned into a suitable vector. The resulting clones are subjected to a selection or a screen for the desired functional property.

Another embodiment of the present invention provides starting with a continuous scaffold strand to which fragments of another gene or genes anneal. The flaps and gaps are trimmed and filled as is described in Coco, et al., Nature Biotech 19 (01)354; United States Patent 6,319,713, and GRAMMR is performed. In this process, GRAMMR would bring about further sequence reassortment by permitting transfer of sequence information between the template strand and the strand resulting from flap and gap trimming and ligation. This method provides the benefits of incorporating specific sequence patches into one continuous strand followed by GRAMMR of residues that mismatch with the scaffold. By annealing many fragments simultaneously to the same sequence or gene, many individual sites can be addressed simultaneously, thereby allowing reassortment of multiple sequences or genes at once. In the present embodiment, the scaffold is not necessarily degraded, rather the duplex can be directly cloned, or amplified by PCR prior to cloning. Exhaustive mismatch resolution will result in a perfectly duplexed DNA. Partial mismatch resolution will result in essentially two different reassorted products per duplex.

As can be appreciated from the present disclosure, GRAMMR can also be applied to a variety of methods that include the annealing of related DNAs as a step in their process. For example, many site-directed mutagenesis protocols call for the annealing of mutant-encoding DNA molecules to a circular DNA in single-stranded form, either phagemid or denatured plasmid. These DNAs are then extended with a polymerase, followed by treatment with ligase to seal the nick, with further manipulation to remove the parental sequence, leaving the desired mutation or mutations incorporated into the parental

genetic background. Though these protocols are generally used to incorporate specific mutations into a particular DNA sequence, it is feasible that the GRAMMR reaction can be applied to the heteroduplexed molecules generated in such a process to reassort sequence variations between the two strands, thereby resulting in a diverse set of progeny with reassorted genetic variation.

Another embodiment provides for sequential rounds of reassortment on only a particular region of the DNA of interest. For example, DNA fragments are annealed to a circular single-strand phagemid DNA, and GRAMMR is performed. The fragments can be treated in order to prevent them from being physically incorporated into the output material. For example, they can be terminated at the 3' end with di-deoxy residues making them non-extendible. Multiple rounds of reassortment can be performed, but only modified molecules from the original input single stranded DNA clone will be recovered. The consequence will be that the DNA fragments used in this reassortment will contribute only sequence information to the final product and will not be physically integrated into the final recoverable product.

GRAMMR can be used for protein, peptide, or aptamer display methods to obtain recombination between library members that have been selected. As fragmentation of the input DNAs is not required for GRAMMR, it may be possible to reassort sequence information between very small stretches of sequence. For instance, DNAs encoding small peptides or RNA aptamers that have been selected for a particular property such as target binding can be reassorted. For annealing to occur between the selected DNA molecules, some level of sequence homology should be

shared between the molecules, such as at the 5' and 3' regions of the coding sequence, in regions of the randomized sequence segment that bear similarity because of similar binding activities, or through the biasing of codon wobble-base identity to a particular set of defaults.

Manipulation of the reaction temperature at which GRAMMR is conducted can be useful. For example, lower temperatures will help to stabilize heteroduplexes allowing GRAMMR to be performed on more highly mismatched substrates. Likewise, additives that affect base-pairing between strands, such as salts, PEG, formamide, etc, can be used to alter the stability of the heteroduplex in the GRAMMR reaction, thereby affecting the outcome of the reaction.

Another embodiment provides for zonal mutagenesis by GRAMMR, that is, random or semi-random mutations at, and in the immediate vicinity of, mismatched residues using nucleotide analogues that have multiple base-pairing potential. This provides for concentration of essentially random mutagenesis at a particular point of interest, and adds another benefit to the present invention. Groups of genes that are similar, but have slightly different functions from one-another, for example, many enzymes, will exhibit moderate sequence differences from one-another in regions that will be operative for their own particular activities. These activities, can include substrate preference, binding partners, regulatory sites, or the like. Gene sequences that govern these functions should be heterogeneous within the population of related genes. Since it is known that the specificity of such function is associated

with these amino acids and their neighbors, GRAMMR mutagenesis, in addition to reassorting sequence information between genes, may also be used to direct random mutagenesis to these regions to evolve their
5 function, while not disturbing other sequences, such as structural framework, invariant residues, and other such important sites, that are potentially less tolerant to randomization.

10 Different enzymes with distinct functions will not differ just in the operative regions, such as active sites and regulatory sites. They are likely to have other differences from one another that arise through genetic drift. Further randomization in the locales of
15 such changes might therefore be considered neutral, minimally important, or deleterious to the outcome of a mutagenesis experiment. In order to direct the random mutagenesis away from such inconsequential sites, and toward sites that might present a better result for
20 random mutagenesis, such as the active site of an enzyme, the codon usage bias of the genes could be manipulated to decrease or increase the overall level of nucleotide complementarity in those regions. If regions of greater complementarity are less susceptible to GRAMMR than
25 regions of lesser complementarity, then the degree of GRAMMER-directed zonal random mutagenesis at a given site can be modulated.

In any DNA shuffling experiment, it is desirable to
30 minimize the proportion of non-shuffled, or parental, DNAs that are obtained within the population of shuffled progeny. Numerous approaches may be used to accomplish this. In a plasmid-on-plasmid DNA shuffling format, where the genes to be shuffled are present on separate,

but otherwise identical plasmids, each plasmid is linearized at one or another different unique restriction sites that are present. After removal of the restriction endonucleases, the linearized DNAs are mixed, melted
5 apart, and allowed to anneal so that populations of heteroduplex DNA form that are either nicked, closed circular heteroduplex molecules, or are double stranded and linear homoduplexes. It is the population of circular double-stranded heteroduplex DNA molecules that
10 represents the desired substrate for the GRAMMR reaction. One can either enrich this desired population by gel fractionation or use one or a number of methods that do not require physical separation of this population, but rather, discourages the recovery of non-shuffled parental
15 molecules. Several such methods are listed below.

First, after GRAMMR reaction of the mixed population of linear parental homoduplex and circular double-stranded heteroduplex, transformation of *E. coli* is
20 generally performed. Since circular DNA is vastly more efficient at transforming *E. coli* than its linearized counterpart, the parental homoduplexes can be strongly discriminated against at this step by preventing their circularization into transformation-competent molecules.
25 The use of *E. coli* DNA ligase as the ligase component of the GRAMMR reaction will serve to prevent recircularization of parental homoduplex, as it more efficiently seals nicks than joins short cohesive termini that result from restriction endonuclease cleavage.
30 Additionally, this enzyme very inefficiently ligates blunt ends. As a result of using this strategy, the progeny resulting from transformation of *E. coli* with the GRAMMR reaction are depleted of non-shuffled parental

genes and enriched for molecules that entered the GRAMMR reaction as heteroduplex substrates.

Another method for excluding parental gene contamination from the population of GRAMMR output molecules is to position the plasmid linearization sites within a selectable marker. The sites should be of sufficient distance from one another to allow annealing to take place between staggered ends of a heteroduplex, and should either have overhangs that can be filled-in or trimmed off, or cause a deletion of sequence upon cleavage. As above, the plasmids containing the genes to be shuffled are linearized at one or other of the sites. After removal of the restriction endonucleases, the linearized DNAs are mixed, melted, and allowed to anneal. The resulting sample is made up of a mixture of circular heteroduplexes and of linear homoduplexes. This sample can then be treated with a proofreading polymerase such as T4 DNA polymerase in the presence of dNTPs. The circular homoduplexes should be unaffected, whereas the linear parental homoduplexes will have been blunted at their termini, effectively adding or deleting bases to the sequence of the selectable marker if that molecule becomes recircularized at any point in the GRAMMR reaction or after transformation into *E. coli*. If the addition or deletion of these sequences results in disruption of the function of the selectable marker, then the resulting molecules will not be recovered under appropriate selection.

Another method one can use to prevent unshuffled parental contamination of the shuffled library is to dephosphorylate the linearized DNAs prior to melting and annealing. Linear homoduplex molecules will be rendered unable to ligate into circular molecules whereas circular

heteroduplexes will simply contain a single nick in each strand, but will still remain circular, and thus competent for efficient transformation into *E. coli*.

5 Another method one can use to prevent unshuffled parental contamination of the shuffled library is to digest with enzymes whose recognition sites are overlapped by mismatches in the heteroduplexed molecules. Digestion of the parental homoduplexes at those sites
10 will render the resulting molecules linear so that they may be subject to any of the treatments described above to reduce parental contamination. The resulting molecules may also be made smaller, facilitating separation from the intact circular heteroduplex
15 molecules.

 If, in addition to excluding unshuffled parental molecules from a shuffling experiment, one desires to prevent shuffling between any two or more genes of a
20 population of two or more parent genes, the same principles described above can be applied.

 In the current invention the random reassortment occurs in an *in vitro* DNA mismatch-resolution reaction.
25 This method does not require any steps of "gene reassembly" that serve as the foundation for the earlier mutation reassortment ("shuffling") methods. Instead, it is based upon the ability of a reconstituted or artificial DNA mismatch resolving system to transmit
30 sequence variations from one or more strands of DNA into another DNA strand by hybridization and mismatch resolution *in vitro*.

In general, standard techniques of recombinant DNA technology are described in various publications, e.g., (Ausubel, 1987; Ausubel, 1999; Sambrook et al., 1989).

Polynucleotide modifying enzymes were used according to
5 the manufacturer's recommendations. If desired, PCR
amplimers for amplifying a predetermined DNA sequence
may be chosen at the discretion of the practitioner.

10

It is noted that each of the activities taught in
the present invention that are involved in the GRAMMR
reaction can be interchanged with a functional equivalent
agent with similar activity, and that such changes are
15 within the scope of the present invention. For instance,
as was indicated in EXAMPLE 6, Taq DNA ligase could
substitute for T4 DNA ligase. Other ligases can be
substituted as well, such as *E. coli* DNA ligase.
Likewise, as shown in EXAMPLE 12, T7 DNA polymerase can
20 be substituted for T4 DNA polymerase. Other enzymes with
appropriate proofreading activity can function in place
of any of these enzymes for the proofreading activity
needed for the GRAMMR reaction. In a similar way, any
polymerase with functionally equivalent activity to those
25 demonstrated to work for GRAMMR can be used for
substitution.

Strand cleavage may be brought about in a number of
ways. In addition to CEL I, a number of functionally
30 equivalent, and potentially similar activities found in
extracts from a variety of plant species (Oleykowski,
Nucleic Acids Res 1998;26:4597-602) may be used. Other
mismatch-directed endonucleases such as T4 endonuclease
VII, T7 endonuclease I, and SP nuclease (Oleykowski,

Biochemistry 1999; 38: 2200-5) may be used. Another particularly useful mismatch-directed endonuclease is RES I.

5 CEL I and RES I can be modified to obtain enzyme variants with desired properties. Homologous nucleic acid sequences from other sources can be found by surveying plants. Alternatively, CEL I or RES I can be mutated, and opposite strands with less than 100% identity can be
10 brought together to form a heteroduplex molecule. The present method disclosed herein or another method can be used to evolve new CEL I and RES I enzyme variants.

15 **CEL I IS A MISMATCH ENDONUCLEASE**

CEL I is a mismatch endonuclease isolated from celery. The use of CEL I in a diagnostic method for the detection of mutations in targeted polynucleotide
20 sequences, in particular, those associated with cancer, is disclosed in U.S. Patent No. 5,869,245. Methods of isolating and preparing CEL I are also disclosed in this patent. However, there is no disclosure in this patent relating to the use of CEL I in DNA sequence
25 reassortment.

Nucleic acid molecules that encode CEL I are disclosed in PCT Application Publication No. WO 01/62974 A1. As with U.S. Patent No. 5,869,245, the use of CEL I
30 in a diagnostic method for the detection of mutations in targeted polynucleotide sequences associated with cancer is disclosed. Also similarly, there is no disclosure relating to the use of CEL I in DNA sequence reassortment.

RES I IS A MISMATCH ENDONUCLEASE

The use of RES I endonuclease is contemplated in
5 diagnostic methods for the detection of mutations in
targeted polynucleotide sequences, in particular, those
associated with cancer. Examples of some of these types
of diagnostic methods are disclosed in U.S. Patent No.
5,869,245, and Del Tito, et al.. The use of RES I is also
10 contemplated in methods for detecting mutations in large
genomic regions Sokurenko, et al., and for mutation
screening methods such as TILLING as disclosed in
McCallum, et al.

15 The reactivity of Endonuclease VII of phage T4 with
DNA-loops of eight, four, or one nucleotide, or any of 8
possible base mismatches *in vitro* is disclosed in
"Endonuclease VII of Phage T4 Triggers Mismatch
Correction *in Vitro*" Solaro, et al., J Mol Biol
20 230(93)868. The publication reports a mechanism where
Endonuclease VII introduces double stranded breaks by
creating nicks and counternicks within six nucleotides 3'
of the mispairing. The publication discloses that a time
delay between the occurrence of the first nick and the
25 counternick was sufficient to allow the 3'-5' exonuclease
activity of gp43 to remove the mispairing and its
polymerase activity to fill in the gap before the
occurrence of the counternick. Nucleotides are erased
from the first nick, which is located 3' of the mismatch
30 on either strand and stops 5' of the mismatch at the
first stable base-pair. The polymerase activity proceeds
in the 5' to 3' direction towards the initial nick, which
is sealed by DNA ligase. As a result, very short repair
tracks of 3 to 4 nucleotides extend across the site of

the former mismatch. The publication concludes with a discussion regarding the various activities Endonuclease VII may have within phage T4. However, the publication does not disclose any practical utility for Endonuclease VII outside of phage T4, and there is no disclosure regarding its applicability in DNA reassortment.

Methods of creating libraries of chimeric DNA sequences *in vivo* in *Escherichia coli* are disclosed in *Nucleic Acids Research*, 1999, Vol 27, No. 18, e18, Volkov, A. A., Shao, Z., and Arnold, F.H. Abostado, J.P., et. Al., (1984). *Proc Natl Acad Sci USA* 81(18): 5792-5796. Cami, B.P., et al., (1984). *Proc Natl Acad Sci USA* 81(2): 503-507. Chang, S.D., et al., (1984). *Gene* 29(3): 255-261. The methods use heteroduplexes formed *in vitro* to transform *E. coli* where repair of regions of non-identity in the heteroduplex creates a library of new, recombined sequences composed of elements of each parent. Although the publications disclose the use of this method as a convenient addition to existing DNA recombination methods, the disclosed methods are limited to the *in vivo* environment of *E. coli*. Volkov, A.A., et al., *Methods in Enzymology* 328: 456-463(2000) Academic Press, Inc., San Diego state that there is more than one mechanism available for mismatch repair in *E. coli*, and speculates that the 'long patch' repair mechanism, which utilizes the MutS/L/H enzyme system, was probably responsible for the heteroduplex repair.

The following non-limiting examples are provided to illustrate the present invention.

EXAMPLE 1
CLONING, EXPRESSION AND PURIFICATION
OF CEL I ENDONUCLEASE

5 This example teaches the preparation of nucleic acid
molecules that were used for expressing CEL I
endonuclease from plants, identified herein as, p1177MP4-
CEL I Avr (SEQ ID NO: 1), and p1177MP4-CEL I 6HIS (SEQ ID
NO: 2). In particular, this example refers to
10 disclosures taught in U.S. Patents No. 5,316,931,
5,589,367, 5,866,785, and 5,889,190.

Celery RNA extraction:

15 Celery was purchased from a local market. Small
amounts of celery tissue (0.5 to 0.75 grams) were
chopped, frozen in liquid nitrogen, and ground in a
mortar and pestle in the presence of crushed glass.
After addition of 400 microliters of Trizol and further
20 grinding, 700 microliters of the extract were removed and
kept on ice for five minutes. Two hundred microliters of
chloroform were then added and the samples were
centrifuged, left at room temperature for three minutes,
and re-centrifuged at 15,000g for 10 minutes. The
25 aqueous layer was removed to a new tube and an equal
volume of isopropanol was added. Tubes were inverted to
mix and left at room temperature for 10 minutes followed
by centrifugation at 15,000g for ten minutes at 4°C. The
pellet was washed twice in 400 microliters of 70%
30 ethanol, once in 100% ethanol, air dried, and resuspended
in 40 microliters of distilled water. One microliter of
RNasin was added and 3.5 microliters was run on a 1%
agarose gel to check the quality of the RNA prep (Gel

picture). The remainder was stored at -70°C until further use.

***CellI* gene cloning and expression by a viral vector:**

5 The total RNA from celery was subjected to reverse transcription followed by PCR to amplify the cDNA encoding the *CellI* gene sequence. In separate reactions, eleven microliters of the total celery RNA prep was mixed with one microliter (50 picomoles) of either *CellI*-Avr-R, 10 *CellI*-6H-R, or with two microliters of oligo dT primer. *CellI*-Avr-R was used to prime cDNA and amplify the native *CellI* sequence at the 3' end of the gene, while *CellI*-6H-R was used to add a sequence encoding linker peptide and a 6-His tag to the 3' terminus of the *CellI* gene. The 15 samples were heated to 70°C for one minute and quick-chilled on ice prior to the addition of 4 microliters of 5X SuperscriptTM II buffer, two microliters of 0.1M DTT, 1 microliter of 10mM each dNTP, and 1 microliter of Superscript II (Gibco/BRL) to each reaction. The 20 reactions were incubated at 42°C for one hour.

 PCR amplification of the *CellI* cDNA sequence was performed using the method of W.M. Barnes (*Proc Natl Acad Sci. USA*, 1994 Mar 15;91(6):2216-20) with a Taq-Pfu 25 mixture or with Pfu alone. The RT reaction primed with *CellI*-Avr-R was used as template for a PCR using primers *CellI*-Pac-F (as the forward primer) paired with *CellI*-Avr-R (as the reverse primer). In other PCRs, the RT reaction that was primed with oligo dT was used as template for 30 both of the above primer pairs. All PCR reactions were performed in 100 microliters with 30 cycles of annealing at 50°C and two minutes of extension at 72°C. Aliquots of the resulting reactions were analyzed by agarose gel electrophoresis. Reactions in which Pfu was used as the

sole polymerase showed no product. All reactions performed with the Taq/Pfu mixtures yielded product of the expected size. However, those amplified from cDNA primed with *CelI* specific primer pairs gave more product than reactions amplified from cDNA primed with oligo-dT. DNAs from the PCR reactions that gave the most product were purified using a Zymoclean DNA spin column kit and digested with PacI and AvrII, gel-isolated, and ligated into PacI and AvrII-digested plasmid pRT130, a tobamovirus-based GENEWARE® vector. 2 microliters of each ligation were transformed into DH5α competent *E. coli* and cultured overnight on LB-amp agar plates. Colonies were picked and grown overnight in liquid culture, and plasmid DNA was isolated using a Qiagen plasmid prep kit. 12 clones from each construct were screened by digestion with PacI and AvrII and 11 of 12 of each set were positive for insert of the correct size. Ten of the clones for each construct were transcribed *in-vitro* and RNA was inoculated to *N. benthamiana* plants. In addition, the CELI gene inserts in both sets of ten clones were subjected to sequence analysis. Several clones containing inserts encoding the native form of CEL I had sequence identical to the published *CelI* sequence in WO 01/62974 A1. One clone containing an insert encoding CEL I fused to a 6-Histidine sequence was identical to the published *CelI* sequence. One clone of each (pRT130-CEL I Avr-B3 and pRT130-CEL I 6His-A9, respectively) was selected for further work. The CEL I-encoding sequences in these clones were subsequently transferred to another GENEWARE® vector. The sequences of these clones, p1177MP4- CEL I Avr-B3, and p1177MP4-CEL I 6His-A9 are provided as SEQ ID NO: 1 and SEQ ID NO: 2, respectively.

Assay of cloned CEL I activities:

To determine whether the GENEWARE® constructs containing *CelI* sequences could produce active CEL I enzyme, samples of pRT130-CEL I Avr and pRT130-CEL I 6His, and GFP-GENEWARE control-infected plants were harvested and homogenized in a small mortar and pestle in Tris-HCl at pH 8.0. Extracts were clarified and assayed for supercoiled DNA nicking activity. Each supercoiled DNA nicking assay was performed in a reaction containing 0.5 micrograms of a supercoiled plasmid prep of a pUC19-derivative in 1X NEB ligase buffer in a total volume of 10 microliters. The amounts of plant extract added to the reactions were 0.1 microliter, 0.01 microliter, or 0.001 microliter, incubated at 42°C for 30 minutes, and run on a 1% TBE-agarose gel in the presence of ethidium bromide. Little or no nicking activity was detected in the GFP-GENEWARE control-infected plant extract whereas extracts from plants infected with the CEL I-GENEWARE constructs showed appreciable amounts of activity against the plasmid DNA substrate.

Additional activity assays were performed on extracts of plants inoculated with pRT130-CEL I Avr-B3 and pRT130-CEL I 6His-A9. In these assays, intracellular fluid was washed from infected leaves and assayed separately from material obtained from the remaining washed leaf tissues. Assays were performed as described above with the exception that the incubation was at 37°C for one hour. Samples were run on a 1% TBE-agarose gel in the presence of ethidium bromide and photographed.

Purification of 6His-tagged CEL I from infected *N. benthamiana* plants:

N. benthamiana plants were inoculated with RNA transcripts from pRT130-CEL I 6His-A9 at 20-21 days post-sowing. Tissues were harvested from 96 infected plants at 10 days post-inoculation and subjected to

5 intracellular fluid washes. Briefly, infected leaf and stem material was vacuum infiltrated for 30 seconds twice with chilled infiltration buffer (50 mM phosphate pH 4 in the presence of 7 mM β -ME). Infiltrated tissues were blotted to adsorb excess buffer and secreted proteins

10 were recovered by centrifugation at 2500x g for 20 min using basket rotor (Beckman). PMSF was added to the extracted intracellular fluid (IF) containing recombinant CEL I to a final concentration of 1 mM, and incubated at 25°C for 15 min with stirring. After addition of Imidazole

15 (pH 6.0) and NaCl to the extract to the final concentration of 5 mM and 0.5 M respectively, IF was adjusted to pH 5.2 and filtered through 1.2 μ Sartorius GF membrane (Whatman) to remove most of the Rubisco and green pigments. Immediately after clarification, pH was

20 adjusted to 7.0 using concentrated NaOH solution and incubated on ice for 20 min to allow non-proteinaceous material to precipitate. IF was further clarified using 0.8 μ or 0.65/0.45 μ Sartorius GF (Whatman). Recombinant CEL I was purified from the clarified IF by metal

25 chelating affinity chromatography using Ni^{2+} Fast Flow Sepharose (Amersham Pharmacia Biotech, NJ) equilibrated with binding buffer (50 mM phosphate, 0.5 M NaCl; pH 7.0) containing 5 mM imidazole, with a linear velocity of 300 cm/hr. Unbound protein was washed with 20 mM

30 imidazole/binding buffer, and CEL I was eluted from Ni^{2+} SepharoseTM with a linear gradient of 20 to 400 M imidazole in the binding buffer. Fractions still containing imidazole were assayed for supercoiled DNA nicking activity as described above but were found to have

negligible activity. The same fractions were then dialyzed against 0.1 M Tris-HCl, pH 8.0 in the presence of ZnCl₂ using 10 kD molecular weight cutoff dialysis tubing (Pierce) and assayed again. The supercoiled DNA
5 nicking activity was restored after this dialysis.

IF and purified CEL I protein were analyzed using Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) precast Tris-glycine gels (Invitrogen,
10 Carlsbad, CA) in the buffer system of Laemmli with a XcellTM II Mini-Cell apparatus (Invitrogen, Carlsbad, CA). The protein bands were visualized by Coomassie brilliant blue and by silver staining. SDS-PAGE Gels were scanned and analyzed using Bio-Rad gel imager.

15

Mass spectrometry of purified CEL I:

The average molecular mass of the purified CEL I was determined by matrix-assisted laser/desorption ionization time-of-flight mass spectrometry (MALDI-TOF). An aliquot
20 of CEL I was diluted 1:10 with 50% acetonitrile/water and mixed with sinapinic acid matrix (1:1 v/v) using a PE Biosystem DE-Pro mass spectrometer. The mass spectrometry was performed using an accelerating voltage of 25 kV and in the positive-linear ion mode.

25

Mass spectrometry of peptides isolated from purified CEL I:

CEL I was separated on SDS-PAGE on a 14% gel and stained with Coomassie brilliant blue. A single
30 homogenous band was visible. This band was excised and de-stained completely. Protein was reduced in the presence of 10 mM DDT in 50% acetonitrile for 30 min at 37°C and reduced sulfhydro groups were blocked in the presence of 28 mM iodoacetamide in 50% acetonitrile for

30 min at 24°C in absence of light. Gel pieces were washed with 50% acetonitrile and after partial dehydration, the excised CEL I band was macerated in a solution of high purity trypsin (Promega). The

5 proteolytic digestion was allowed to continue at 37°C for 16 h. The resulting peptides were eluted from gel pieces with a 50% acetonitrile and 0.1 % tri-fluoro-acetic acid (TFA) concentrated in a SpeedVacTM. The peptides were analyzed by MALDI-TOF. Mixed tryptic digests were

10 crystallized in a matrix of α -cyano-4-hydroxycinnamic acid and analyzed by using a PerSeptive Biosystem DE-STR MALDI-TOF mass spectrometer equipped with delayed extraction operated in the reflector-positive ion mode and accelerating voltage of 20 kV. Expected theoretical

15 masses were calculated by MS-digest (Protein Prospector) or GPMW program (Lighthouse Data, Odense, Denmark). For tandem mass spectrometry (nano electrospray ionization (ESI), peptide samples were diluted with 5% acetonitrile/0.1% formic acid and subjected to LC MS/MS,

20 analyzed on a quadropole orthogonal time-of-flight mass spectrometry instrument (micromass, inc., Manchester, UK). The data were processed by Mslynx and database was searched by SonarTM.

25 Virally expressed, recombinant CEL I was secreted to the IF. Clarified IF-extracted material was used to purify the His-tag CEL I activity. CEL I was purified using one step Ni²⁺ affinity chromatography separation. A highly purified homogeneous single protein band was

30 purified as determined by Coomassie stained SDS-PAGE and mass spectrometry. The size of mature proteins and percent glycosylation concur with what has been reported for the CEL I protein isolated from celery (Yang et al., 2000). The purified CEL I has an average molecular mass

of 40 kD as determined by MALDI-TOF mass spectrometry, indicates 23.5% glycosylation by mass. CEL I has four potential glycosylation sites at amino acid positions 58, 116, 134, and 208. A mono-isotopic mass of 2152.6086 (2152.0068 Theoretical) Da corresponding to the mass of the peptide 107-125 (K)**DMCVAGAIQNFTSQLGHFR**(H) (SEQ ID NO: 35) that was recovered by MALDI-TOF, indicates that asparagine 116 is not glycosylated. Together, these gel analyses and mass spectrometry data indicate that a significant fraction of the CEL I protein was recoverable, and that the protein was correctly processed in the *N. benthamiana* plant.

For subsequent experiments, the 6-His tagged CEL I enzyme was produced using p1177MP4-CEL I 6His-A9. This clone was transcribed and inoculated onto *N. benthamiana* plants, which were harvested 8 days post infection. The plant material was combined with 2 volumes of extraction buffer (500 mM NaCl, 100 mM NaPi, 25 mM Tris pH 8.0, 7mM Beta-mercaptoethanol, 2mM PMSF) and vacuum infiltrated. Following buffer infiltration the tissue was macerated in a juice extractor, the resulting green juice adjusted to 4% w/v polyethyleneglycol, and let stand at 4°C for one hour. The green juice was clarified by either centrifugation at low speed (3500 x g) for 20 minutes or combined with perlite (2% w/v) and filtered through a 1.2 µm filter. The tagged CEL I can be selectively purified from the clarified green juice by metal affinity chromatography. The green juice was either combined with nickel-NTA resin, and batch binding of the CEL I performed, or purification was performed in column format, where the green juice was permitted to flow through a bed of nickel-NTA resin. For binding, the clarified green juice was adjusted to 10% w/v glycerol

and 10 mM imidazole. Following binding the resin was washed extensively with wash buffer (330 mM NaCl, 100 mM NaPi, pH 8.0, 10 mM imidazole) and the bound CEL I enzyme eluted from the nickel-NTA resin in 2 resin-bed volumes of 1X phosphate-buffered saline (PBS) containing 400 mM imidazole. The CEL I preparation was subsequently dialyzed against 1X PBS to remove the imidazole, assayed for activity, and stored at 4 °C or at -20 °C with or without glycerol until use.

10

Example 2

Cleavage of Mismatched DNA Substrate by CEL I

15 This example teaches the preparation of CEL I enzyme and its use in the cleavage of mismatched DNA substrate.

CEL I enzyme was prepared from celery stalks using the homogenization, ammonium sulfate, and Concanavalin A-Sepharose protocol described by Yang et al. (Biochemistry, 39:3533-3541 (2000)).

20 A 1.5 kg sample of chilled celery stalks was homogenized with a juice extractor. One liter of juice was collected, adjusted to 100 mM Tris-HCL, pH 7.7 with 100 micromolar phenylmethylsulfonyl fluoride (PMSF), and filtered through two layers of miracloth. Solid (NH₄)₂SO₄ was slowly added to 25% saturation while stirring on ice. After 30 minutes, the suspension was centrifuged at 27,000g for 1.5 hours at 4°C. The supernatants were collected and adjusted with solid (NH₄)₂SO₄ to 80% saturation while stirring on ice followed by centrifugation at 27,000g for 2 hours. The pellets were re-suspended in buffer B (0.1 M Tris-HCL, pH 7.7, 0.5 M KCl, 100 micromolar PMSF) and dialyzed against the

25
30

same buffer.

Concanavalin A (ConA) Sepharose affinity chromatography was performed by first incubating the
5 dialyzed sample with 2 ml of ConA resin overnight with gentle agitation. The ConA resin was then packed into a 0.5 cm diameter column and washed with several column volumes of buffer B. Elution was performed using 0.3 M alpha-methyl-mannoside in buffer B. Fractions were
10 collected in 1 ml aliquots. Fractions were assayed for mismatch cleavage activity on a radiolabeled mismatch substrate by incubating 0.1 microliter of each fraction with the mismatched probe in buffer D (20 mM Tris-HCL, pH 7.4, 25 mM KCL, 10 mM MgCl₂) for 30 minutes at 45°C as
15 described by Oleykowski et al. (Nucleic Acids Research 26: 4597-4602 (1998)).

Reaction products were visualized by separation on 10% TBE-PAGE gels containing 7% urea (Invitrogen), followed by autoradiography. Aliquots of the CEL I fractions
20 having mismatch cleavage activity were stored frozen at -20°C. A series of five-fold dilutions of CEL I fraction #5 were then analyzed for mismatch cleavage of radiolabeled mismatch substrate. Reactions were performed either in buffer D, New England BioLabs (NEB)
25 T4 DNA ligase buffer (50 mM Tris-HCL, pH 7.5, 10 mM MgCl₂, 10 mM dithiothreitol (DTT), 1 mM ATP, 25 microgram/ml BSA), or Gibco/BRL T4 DNA ligase buffer (50 mM Tris-HCL, pH 7.6, 10mM MgCl₂, 1 mM DTT, 1 mM ATP, 5%(w/v) polyethylene glycol-8000). Reaction products were
30 visualized as above. Cleavage activity in buffer D and in NEB T4 DNA ligase buffer were found to be roughly equivalent, whereas cleavage in the PEG-containing Gibco/BRL ligase buffer was enhanced by five to ten-fold compared to the other buffers.

Additional analysis of CEL I activity was carried out using defined heteroduplex DNAs from two different Green Fluorescent Protein (GFP) genes as substrate. This GFP heteroduplex substrate was prepared by annealing single stranded DNAs corresponding to cycle 3 GFP (SEQ ID NO: 30) on the sense strand and wild-type GFP (SEQ ID NO: 29) on the antisense strand. The single-stranded DNAs had been synthesized by asymmetric PCR and isolated by agarose gel electrophoresis. After annealing by heating to 90°C and cooling to room-temperature in the presence of 1X NEB restriction enzyme buffer 2 (10 mM Tris-HCL, pH 7.9, 10 mM MgCl₂, 50 mM NaCl, 1 mM dithiothreitol), the heteroduplex DNA was isolated by agarose gel electrophoresis followed by excision of the heterduplex band and extraction using Qiaquick DNA spin columns. A total of twenty eight mismatches, one or two nucleotides in length, occur throughout the length of the heteroduplex molecule. The distribution of the mismatches ranges from small clusters of several mismatches separated by one or two nucleotides to mismatches separated by more than thirty base pairs on either side.

A series of three-fold dilutions of CEL I in 1X NEB T4 DNA ligase buffer were prepared and one microliter aliquots of each were incubated in two separate series of 10 microliter reactions, each containing as substrate either 0.5 microgram of a supercoiled plasmid preparation or one hundred nanograms of the cycle3/wild-type GFP heteroduplex. All reactions took place in 1X NEB T4 DNA ligase buffer. Reactions were incubated at 45°C for 30 minutes and run on 1.5% TBE-agarose gel in the presence of ethidium bromide.

Treatment of the supercoiled plasmid preparation with increasing amounts of CEL I resulted in the conversion of supercoiled DNA to nicked circular, then linear molecules, and then to smaller fragments of DNA of random size. Treatment of the mismatched GFP substrate with the CEL I preparation resulted in the digestion of the full-length heteroduplex into ladder DNA bands which are likely to represent cleavage on opposite DNA strands in the vicinity of clusters of mismatches. Further digestion resulted in the conversion of the mismatched GFP substrate to smaller DNAs that may represent a limit digest of the heteroduplex DNA by the CEL I preparation.

15

EXAMPLE 3**Use of Cloned CEL I in the GRAMMR Reaction**

This example teaches that CEL I from a cloned source can be used in place of native CEL I enzyme purified from celery in Genetic Reassortment By DNA Mismatch Resolution without any noticeable change in results.

The cDNA of *CelI* was cloned from celery RNA. The gene was inserted into a TMV viral vector and expressed. Transcripts of the construct were used to infect *Nicotiana benthamiana* plants. Infected tissue was harvested, and the CEL I enzyme was purified. The results of the GRAMMR reaction obtained using the purified enzyme were compared to those using CEL I purified from celery, and were found to be similar.

Reactions were set up using twenty-one nanograms of the circular double-stranded heteroduplexed GFP plasmid substrate, as described in EXAMPLE 3, in ten microliters

containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.2 units T4 DNA ligase (Gibco/BRL), 1 unit of T4 DNA polymerase, and either 1.0 microliter of CEL I purified from celery (fraction 5, described in EXAMPLE 2), or 0.3
5 microliters of CEL I purified from a cloned source. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and
10 examined by RFLP analysis followed by sequence analysis of the GFP gene sequences. The RFLP results showed that reassortment of restriction sites, that is, GRAMMR had occurred in both celery-derived CEL I, as well as cloned CEL I-containing reactions. DNA sequence analysis
15 confirmed these results. Therefore, the data shows CEL I from a cloned source can be used in lieu of CEL I from celery for GRAMMR. In addition, the data demonstrates that it is CEL I activity that is part of the GRAMMR reaction, rather than a coincidental effect resulting
20 from the purifying steps used in extracting CEL I from celery.

EXAMPLE 4

Cloning, Expression And Use

25 Of RES I Endonuclease

.This example teaches the construction of a cDNA library from *Selaginella lepidophylla*, the identification of a nucleic acid sequence from the library that encodes
30 an endonuclease, and the expression of the new endonuclease, herein designated as "RES I." RNA was extracted from tissues of the resurrection plant, *Selaginella lepidophylla*, using the Trizol method, and oligo-dT primed cDNA that was prepared using standard

methodology. Resulting cDNAs were ligated into a GENEWARE®-based cloning vector and the ligation products were transformed into competent *E. coli* cells. Bacterial colonies containing GENEWARE® cDNA clones were picked at
5 random and grown as liquid cultures prior to DNA prepping and determination of the cloned cDNA sequences. The sequence files for the cloned *Selaginella* cDNAs were loaded into a database which was then searched by BLAST analysis for sequences that had similarity to the DNA
10 sequence of the *CellI* gene. BLAST analysis was also performed on other DNA sequence databases containing sequences of cDNAs obtained from other species.

BLAST hits that showed some level of homology to the
15 celery *CellI* sequence were identified in libraries from several species and the corresponding GENEWARE®-cDNA clones were re-arrayed into a single set of GENEWARE®-cDNA clones. This set of cDNA clones was then transcribed in vitro to generate infectious GENEWARE®
20 transcripts which were then inoculated onto leaves on *Nicotiana benthamiana* plants for expression analysis of the cDNA sequences encoded within the GENEWARE® viral genome. At seven days post-inoculation, leaf samples were taken from the infected plants and homogenized in
25 two volumes of water. The extracts were then assayed for supercoiled DNA nicking and cleavage activity.

Each supercoiled DNA nicking assay was performed in a reaction containing 0.5 micrograms of a supercoiled
30 plasmid prep of a pUC19-derivative in 1X NEB T4 DNA ligase buffer in a total volume of 10 microliters. The amounts of plant extract added to the reactions were 1 microliter, 0.33 microliter, or 0.011 microliter, incubated at 37°C for 30 minutes, and run on a 1% TAE-

agarose gel in the presence of Gelstar fluorescent DNA staining reagent. Little or no nicking activity was detected in uninfected plant extracts whereas only extracts from plants infected with GENEWARE® constructs containing cDNAs for a single gene from *Selaginella lepidophylla* showed appreciable amounts of activity against the plasmid DNA substrate.

The complete gene sequences of these clones were determined and PCR primers were designed to amplify the open reading frame minus any non-coding 5' and 3' sequences and to add a six histidine tail to the C-terminus of the encoded protein. The primers were then used to amplify the ORF from one of the active full-length *Selaginella* clones. The resulting PCR product was then cloned into the GENEWARE® vector pDN4 between the PacI and AvrII sites for expression *in planta*. The resulting clone, pLSB2225, which contains the *ResI* ORF (SEQ ID NO: 16), and which encodes the RES I protein (SEQ ID NO: 34), was sequenced to confirm that the gene had been inserted correctly, and then transcribed *in vitro* followed by inoculation of the infectious transcripts onto *N. benthamiana* plants. Seven days post inoculation, infected plant extracts were made as above and assayed for supercoiled DNA nicking and digestion activity to confirm the activity of the cloned enzyme.

Each supercoiled DNA nicking assay was performed in a reaction containing 0.5 micrograms of a supercoiled plasmid prep of a pUC19-derivative in 1X NEB *E. coli* DNA ligase buffer in the presence of 50 mM KCl in a total volume of 10 microliters. The amounts of plant extract added to the reactions were 0.2 microliter, 0.04 microliter, 0.008 microliter, or 0.0016 microliter,

incubated at 37°C for 30 minutes, and run on a 0.8% TAE-
agarose gel in the presence of Gelstar fluorescent DNA
staining reagent. Little or no nicking activity was
detected in uninfected plant extracts whereas extracts
5 from plants infected with the GENEWARE®-*Selaginella*
construct pLSB2225 showed appreciable amounts of activity
against the plasmid DNA substrate.

After positive results were obtained in that assay,
10 extracts of pLSB2225 infected plants were used in a
GRAMMR reaction to test the ability of this enzyme to
operate as a component of the mismatch resolution
reaction in place of the GENEWARE®-produced CEL I enzyme.

15

EXAMPLE 5

Use of RES I in the GRAMMR Reaction

This example teaches that RES I can be used in place
of native CEL I enzyme purified from celery in Genetic
20 Reassortment By DNA Mismatch Resolution without any
noticeable change in results.

GRAMMR was performed between the wild-type *Aequorea*
victoria GFP gene (Prasher, et al., Gen111(92)229) in a
25 pBS derivative (Stratagene, La Jolla, CA) encoded by
pBSWTGFP (SEQ ID NO:31) and a variant with mutations to
increase fluorescence intensity in *E. coli*, and to alter
the emission wavelength to blue light emission (Cramer,
et al., Nat Biotechnol 14(96)315; Heim et al.,
30 PNAS91(94)12501; Yang, et al., J Biol Chem 273(98)8212).
This variant gene (SEQ ID NO: 33), encoded by the plasmid
pBSC3BFP, as shown in Figure 5 (SEQ ID NO: 32), encodes a
fluorescent protein that emits bright blue light when
excited by longwave UV light.

The GRAMMR reactions were performed on GFP/c3BFP heteroduplexes in a circular, double-stranded plasmid DNA context. The circular, whole-plasmid heteroduplex DNA substrates were prepared by first linearizing pBSWTGFP (SEQ ID NO:31) and pBSC3BFP (Figure 5, SEQ ID NO: 32) by digestion with Kpn I and NgoM IV, respectively, then purifying the digested DNA using DNA spin columns. Next, 200 nanograms of each of the two linearized plasmids were mixed and brought to 1X SSPE (180 nM NaCl, 10 mM NaH₂PO₄, 1mM EDTA at pH 7.4) in a volume of 20 microliters. The mixture was then incubated at 95 degrees Celsius for 4 minutes, plunged into icewater where it remained for 10 minutes prior to incubation at 37 degrees Celsius. After 30 minutes, the annealed DNA sample was then transferred back to ice where it was held until use in GRAMMR reactions.

Two independent series of shuffling reactions were performed to compare CEL I with RES I in their abilities to facilitate sequence shuffling by GRAMMR. Each GRAMMR reaction contained 1 unit of T4 DNA polymerase, 2 units of *E. coli* DNA ligase, and 5 nanomoles of each dNTP in 1X NEB *E. coli* ligase buffer supplemented with KCl to 50 mM. Two separate enzyme dilution series were then performed. To each of two series of tubes containing aliquots of the above cocktail, one microliter aliquots of GENEWARE®-expressed CEL I or RES I extracts at dilutions of 1/3, 1/9, 1/27, 1/81, or 1/243 were added. An endonuclease-free control reaction was also prepared. To each of the reactions, one microliter aliquots containing 20 nanograms of the annealed DNA heteroduplex substrate were added and the reactions incubated at room temperature for

one hour and on ice for 30 minutes prior to transformation into competent *E. coli*.

Green fluorescent protein (GFP) and blue fluorescent
5 protein (BFP) could be visualized in the resulting colonies by long wave UV illumination. The parental wild-type GFP has dim green fluorescence, and the parental c3BFP gave bright blue fluorescence. In the genes encoding these fluorescent proteins, the sequences
10 that determine the emission color and those that govern fluorescence intensity are at different positions from one another. It is expected that DNA shuffling would result in the "de-linking" of the sequences that determine the emission color from those that govern
15 fluorescence intensity. As a consequence, the resultant progeny would be expected to exhibit reassortment of the functional properties of emission color and intensity. Therefore a measure of the extent of the DNA shuffling that had taken place in each reaction could be scored by
20 examining the color and intensity of fluorescence from the bacterial colonies on the corresponding plates. In the zero-nuclease control, only dim green and bright blue colonies were observed. However, on plates with cells transformed with DNAs from the reactions containing
25 either CEL I or RES I, some bright green as well as some dim blue colonies were observed, indicating that shuffling of DNA sequences had taken place. DNA sequence analysis confirmed that this was indeed the case and that on average, the recovery of shuffled clones was greater
30 than 85% for both CEL I and RES I and that the number and distribution of information transfer events was similar for both enzymes. However, it appeared that the activity of RES I in this experiment was several-fold higher than that of CEL I, as indicated by the low transformation

efficiency of reactions treated with the higher concentrations of the RES I preparation.

EXAMPLE 6

5 **Conservation of Full Length GFP Gene with Mismatch Resolution Cocktails**

This example teaches various mismatch resolution cocktails that conserve the full length GFP Gene.

10

Mismatched GFP substrate was treated with various concentrations of CEL I in the presence of cocktails of enzymes that together constitute a synthetic mismatch resolution system. The enzymes used were CEL I, T4 DNA polymerase, Taq DNA polymerase and T4 DNA ligase. CEL I activity should nick the heteroduplex 3' of mismatched bases. T4 DNA polymerase contains 3'-5' proofreading activity for excision of the mismatched base from the nicked heteroduplex. T4 DNA polymerase and Taq DNA polymerase contain DNA polymerase capable of filling the gap. T4 DNA ligase seals the nick in the repaired molecule. Taq DNA polymerase also has 5' flap-ase activity.

25 Matrix experiments were performed to identify the reaction conditions that would serve to resolve mismatches in the GFP heteroduplex substrate. In one experiment, cycle 3/wild-type GFP heteroduplex was incubated in a matrix format with serial dilutions of CEL
30 I fraction number five (described above) at eight different concentrations. Each reaction contained 100 nanograms of heteroduplex substrate and 0.2 microliters of T4 DNA ligase (Gibco BRL) in 1X NEBT4 DNA ligase buffer and dNTPs at 250 micromolar each, in a reaction

volume of 10 microliters. In all, the matrix contained 96 individual reactions. One full set of reactions was incubated at room temperature for 30 minutes while another full set was incubated at 37°C for 30 minutes.

5

After incubation, PCR was used to amplify the GFP gene from each reaction. Aliquots from each PCR were then digested with *Hind*III and *Hpa*I and electrophoresed on 3% agarose gels with ethidium bromide. Only cycle 3
10 GFP has a *Hind*III site and only wild-type encodes an *Hpa*I site.

If DNA mismatch resolution occurred at either the *Hind*III or *Hpa*I mismatched sites, then a proportion of
15 the PCR product would be expected to contain both sites, yielding a novel band. The band was observed in all samples, including the negative control samples that had neither CEL I, nor T4 DNA polymerase, nor Taq DNA polymerase. The results suggested that a basal level of
20 background recombination might have occurred at some point in the experiment other than in the GRAMMR reaction; possibly in the PCR step. PCR-mediated recombination is known to occur at some frequency between related sequences during amplification Paabo, et al., J
25 Biol Chem 265(90)4718-4721.

In another experiment, 200 nanograms of cycle 3/wild-type GFP heteroduplex was treated with CEL I and T4 DNA polymerase in various concentrations along with
30 2.5 units of Taq DNA polymerase in the presence or absence of T4 DNA ligase (0.2 units; Gibco BRL). Each reaction contained 1X NEB T4 DNA ligase buffer with 0.05 mM each dNTP in a final volume of 20 microliters. Reactions were incubated for 30 minutes at 37°C and 10

microliters were run on a 2% TBE-agarose gel in the presence of ethidium bromide. Results showed that in the presence of DNA ligase, but in the absence of T4 DNA polymerase, increasing amounts of CEL I caused greater degradation of the heteroduplexed DNA, but that this effect could be counteracted by increasing the amount of T4 DNA polymerase in the reaction. These results indicated that the various components of the complete reaction could act together to conserve the integrity of the full-length gene through DNA mismatch resolution.

Another matrix experiment was conducted to expand on these results and to identify additional conditions for DNA mismatch resolution for this synthetic system. 60 nanograms of cycle3/wild-type GFP heteroduplex were treated with CEL I and T4 DNA polymerase at various concentrations in the presence of 2.5 units of Taq DNA polymerase and 0.2 units of T4 DNA ligase in 1 X NEB T4 DNA ligase buffer containing 0.5 mM of each dNTP in a reaction volume of 10 microliters. Each set of reactions was incubated for 1 hour at 20°C, 30°C, 37°C, or 45°C. All reactions were then run on a 1.5% TBE-agarose gels in the presence of ethidium bromide. The results showed that the GFP heteroduplex was cleaved into discrete fragments by the CEL I preparation alone. The success of DNA mismatch resolution was initially gauged by the degree to which the apparent full-length integrity of the GFP sequence was maintained by the other components of the mismatch resolution system in the presence of CEL I. Conditions of enzyme concentration and temperature were identified that conserved a high proportion of the DNA as full-length molecules in this assay. Namely, one microliter of the CEL I fraction five preparation (described in EXAMPLE 2) with one microliter (1 unit) of

the T4 DNA polymerase in the presence of the other reaction components which were held constant in the experiment. It was found that as the reaction temperature increased, the degradative activity of CEL I increased accordingly. Furthermore, it was shown that the other components of the repair reaction acted to conserve the integrity of the full-length DNA at 20°C, 30°C, and 37°C, but was remarkably less efficient at conserving the full-length DNA at 45°C. From these results, we concluded that under these experimental conditions, incubation at 45°C was not optimal for the process of GRAMMR, and that incubation at 20°C, 30°C, and 37°C were permissible.

15

EXAMPLE 7**Restoration of Restriction Sites to GFP Heteroduplex DNA after DNA Mismatch Resolution (GRAMMR)**

This experiment teaches the operability of genetic reassortment by DNA mismatch resolution (GRAMMR) by demonstrating the restoration of restriction sites.

The full-length products of a twenty-fold scale-up of the GRAMMR reaction, performed at 37°C for one hour, using the optimal conditions found above (the 1x reaction contained sixty nanograms of heteroduplex DNA, one microliter of CEL I fraction five (described in EXAMPLE 2), one unit T4 DNA polymerase in the presence of 2.5 units of Taq DNA polymerase and 0.2 units of T4 DNA ligase in 1 X NEB T4 DNA ligase buffer containing 0.5 mM of each dNTP in a reaction volume of 10 microliters) were gel-isolated and subjected to restriction analysis by endonucleases whose recognition sites overlap with

mismatches in the GFP heteroduplex, thereby rendering those sites in the DNA resistant to restriction enzyme cleavage. The enzymes used were *Bam*HI, *Hind*III, *Hpa*I, and *Xho*I. Negative controls consisted of untreated GFP
5 heteroduplex. Positive controls consisted of Cycle 3 or wild type GFP sequences, individually. All controls were digested with the same enzymes as the product of the DNA mismatch resolution reaction. All samples were run on a 2% TBE-agarose gel in the presence of ethidium bromide.

10

After treatment with the mismatch resolution cocktail, a proportion of the DNA gained sensitivity to *Bam*HI and *Xho*I restriction endonucleases, indicating that DNA mismatch resolution had occurred. The *Hpa*I-cut
15 samples could not be interpreted since a low level of cleavage occurred in the negative control. The *Hind*III, *Bam*HI and *Xho*I sites displayed different degrees of cleavage in the GRAMMR-treated samples. Restoration of the *Xho*I site was more extensive than that of the *Bam*HI
20 site, which was in turn, more extensive than restoration at *Hind*III site.

The extent to which cleavage occurs is indicative of the extent to which mismatches in the DNA have been
25 resolved at that site. Differences in mismatch resolution efficiency may relate to the nature or density of mismatches present at those sites. For example, the *Xho*I site spans a three-mismatch cluster, whereas the *Bam*HI site spans two mismatches and the *Hind*III site
30 spans a single mismatch.

EXAMPLE 8**GRAMMR-treated GFP Genes**

5 This example demonstrates that GRAMMR can reassort sequence variation between two gene sequences in a heteroduplex and that there are no significant differences in GRAMMR products that were directly cloned, or PCR amplified prior to cloning.

10 The GRAMMR-treated DNA molecules of EXAMPLE 7 were subsequently either directly cloned by ligation into pCR-Blunt II-TOPO (Invitrogen), or amplified by PCR and ligated into pCR-Blunt II-TOPO according to the manufacturer's instructions, followed by transformation
15 into *E. coli*. After picking individual colonies and growing in liquid culture, DNA was prepared and the sequences of the GFP inserts were determined. As negative controls, the untreated GFP heteroduplex substrate was either directly cloned or PCR amplified
20 prior to cloning into the plasmid.

In GRAMMR, reassortment of sequence information results from a process of information transfer from one strand to the other. These sites of information transfer
25 are analogous to crossover events that occur in recombination-based DNA shuffling methods. For the purposes of relating the results of these reassortment experiments, however, the GRAMMR output sequences are described in terms of crossovers. Sequences of twenty
30 full-length GFP clones that were derived from the GRAMMR-treated GFP genes were analyzed. Four of these clones were derived from DNA that had been directly cloned into pZeroBlunt (Invitrogen) following GRAMMR reaction (no PCR amplification). The other sixteen sequences were cloned

after PCR amplification. Analysis of these full-length GFP sequences revealed that all twenty sequences had undergone sequence reassortment having between one and ten crossovers per gene. A total of 99 crossovers were
5 found in this set of genes, giving an average of about 5 crossovers per gene. With the distance between the first and last mismatches of about 590 nucleotides, an overall frequency of roughly one crossover per 120 base-pairs was calculated. Within this set of twenty clones, a total of
10 seven point mutations had occurred within the sequences situated between the PCR primer sequences, yielding a mutation frequency of roughly 0.05%.

Thirty-five clones that had not been subjected to the GRAMMR reaction were sequenced. Of these controls, fourteen were derived from direct cloning and twenty-one were obtained after PCR amplification using the GFP heteroduplex as template. Of these thirty-five non-GRAMMR treated control clones, eight were recombinants,
20 ranging from one to three crossovers, with most being single crossover events. A total of twenty-five point mutations had occurred within the sequences situated between the PCR primers, yielding a mutation frequency of roughly 0.1%.

25

No significant differences were observed between the GRAMMR-treated products that were either directly cloned or PCR amplified. Notably, though, in the non-GRAMMR-treated controls, the frequency of recombinants was
30 higher in the PCR amplified DNAs than in the directly cloned DNAs. This higher frequency is consistent with results obtained by others in which a certain level of recombination was found to be caused by "jumping PCR." (Paabo, et al., DNA damage promotes jumping between

templates during enzymatic amplification. J Biol Chem
265(90)4718-4721).

EXAMPLE 9

5 **Heteroduplex Substrate Preparation for**
 Plasmid-on-Plasmid Genetic Reassortment
 By DNA Mismatch Resolution (POP GRAMMR) of GFP
 Plasmids

10 This example teaches that heteroduplex substrate for
Genetic Reassortment by DNA Mismatch Resolution can be in
the form of intact circular plasmids. Cycle 3-GFP and
wild-type GFP heteroduplex molecules were prepared
plasmid-on-plasmid (POP) format. In this format, the GFP
15 sequences were reassorted within the context of a
circular double-stranded plasmid vector backbone. This
made possible the recovery of the reassorted product by
direct transformation of *E. coli* using an aliquot of the
GRAMMR reaction. Consequently, neither PCR amplification
20 nor other additional manipulation of the GRAMMR-treated
DNA was necessary to obtain reassorted clones.

Mismatched DNA substrate for POP-GRAMMR reactions
was generated containing wild-type GFP (SEQ ID NO: 29)
25 and Cycle 3 GFP (SEQ ID NO: 30), resulting in the two
pBluescript-based plasmids, pBSWTGFP (SEQ ID NO: 31) and
pBSC3GFP (SEQ ID NO: 17), respectively. The GFPs were
inserted between the *KpnI* and *EcoRI* sites of the
pBluescript polylinker so that the only sequence
30 differences between the two plasmids occurred at sites
where the wild-type and Cycle 3 GFPs differ from one-
another. Both plasmids were linearized by digestion of
the plasmid backbone with *SapI*, cleaned up using a DNA
spin-column, mixed, amended to 1X PCR buffer (Barnes,

1994; *PNAS*, **91**, 2216-2220), heated in a boiling water bath for three minutes, and slow-cooled to room temperature to anneal the denatured DNA strands. Denaturing and annealing these DNAs led to a mixture of duplexes; the re-formation of parental duplexes, and the formation of heteroduplexes from the annealing of strands from each of the two input plasmids. Parental duplexes were deemed undesirable for GRAMMR and were removed by digestion with restriction enzymes that cut in one or the other parental duplex but not in the heteroduplexed molecules. *PmlI* and *XhoI* were chosen for this operation since *PmlI* cuts only in the wild-type GFP sequence and *XhoI* cuts only Cycle 3 GFP. After treatment with these enzymes, the products were resolved on an agarose gel. The full-length, uncut heteroduplex molecules were resolved from the *PmlI*- and *XhoI*-cut parental homoduplexes in an agarose gel and purified by excision of the band and purification with a DNA spin column.

The resulting population of heteroduplexed molecules was treated with DNA ligase to convert the linear DNA into circular, double-stranded DNA heteroduplexes. After confirmation by agarose gel-shift analysis, the circular double-stranded GFP heteroduplexed plasmid was used as substrate for GRAMMR reactions. Examples of the resulting clones are included as SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

EXAMPLE 10

Exemplary Reaction Parameters for Genetic Reassortment by DNA Mismatch Resolution CEL I and T4 DNA Polymerase Concentrations Compared

The GRAMMR reaction involves the interaction of

numerous enzymatic activities. Several parameters associated with the GRAMMR reaction were examined, such as CEL I concentration, T4 DNA polymerase concentration, reaction temperature, substitution of T4 DNA polymerase with T7 DNA polymerase, the presence of Taq DNA polymerase, and the source of the CEL I enzyme. A matrix of three different CEL I concentrations versus two concentrations of T4 DNA polymerase was set up to examine the limits of the *in vitro* DNA mismatch resolution reaction.

Twenty-one nanograms (21ng) of the circular double-stranded heteroduplexed plasmid, prepared as described in EXAMPLE 9, was used as substrate in a series of ten microliter reactions containing 1X NEB ligase buffer, 0.5 mM each dNTP, 1.0 unit Taq DNA polymerase, 0.2 units T4 DNA ligase (Gibco/BRL), either 1.0 or 0.2 units T4 DNA polymerase, and either 0.3, 0.1, or 0.03 microliters of a CEL I preparation (fraction 5, described in EXAMPLE 2). Six reactions representing all six combinations of the two T4 DNA polymerase concentrations with the three CEL I concentrations were prepared, split into equivalent sets of five microliters, and incubated at either 20 degrees C or 37 degrees C. A control reaction containing no CEL I and 0.2 unit of T4 DNA polymerase with the other reaction components was prepared and incubated at 37 degrees C. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by restriction fragment length polymorphism analysis (RFLP) followed by sequence analysis of the GFP gene sequences. RFLP analysis was based on differences in several restriction enzyme recognition sites between

the wild-type and Cycle 3 GFP genes. The RFLP results showed that throughout the CEL I /T4 DNA polymerase/temperature matrix, reassortment of restriction sites, that is GRAMMR, had occurred, and that
5 no such reassortment had occurred in the zero CEL I control clones. DNA sequence analysis confirmed that reassortment had occurred in all of the CEL I -containing samples. Sequencing also confirmed that the zero- CEL I controls were not reassorted, with the exception of a
10 single clone of the 16 control clones, which had a single-base change from one gene sequence to the other, presumably resulting either from repair in *E. coli* or from random mutation. The sequences of several exemplary GRAMMR-output GFP clones are shown; all of which came
15 from the reaction containing 0.3 microliters of the CEL I preparation and 1.0 unit of T4 DNA polymerase incubated at 37 degrees C. The parental wild-type and Cycle 3 GFP genes are shown first for reference.

20

EXAMPLE 11**Taq DNA Polymerase is Not Required
for Genetic Reassortment by DNA Mismatch Resolution**

This experiment teaches that Taq DNA Polymerase does
25 not dramatically, if at all, contribute or interfere with the functioning of GRAMMR. Taq DNA polymerase is reported to have a 5' flap-ase activity, and had been included in the teachings of the previous examples as a safeguard against the possible formation and persistence
30 of undesirable 5' flaps in the heteroduplexed DNA undergoing the GRAMMR reaction.

GRAMMR reactions were set up, as in EXAMPLE 10, with twenty-one nanograms of the circular double-stranded

heteroduplexed GFP plasmid substrate in ten microliter reactions containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.2 units T4 DNA ligase, 1.0 unit T4 DNA polymerase, 1.0 microliter of a CEL I preparation (fraction 5, described in EXAMPLE 2), and either 2.5 units, 0.5 units of Taq DNA polymerase, or no Taq DNA polymerase. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by RFLP analysis followed by sequence analysis of the GFP gene sequences. The RFLP results showed that reassortment of restriction sites, that is, GRAMMR, had occurred both in the presence and the absence of Taq DNA polymerase in the GRAMMR reaction. DNA sequence analysis confirmed these results. Therefore, the data shows that Taq DNA polymerase was unnecessary for GRAMMR.

20

EXAMPLE 12**Alternate Proofreading DNA Polymerases
for Genetic Reassortment by DNA Mismatch Resolution**

This experiment teaches that Genetic Reassortment by DNA Mismatch Resolution is not limited to the use of T4 DNA polymerase, and that alternate DNA polymerases can be substituted for it.

Reactions were set up, as in EXAMPLE 10, with twenty-one nanograms of the circular double-stranded heteroduplexed GFP plasmid substrate in ten microliter reactions containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.2 units T4 DNA ligase (Gibco/BRL), 10 units or 2 units of T7 DNA polymerase, 1.0 microliter of a CEL I

preparation (fraction 5, described in EXAMPLE 2), and 2.5 units of Taq DNA polymerase. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by RFLP analysis followed by sequence analysis of the GFP gene sequences. The RFLP results showed that reassortment of restriction sites, that is GRAMMR, had occurred in both T7 DNA polymerase-containing reactions. DNA sequence analysis confirmed these results. Therefore, the data shows that T7 DNA polymerase can substitute for T4 DNA polymerase for GRAMMR. In addition, it shows that individual components and functionalities can be broadly substituted in GRAMMR, while still obtaining similar results.

EXAMPLE 13

Molecular Breeding of Tobamovirus 30K Genes in a Viral Vector.

In the preceding examples, Genetic Reassortment by DNA Mismatch Resolution has been taught to be useful for reassorting sequences that are highly homologous, for example, wtGFP and Cycle 3 GFP are 96% identical. The present example teaches that GRAMMR can be used to reassort more divergent nucleic acid sequences, such as genes encoding tobamovirus movement protein genes.

Heteroduplexes of two tobamovirus movement protein (MP) genes that are approximately 75% identical were generated. The heteroduplex substrate was prepared by annealing partially-complementary single-stranded DNAs of opposite strandedness synthesized by asymmetric PCR; one

strand encoding the movement protein gene from the tobacco mosaic virus U1 type strain (TMV-U1) (SEQ ID NO: 9), and the other strand encoding the movement protein gene from tomato mosaic virus (ToMV) (SEQ ID NO: 10).

5 The sequences of the two partially complementary movement protein genes were flanked by 33 nucleotides of absolute complementarity to promote annealing of the DNAs at their termini and to facilitate PCR amplification and cloning. The annealing reaction took place by mixing 2.5

10 micrograms of each single-stranded DNA in a 150 microliter reaction containing 333 mM NaCl, 33 mM MgCl₂, 3.3 mM dithiothreitol, 166mM Tris-HCl, pH 7, and incubating at 95°C for one minute followed by slow cooling to room-temperature. GRAMMR was performed by incubating

15 5 microliters of the heteroduplex substrate in a 20 microliter reaction containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.4 units T4 DNA ligase (Gibco/BRL), 2.0 units of T4 DNA polymerase, and CEL I. The CEL I was from a cloned preparation and the amount that was used

20 varied from 2 microliters of the prep, followed by five serial 3-fold dilutions. A seventh preparation with no CEL I was prepared, which served as a control.

After one hour at room-temperature, DNA was purified

25 from the reactions using Strataprep spin DNA purification columns (Stratagene, LaJolla, CA) and used as templates for PCR reactions using primers designed to anneal to the flanking primer-binding sites of the two sequences. PCR products from each reaction were purified using

30 Strataprep columns, digested with AvrII and PacI, and ligated into the movement protein slot of similarly-cut pGENEWARE®-MP-Avr-Pac. This plasmid contained a full-length infectious tobamovirus-GFP clone modified with AvrII and PacI sites flanking the movement protein gene

to permit its replacement by other movement protein genes. After transformation of DH5-alpha *E. coli* and plating, colonies were picked, cultures grown, and DNA was extracted. The movement protein inserts were
5 subjected to DNA sequence analysis from both directions and the sequence data confirmed that in the majority of inserts derived from the GRAMMR-treated material were reassorted sequences made up of both TMV-U1 and ToMV movement protein gene sequences. The DNA sequences of
10 several exemplary GRAMMR output MP clones are shown as SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, and SEQ ID NO: 15.

Example 14

15 **Molecular Breeding of Highly Divergent
Tobamovirus 30K Genes in Viral Vectors using
Plasmid-on-Plasmid Genetic Reassortment
By DNA Mismatch Resolution (POP GRAMMR)**

20 EXAMPLE 4 taught the reassortment of movement protein (MP) genes from several divergent strains of tobamovirus (approximately 75% identical; cloned into the pGENEWARE-MP-Avr-Pac vector) using GRAMMR. This example teaches the use of Plasmid-on-plasmid GRAMMR (POP GRAMMR)
25 for reassorting even more highly divergent species.

Starting parental MP genes from the tobamoviruses TMV-Cg (Figure 6, SEQ ID NO: 18), TMV-Ob (Figure 7, SEQ ID NO: 19), TMV-U2 (Figure 8, SEQ ID NO: 20), TMV-U1 (SEQ
30 ID NO: 9), and tomato mosaic virus (ToMV) (SEQ ID NO: 10) were used. The plasmid of pGENEWARE-ToMV MP was linearized by digestion with Sma I. The plasmids of pGENEWARE containing the MP genes from either TMV-Cg, TMV-Ob, TMV-U2, or TMV-U1 were digested with Stu I. The

digested pGENEWARE-MP constructs were purified using DNA spin columns. The following heterduplex pairs were generated: pGENEWARE-Cg MP and pGENEWARE-ToMV MP, pGENEWARE-TMV-Ob MP and pGENEWARE-ToMV MP, pGENEWARE-TMV-U2 MP and pGENEWARE-ToMV MP, pGENEWARE-TMV-U1 MP and pGENEWARE-ToMV MP. The heteroduplexes of these MP gene sequences are approximately 47%, 58%, 62%, and 75% identical, respectively. Heteroduplex DNA was generated by mixing 200 nanograms of each of the two linearized plasmids in 1X SSPE (180 mM NaCl, 10 mM NaH₂PO₄, 1 mM EDTA, at pH 7.4) in a volume of 20 microliters. The mixture was incubated at 95 degrees Celsius for 4 minutes, plunged into ice water where it remained for 10 minutes prior to incubation at 37 degrees Celsius. After 30 minutes, the annealed DNA sample was then transferred back to ice where it was held until use in GRAMMR reactions.

Each 10 microliter GRAMMR reaction contained 1 unit of T4 DNA polymerase, 2 units of *E. coli* DNA ligase, and 0.5 mM of each dNTP in 1X NEB *E. coli* DNA ligase buffer supplemented with KCl to 50 mM. A one microliter aliquot of CEL I (diluted 1/3, 1/9, 1/27, 1/81, 1/243, or 1/729) was next added. An endonuclease-free control reaction was also prepared. To each of the reactions, a one microliter aliquot containing 20 nanograms of the annealed DNA heteroduplex substrate was added and the reactions were incubated at room temperature for one hour and on ice for 30 minutes prior to transformation into competent *E. coli*.

DNA sequence analysis was performed from both directions, and the sequence data showed that a significant number of clones derived from the GRAMMR-

treated material were reassorted sequences containing information from both parental movement protein gene sequences. The DNA sequences of several exemplary output pGENEWARE-MP clones from the GRAMMR reaction are shown as follows, TMV-Cg/ToMV clones, Figure 9, SEQ ID NO: 21, and Figure 10, SEQ ID NO: 22; TMV-Ob/ToMV clones, Figure 11, SEQ ID NO: 23, and Figure 12, SEQ ID NO: 24; TMV-U2/ToMV clones, Figure 13, SEQ ID NO: 25, and Figure 14, SEQ ID NO: 26; and TMV-U1/ToMV clones, Figure 15, SEQ ID NO: 27, and Figure 16, SEQ ID NO: 28.

Example 15

GRAMMR On Linearized DNA Substrate Using Endonucleases That Cleave Within A Selectable Marker

This example teaches a GRAMMR reaction where DNA substrate molecules are linearized with restriction endonucleases that cleave within a selectable marker gene.

GRAMMR is performed between the wild-type *Aequorea victoria* GFP gene (Prasher, et al., Gene111(92)229) in a pBS derivative (Stratagene, La Jolla, CA) encoded by pBSWTGFP (SEQ ID NO:31) and a variant with mutations to increase fluorescence intensity in *E. coli*, and to alter the emission wavelength to blue light emission (Cramer, et al., Nat Biotechnol 14 (96) 315; Heim et al., PNAS91(94)12501; Yang, et al., J Biol Chem 273(98)8212). This variant gene, encoded by the plasmid pBSC3BFP (SEQ ID NO: 32), encodes a fluorescent protein that emits bright blue light when excited by longwave UV light.

The GRAMMR reactions are performed on GFP/c3BFP heteroduplexes in a circular, double-stranded plasmid DNA context. The circular, whole-plasmid heteroduplex DNA substrates are prepared by first linearizing pBSWTGFP (SEQ ID NO:31) and pBSC3BFP (SEQ ID NO: 32) by digestion with *Ahd* I and *Bcg* I, respectively, then purifying the digested DNA using DNA spin columns. Next, 200 nanograms of each of the two linearized plasmids are mixed and brought to 1X SSPE (180 nM NaCl, 10 mM NaH₂PO₄, 1mM EDTA at pH 7.4) in a volume of 20 microliters. The mixture is then incubated at 95 degrees Celsius for 4 minutes, plunged into icewater where it remains for 10 minutes prior to incubation at 37 degrees Celsius. After 30 minutes, the annealed DNA sample is then transferred back to ice where it is held until use in GRAMMR reactions.

Two independent series of reassortment reactions are performed to compare CEL I with RES I in their abilities to facilitate sequence reassortment by GRAMMR. Each reaction is first treated for 10 minutes at room-temperature with 1 unit of T4 DNA polymerase in the presence of 5 nanomoles of each dNTP in 1X NEB E. coli ligase buffer supplemented with KCl to 50 mM. Subsequently, 2 units of E. coli DNA ligase are added. Two separate enzyme dilution series are then performed. To each of two series of tubes containing aliquots of the above cocktail, one microliter aliquots of GENEWARE®-expressed CEL I or RES I extracts at dilutions of 1/3, 1/9, 1/27, 1/81, or 1/243 are added. An endonuclease-free control reaction is also prepared. To each of the reactions, one microliter aliquots containing 20 nanograms of the annealed DNA heteroduplex substrate are added and the reactions incubated at room temperature for

one hour and on ice for 30 minutes prior to transformation into competent *E. coli*.

Green fluorescent protein (GFP) and blue fluorescent protein (BFP) is visualized in the resulting colonies by long wave UV illumination. The parental wild-type GFP gives dim green fluorescence, and the parental c3BFP gives bright blue fluorescence. In the genes encoding these fluorescent proteins, the sequences that determine the emission color and those that govern fluorescence intensity are at different positions from one another.

It is expected that DNA reassortment would result in the "de-linking" of the sequences that determine the emission color from those that govern fluorescence intensity. As a consequence, the resultant progeny would be expected to exhibit reassortment of the functional properties of emission color and intensity. Therefore a measure of the extent of the DNA reassortment that had taken place in each reaction can be scored by examining the color and intensity of fluorescence from the bacterial colonies on the corresponding plates.

EXAMPLE 18

25 **Detection of DNA mutations Using RES I and Multiplex Analysis**

The sensitivity of RES I for mismatch detection is illustrated by its ability to detect mutations in pooled DNA samples. DNA is obtained from peripheral blood lymphocytes from individuals undergoing genetic screening. Samples are obtained from breast cancer-only, ovarian cancer-only, breast/ovarian cancer syndrome families or from non-breast/ovarian cancer control

samples. Unlabeled primers specific for exon 2 of BRCA1 are utilized to PCR amplify this region of the gene. The wild-type PCR products of exon 2 are labeled with γ - ^{32}P -ATP. Briefly, 10 picomoles of PCR product are purified
5 by the Wizard procedure (Promega). Exon 2 wild-type products are then phosphorylated using T4 kinase and 15 picomoles of γ - ^{32}P -ATP at 6,000 Ci/mmol in 30 μl 1X kinase buffer (70 mM Tris-HCl (pH 7.6), 10 mM MgCl_2 , 5 mM dithiothreitol) at 37°C. for 1 hour. The reactions are
10 stopped with 1 μl 0.5M EDTA. The reaction volume is brought up to 50 μl with 1X STE buffer (100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM EDTA) and processed through a Pharmacia Probe Quant column. Labeled DNA (1 pmol/ μl in 100 μl) is then used for hybridization with individual
15 unlabeled PCR amplified experimental samples. For each individual sample, 100 fmol of the unlabeled PCR amplified product is incubated with 200 fmol of the ^{32}P -labeled wild-type PCR product in RES I reaction buffer (25 mM KCl, 10 mM MgCl_2 , 20 mM Tris-HCl, pH 7.5).
20 Following denaturation and renaturation, heteroduplexed, radiolabeled PCR products are exposed to RES I for 30 minutes at 37°C in 1X RES reaction buffer and stopped via the addition of 10 μl stop mix (75% formamide, 47 mM EDTA, 1.5% SDS, xylene cyanol and bromophenol blue). The
25 heteroduplexes are treated with the enzyme individually or pooled in one sample tube and treated. The products of the reaction are loaded onto a 15% polyacrylamide gel containing 7M urea.

30 To further illustrate the ability of RES I to detect mutations in pooled DNA samples, 1, 2, 3, 5, 10 or 30 heteroduplexed, radiolabeled PCR products, (again amplified from exon 2 of the BRCA1 gene), are exposed to RES-I in a single reaction tube and the products run on a

6% polyacrylamide gel containing 7M urea. Samples are amplified and radiolabeled as described above. Each pool contains only one sample that has a mutation (AG deletion). The other samples in each pool are wild-type. Control samples are not exposed to RES I. In the pooled samples where a mutation is present, RES-I consistently cleaves the PCR products illustrating the sensitivity of the enzyme in the presence of excess wild-type, non-mutated DNA. As a control, heteroduplexed PCR products containing no mutations are analyzed and no cut band corresponding to a mutation appears.

EXAMPLE 17

Detection of Mutations and Polymorphisms by RES-I in Samples Obtained from High Risk Families

PCR primer sets specific for the exons in the BRCA1 gene are synthesized. The gene sequence of BRCA1 is known. The exon boundaries and corresponding base numbers are shown in table II. Primers to amplify desired sequences can be readily designed by those skilled in the art following the methodology set forth in Current Protocols in Molecular Biology, Ausubel et al., eds, John Wiley and Sons, Inc. (1995). These primers are planned such that in each PCR reaction, one primer is labeled at the 5' termini with a fluorescent-label, 6-FAM, while the other primer is similarly labeled with a label of another color, TET. A PCR product will thus be labeled with two colors such that DNA nicking events in either strand can be observed independently and the measurements corroborated.

TABLE II

EXON BOUNDARIES AND CORRESPONDING BASE NUMBERS IN BRCA1		
5	EXON	BASE #'s
	1	1-100
	2	101-199
	3	200-253
10	5	254-331
	6	332-420
	7	421-560
	8	561-665
	9	666-712
15	10	713-788
	11	789-4215
	11B	789-1591
	11C	1454-2459
	11A	2248-3290
20	11D	3177-4215
	12	4216-4302
	13	4303-4476
	14	4477-4603
	15	4604-4794
25	16	4795-5105
	17	5106-5193
	18	5194-5273
	19	5274-5310
	20	5311-5396
30	21	5397-5451
	22	5452-5526
	23	5527-5586
	24	5587-5711

35

Peripheral blood samples from individuals in high-risk families are collected and the DNA isolated. The PCR products are amplified using Elongase (BRL) and purified using Wizard PCR Preps (Promega). The DNA is heated to 94°C and slowly cooled in 1X RES I buffer (20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM MgCl₂) to form heteroduplexes. The heteroduplexes are incubated in 20 µl 1X RES I buffer with 0.2 µl of RES I and 0.5 units of AmpliTaq at 45°C. for 30 minutes. The reactions are stopped with 1 mM phenanthroline and incubated for an

45

additional 10 minutes at 45°C. The sample is processed through a Centricon column (Princeton Separations) and dried down. One microliter of ABI loading buffer (25 mM EDTA, pH 8.0, 50 mg/ml Blue dextran), 4 µl deionized formamide and 0.5 µl TAMRA internal lane standard are added to the dried DNA pellet. The sample is heated at 90°C. for 2 minutes and then quenched on ice prior to loading. The sample is then loaded onto a 4.25% denaturing 34 cm well-to-read acrylamide gel and analyzed on an ABI 373 Sequencer using GENESCAN 672 software. The 6-FAM labeled primer in this experimental sample is at nucleotide 3177 of the BRCA1 cDNA (region 11D), the TET labeled primer is 73 nucleotides into the intron between exon 11 and exon 12. Each spike represents the presence of a DNA band produced by the cleavage of the heteroduplex by RES-I where a mutation or a polymorphism is present. One spike represents the size of the RES I produced fragment from the 3' side of the mismatch site to the 5' 6-FAM label of the top strand. The other spike represents the corresponding fragment in the bottom strand from the 3' side of the mismatch to the 5' TET label. The sum of the two fragments equals one base longer than the length of the PCR product. The 6-FAM panel shows a spike at base #645 from the 6-FAM label and the TET panel shows a spike at base #483 from the TET label, both corresponding to the site of the 5 base deletion at nucleotide 3819 of the BRCA1 cDNA.

Analysis of exon 11 in another individual is performed using a 6-FAM-labelled primer at nucleotide 1454 of the BRCA1 cDNA. The TET-labeled primer is at nucleotide 2459 (region 11C). The PCR amplified products are amplified and prepared as described above. In this individual, the 6-FAM panel shows a spike at base #700

and the TET panel shows a spike at #305, each spike corresponding to the site of RES I incision in the respective DNA strand at a nonsense mutation of A>T at nucleotide 2154 of the BRCA1 cDNA. The 6-FAM panel also shows a spike at base #747 and the TET panel shows a spike at #258 corresponding to the site of a polymorphism C>T at nucleotide 2201 of the BRCA1 cDNA. The nonsense mutation and polymorphism can be confirmed by sequencing the sample using the ABI 377 Sequencer.

10

Certain individuals have mutations in another region of exon 11, region 11A, on the schematic in FIG. n. A 6-FAM-labelled primer at nucleotide 2248 of the BRCA1 cDNA and a TET labeled primer at nucleotide 3290 are used to amplify this region of exon 11. Following amplification, the samples are processed as described above.

15

Deposits with The American Type Culture Collection (ATCC)

20

Three deposits have been made in the American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA. A deposit has been made of a plasmid DNA construct containing a derivative of tobacco mosaic virus and cDNA of the CEL I mismatch-endonuclease gene from celery, tagged with 6HIS. The construct is internally designated P1177MP4-CEL I 6HIS, and has been assigned ATCC Number PTA-3927. A deposit has been made of a plasmid DNA construct containing a derivative of tobacco mosaic virus and cDNA of the CEL I mismatch-endonuclease gene from celery. The construct is internally designated P1177MP4-CEL I Avr, and has been assigned ATCC Number PTA-3926. A deposit has been made of a plasmid DNA construct containing a derivative of tobacco mosaic virus

25

30

and a cDNA insert encoding a 34kDa protein from
Selaginella lepidophylla. The cDNA insert is referred to
as RES I-6HIS. RES I is a mismatch endonuclease gene.
The construct is internally designated pLSB-2225, and has
5 been assigned ATCC Number PTA-4562.

These deposits were made in accordance with the
terms and provisions of the Budapest Treaty relating to
deposit of microorganisms and was made for a term of at
10 least thirty (30) years and at least five (05) years
after the most recent request for the furnishing of a
sample of the deposit is received by the depository, or
for the effective term of a patent to issue from this
application or a subsequent application citing any of
15 these deposits, whichever is longer. Each deposit will
be replaced if it becomes non-viable during that period.

It should be noted that applicant's designations for
each of the clones were shortened in the deposit to the
20 aforementioned deposit with the American Type Culture
Collection, that is, p1177MP4- CEL I Avr-B3 is referred
to as p1177MP4- CEL I Avr; and p1177MP4- CEL I 6His-A9 is
referred to as p1177MP4-CEL I 6His. The clone p1177MP4-
CEL I Avr (SEQ ID NO:01) contained the CEL I open reading
25 frame extending from nucleotide 5765 to 6655 (SEQ ID
NO:03); and the clone p1177MP4- CEL I 6His-A9 (SEQ ID
NO:02) contained the CEL I-6His open reading frame
extending from nucleotide 5765-6679 (SEQ ID NO:04).

30

1/38

SEQUENCE LISTING

<110> Large Scale Biology Corporation
 <120> MISMATCH ENDONUCLEASES AND METHODS OF USE THEREOF
 <130> O8900840CA
 <140>
 <141> 2003-01-31
 <150> 60/353,722
 <151> 2002-02-01
 <150> 10/098,155
 <151> 2002-03-14
 <150> 10/211,079
 <151> 2002-08-01
 <160> 35
 <170> PatentIn version 3.1
 <210> 1
 <211> 10600
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> TMV infectious clone containing CEL I gene

<400> 1
 gtattttttac aacaattacc aacaacaaca aacaacaaac aacattacaa ttactatttta 60
 caattacaat ggcatacaca cagacagcta ccacatcagc tttgctggac actgtccgag 120
 gaaacaactc cttgggtcaat gatctagcaa agcgtcgtct ttacgacaca gcggttgaag 180
 agtttaacgc tcgtgaccgc aggcccaagg tgaacttttc aaaagtaata agcgaggagc 240
 agacgcttat tgctaccgag gcgtatccag aattccaaat tacattttat aacacgcaaa 300
 atgccgtgca ttcgcttgca ggtggattgc gatctttaga actggaatat ctgatgatgc 360
 aaattcccta cggatcattg acttatgaca taggcgggaa ttttgcacgc catctgttca 420
 agggacgagc atatgtacac tgctgcatgc ccaacctgga cgttcgagac atcatgcggc 480
 acgaaggcca gaaagacagt attgaactat acctttctag gctagagaga ggggggaaaa 540
 cagtcccca cttccaaaag gaagcatttg acagatacgc agaaattcct gaagacgctg 600
 tctgtcacia tactttccag acaatgacgc atcagccgat gcagcaatca ggcagagtgt 660
 atgccattgc gctacacagc atatatgaca taccagccga tgagttcggg gcggcactct 720
 tgaggaaaaa tgtccatacg tgctatgccg ctttccactt ctctgagaac ctgcttcttg 780
 aagattcata cgtcaatttg gacgaaatca acgcgtgttt ttgcgcgat ggagacaagt 840
 tgaccttttc ttttgcacga gagagtactc ttaattattg tcatagttat tctaatttc 900

ttaagtatgt gtgcaaaact tacttcccgg cctctaatag agaggtttac atgaaggagt	960
ttttagtcac cagagttaat acctgggtttt gtaagttttc tagaatagat acttttcttt	1020
tgtacaaagg tgtggcccat aaaagtgtag atagttagca gttttatact gcaatggaag	1080
acgcatggca ttacaaaaag actcttgcaa tgtgcaacag cgagagaatc ctccttgagg	1140
attcatcatc agtcaattac tggtttccca aaatgagga tatgggtcatc gtaccattat	1200
tcgacatttc tttggagact agtaagagga cgcgcaagga agtcttagtg tccaaggatt	1260
tcgtgtttac agtgcttaac cacattcgaa cataccaggc gaaagctctt acatacgcaa	1320
atgttttgtc ctttgtcgaa tcgattcgat cgagggtaat cattaacggt gtgacagcga	1380
gggccgaatg ggatgtggac aaatctttgt tacaatcctt gtccatgacg ttttacctgc	1440
atactaagct tgccgttcta aaggatgact tactgattag caagttagt ctcgggtcga	1500
aaacggtgtg ccagcatgtg tgggatgaga ttctgctggc gtttggaac gcatttcct	1560
ccgtgaaaga gaggtcttg aacaggaaac ttatcagagt ggcaggcgac gcattagaga	1620
tcagggtgcc tgatctatat gtgaccttc acgacagatt agtgactgag tacaaggcct	1680
ctgtggacat gcctgctgtt gacattagga agaagatgga agaaacggaa gtgatgtaca	1740
atgcactttc agagttatcg gtgttaaggg agtctgacaa attcgatgtt gatgtttttt	1800
cccagatgtg ccaatctttg gaagttgacc caatgacggc agcgaagggt atagtcgcgg	1860
tcattgagcaa tgagagcggc ctgactctca catttgaacg acctactgag gcgaatgttg	1920
cgctagcttt acaggatcaa gagaaggctt cagaagggtc tttggtagtt acctcaagag	1980
aagttgaaga accgtccatg aagggttcga tggccagagg agagttacaa ttagctggtc	2040
ttgctggaga tcatccggag tcgtcctatt ctaagaacga ggagatagag tcttttagagc	2100
agtttcatat ggcaacggca gattcgttaa ttcgtaagca gatgagctcg attgtgtaca	2160
cgggtccgat taaagttcag caaatgaaaa actttatcga tagcctggta gcatcactat	2220
ctgctgcggc gtcgaatctc gtcaagatcc tcaaagatac agctgctatt gaccttgaaa	2280
cccgtcaaaa gtttgagtc ttggatgttg catctaggaa gtgggttaac aaaccaacgg	2340
ccaagagtca tgcattgggt gttgttgaaa ccacgcgag gaagtatcat gtggcgcttt	2400
tggaatatga tgagcagggt gtggtgacat gcgatgattg gagaagagta gctgtcagct	2460
ctgagtctgt tgtttattcc gacatggcga aactcagaac tctgcgcaga ctgcttcgaa	2520
acggagaacc gcatgtcagt agcgcaaagg ttgttcttgt ggacggagtt ccgggctgtg	2580
ggaaaaccaa agaaattctt tccagggtta attttgatga agatctaatt ttagtacctg	2640
ggaagcaagc cgcggaaatg atcagaagac gtgcgaattc ctcagggatt attgtggcca	2700
cgaaggacaa cgttaaaacc gttgattctt tcatgatgaa ttttgggaaa agcacacgct	2760

gtcagttcaa gaggttattc attgatgaag ggttgatggt gcatactggg tgtgttaatt	2820
ttcttgtggc gatgtcattg tgcgaaattg catatgttta cggagacaca cagcagattc	2880
catacatcaa tagagtttca ggattcccgt accccgcccc ttttgccaaa ttggaagttg	2940
acgaggtgga gacacgcaga actactctcc gttgtccagc cgatgtcaca cattatctga	3000
acaggagata tgagggtttt gtcattgagca cttcttcggg taaaaagtct gtttcgcagg	3060
agatggtcgg cggagccgcc gtgatcaatc cgatctcaaa acccttgcat ggcaagatcc	3120
tgacttttac ccaatcggat aaagaagctc tgctttcaag agggatttca gatgttcaca	3180
ctgtgcatga agtgcaaggc gagacatact ctgatgtttc actagttagg ttaacccta	3240
caccagtctc catcattgca ggagacagcc cacatgtttt ggtcgcattg tcaaggcaca	3300
cctgttcgct caagtactac actgttggtt tggatccttt agttagtatc attagagatc	3360
tagagaaact tagctcgtag ttgttagata tgtataaggc cgatgcagga acacaatagc	3420
aattacagat tgactcgggtg ttcaaagggt ccaatctttt tggtgcagcg ccaaagactg	3480
gtgatatttc tgatatgcag ttttactatg ataagtgtct cccaggcaac agcaccatga	3540
tgaataattt tgatgctggt accatgaggt tgactgacat ttcattgaat gtcaaagatt	3600
gcatattgga tatgtctaag tctgttgctg cgcctaagga tcaaatcaaa ccactaatac	3660
ctatggtacg aacggcggca gaaatgccac gccagactgg actattggaa aatttagtgg	3720
cgatgattaa aaggaacttt aacgcacccg agttgtctgg catcattgat attgaaaata	3780
ctgcatcttt agttgtagat aagttttttg atagttattt gcttaaagaa aaaagaaaac	3840
caaataaaaa tgtttctttg ttcagtagag agtctctcaa tagatgggtta gaaaagcagg	3900
aacaggtaac aataggccag ctgcagatt ttgatattgt agatttgcca gcagttgatc	3960
agtacagaca catgattaa gcacaacca agcaaaaatt ggacacttca atccaaacgg	4020
agtaccggc tttgcagacg attgtgtacc attcaaaaaa gatcaatgca atatttgcc	4080
cgttgtttag tgagcttact aggcaattac tggacagtgt tgattcgagc agatttttgt	4140
ttttcacaag aaagacacca gcgcagattg aggatttctt cggagatctc gacagtcatg	4200
tgccgatgga tgtcttgag ctggatatat caaaatacga caaatctcag aatgaattcc	4260
actgtgcagt agaatacgag atctggcgaa gattgggttt tgaagacttc ttgggagaag	4320
tttggaaca agggcataga aagaccacc tcaaggatta taccgcagg ataaaaactt	4380
gcatctggta tcaaagaaag agcggggacg tcacgacgtt cattggaaac actgtgatca	4440
ttgctgcatg tttggcctcg atgcttccga tggagaaaat aatcaaagga gccttttgcg	4500
gtgacgatag tctgctgtac tttccaaagg gttgtgagtt tccggatgtg caacactccg	4560

cgaatcttat	gtggaatttt	gaagcaaaac	tgtttaaaaa	acagtatgga	tacttttgcg	4620
gaagatatgt	aatacatcac	gacagaggat	gcatttgtga	ttacgatccc	ctaaagttga	4680
tctcgaaact	tggtgctaaa	cacatcaagg	attgggaaca	cttggaggag	ttcagaaggt	4740
ctcttttgtga	tgttgctggt	tcgttgaaca	attgtgcgta	ttacacacag	ttggacgacg	4800
ctgtatggga	ggttcataag	accgcccctc	caggttcggt	tgtttataaa	agtctgggtga	4860
agtatttgct	tgataaagtt	cttttttagaa	gtttgtttat	agatggctct	agttgttaaa	4920
ggaaaagtga	atatcaatga	gtttatcgac	ctgacaaaaa	tggagaagat	cttaccgtcg	4980
atgtttaccc	ctgtaaagag	tgttatgtgt	tccaaagttg	ataaaataat	ggttcatgag	5040
aatgagtcac	tgtcagaggt	gaaccttctt	aaaggagtta	agcttattga	tagtggatac	5100
gtctgttttag	ccggtttggt	cgtcacgggc	gagtggaaact	tgcctgacaa	ttgcagagga	5160
ggtgtgagcg	tgtgtctggt	ggacaaaagg	atggaaagag	ccgacgaggc	cactctcgga	5220
tcttactaca	cagcagctgc	aaagaaaaga	tttcagttca	aggctggtcc	caattatgct	5280
ataaccaccc	aggacgcgat	gaaaaacgtc	tggcaagttt	tagttaatat	tagaaatgtg	5340
aagatgtcag	cgggtttctg	tccgctttct	ctggagtttg	tgtcgggtgtg	tattgtttat	5400
agaaataata	taaaattagg	tttgagagag	aagattacaa	acgtgagaga	cggagggccc	5460
atggaactta	cagaagaagt	cgttgatgag	ttcatggaag	atgtccctat	gtcgatcagg	5520
cttgcaaagt	ttcgatctcg	aaccggaaaa	aagagtgatg	tccgcaaagg	gaaaaatagt	5580
agtaatgatc	ggtcagtgcc	gaacaagaac	tatagaaatg	ttaaggatgt	tggaggaatg	5640
agttttaaaa	agaataatgt	aatcgatgat	gattcggagg	ctactgtcgc	cgaatcggat	5700
tcgttttaaa	tagatcttac	agtatcacta	ctccatctca	gttcgtgttc	ttgtcattaa	5760
ttaaatagacg	cgattatatt	ctgtgttctt	tcttttggtg	gctctttag	ttgaaccggg	5820
tgttagagcc	tggagcaaag	aaggccatgt	catgacatgt	caaattgcgc	aggatctggt	5880
ggagccagaa	gcagcacatg	ctgtaaagat	gctgttaccg	gactatgcta	atggcaactt	5940
atcgtcgtcg	tgtgtgtggc	ctgatcaaat	tcgacactgg	tacaagtaca	ggtggactag	6000
ctctctccat	ttcatcgata	cacctgatca	agcctgttca	tttgattacc	agagagactg	6060
tcatgatcca	catggaggga	aggacatgtg	tgttgctgga	gccattcaaa	atttcacatc	6120
tcagcttgga	catttccgcc	atggaacatc	tgatcgtcga	tataatatga	cagaggcttt	6180
gttattttta	tcccacttca	tgggagatat	tcatcagcct	atgcatgttg	gatttacaag	6240
tgatatggga	ggaaacagta	tagatttgcg	ctggtttcgc	cacaaatcca	acctgcacca	6300
tgtttgggat	agagagatta	ttcttacagc	tgcagcagat	taccatggta	aggatatgca	6360
ctctctccta	caagacatac	agaggaactt	tacagagggt	agttggttgc	aagatgttga	6420

atcctggaag gaatgtgatg atatctctac ttgcgccaat aagtatgcta aggagagtat 6480
aaaactagcc tgtaactggg gttacaaaga tggtgaatct ggcgaaactc tgtcagataa 6540
atacttcaac acaagaatgc caattgtcat gaaacggata gctcaggggtg gaatccgttt 6600
atccatgatt ttgaaccgag ttcttggaag ctccgcagat cattctttgg catgacctag 6660
gccagtagtt tggtttaaac ccaactgcga ggggtagtca agatgcataa taaataacgg 6720
attgtgtccg taatcacacg tgggtgcgtac gataacgcat agtggttttc cctccactta 6780
aatcgaaggg ttgtgtcttg gatcgcgcggt gtcaaatgta tatggttcat atacatccgc 6840
aggcacgtaa taaagcgagg ggttcggggtc gaggtcgggt gtgaaactcg aaaagggtcc 6900
ggaaaacaaa aaagagatgg taggtaatag tgtaataat aagaaaataa ataatagtgg 6960
taagaaaggt ttgaaagttg aggaaattga ggataatgta agtgatgacg agtctatcgc 7020
gtcatcgagt acgtttttaat caatatgcct tatacaatca actctccgag ccaatttggt 7080
tacttaagtt ccgcttatgc agatcctgtg cagctgatca atctgtgtac aaatgcattg 7140
ggtaaccagt ttcaaacgca acaagctagg acaacagtc aacagcaatt tgcggatgcc 7200
tggaacactg tgcctagtat gacagtgaga tttcctgcat cggatttcta tgtgtataga 7260
tataattcga cgcttgatcc gttgatcacg gcgttattaa atagcttcga tactagaaat 7320
agaataatag aggttgataa tcaaccgcga ccgaatacta ctgaaatcgt taacgcgact 7380
cagagggtag acgatgcgac tgtagctata agggcttcaa tcaataattt ggctaataa 7440
ctgggttcgtg gaactggcat gttcaatcaa gcaagctttg agactgctag tggacttgtc 7500
tggaccacaa ctccggctac ttagctattg ttgtgagatt tcctaaaata aagtcactga 7560
agacttaaaa ttcaggggtg ctgataccaa aatcagcagt ggttggtcgt ccacttaaat 7620
ataacgattg tcatatctgg atccaacagt taaacatgt gatgggtgat actgtggtat 7680
ggcgtaaaac aacggaaaag tcgctgaaga cttaaaattc aggggtggctg ataccaaaat 7740
cagcagtggg tggtcgcca cttaaaaata acgattgtca tatctggatc caacagttaa 7800
accatgtgat ggtgtatact gtggtatggc gtaaaacaac ggagaggttc gaatcctccc 7860
ctaaccgcgg gtagcggccc aggtaccgg atgtgttttc cgggctgatg agtccgtgag 7920
gacgaaaccc ggcatgcaag cttggcgtaa tcatgggtcat agctgtttcc tgtgtgaaat 7980
tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagt taaagcctgg 8040
ggtgcctaata gagtgagcta actcacatta attgcgttgc gtcactgcc cgctttccag 8100
tcgggaaacc tgctgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt 8160
ttgcgtattg ggcgtcttc cgcttcctcg ctactgact cgctgcgctc ggtcggttcgg 8220

ctgcggcgag	cggtatcagc	tcactcaaag	gcggtaatac	ggttatccac	agaatcaggg	8280
gataacgcag	gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	8340
gccgcgttgc	tggcgttttt	ccataggctc	cgccccctg	acgagcatca	caaaaatcga	8400
cgctcaagtc	agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	8460
ggaagctccc	tcgtgcgctc	tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	8520
tttctccctt	cggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	8580
gtgtaggtcg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	gcccgaccgc	8640
tgcgccttat	ccggtaaacta	tcgtcttgag	tccaaccgcg	taagacacga	cttatcgcca	8700
ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	8760
ttcttgaagt	ggtaggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	8820
ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	caaacaacc	8880
accgctggta	gcggtggttt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	8940
tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtggaa	cgaaaactca	9000
cgtaagggga	ttttggtcac	gagattatca	aaaaggatct	tcacctagat	ccttttaa	9060
taaaaatgaa	gttttaaatac	aatctaaagt	atatatgagt	aaacttggtc	tgacagttac	9120
caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcgttc	atccatagtt	9180
gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	tggtcccgag	9240
gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atttatcagc	aataaaccag	9300
ccagccggaa	gggcccagcg	cagaagtggc	cctgcaactt	tatccgcctc	catccagtct	9360
attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatagttt	gcgcaacggt	9420
gttgccattg	ctacaggcat	cgtgggtgtc	cgctcgtcgt	ttgggtatggc	ttcattcagc	9480
tccggttccc	aacgatcaag	gcgagttaca	tgatccccca	tggtgtgcaa	aaaagcggtt	9540
agctccttcg	gtcctccgat	cgttgtcaga	agtaagttgg	ccgcagtgtt	atcactcatg	9600
gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	cttttctgtg	9660
actggtgagt	actcaaccaa	gtcattctga	gaatagtgtg	tgccggcgacc	gagttgctct	9720
tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	agtgctcatc	9780
attggaaaac	gttcttcggg	gcgaaaactc	tcaaggatct	taccgctgtt	gagatccagt	9840
tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	caccagcggt	9900
tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	agggaataag	ggcgacacgg	9960
aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	tcagggttat	10020
tgtctcatga	gcggatacat	atgtgaatgt	athtagaaaa	ataaacaat	aggggttccg	10080

cgcacatttc cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta 10140
 acctataaaa ataggcgtat cacgaggccc ttctgtctcg cgcgtttcgg tgatgacggg 10200
 gaaaacctct gacacatgca gctcccggag acggtcacag cttgtctgta agcggatgcc 10260
 gggagcagac aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg gggctggctt 10320
 aactatgcgg catcagagca gattgtactg agagtgcacc atatgcgggtg tgaaataccg 10380
 cacagatgcg taaggagaaa ataccgcac aggcgccatt cgccattcag gctgcgcaac 10440
 tgttgggaag ggcgatcggg gcgggcctct tcgctattac gccagctggc gaaaggggga 10500
 tgtgctgcaa ggcgattaag ttgggtaacg ccagggtttt ccaggtcacg acgttgtaaa 10560
 acgacggcca gtgaattcaa gcttaatacg actcactata 10600

<210> 2

<211> 10624

<212> DNA

<213> Artificial Sequence

<220>

<223> TMV infectious clone containing CEL I gene fused to a 6HIS encoding sequence

<400> 2

gtattttttac aacaattacc aacaacaaca aacaacaaac aacattacaa ttactattta 60
 caattacaat ggcatacaca cagacagcta ccacatcagc ttgctggac actgtccgag 120
 gaaacaactc cttgggtcaat gatctagcaa agcgtcgtct ttacgacaca gcggttgaag 180
 agtttaacgc tcgtgaccgc aggcccaagg tgaacttttc aaaagtaata agcgaggagc 240
 agacgcttat tgctaccgag gcgtatccag aattccaaat tacattttat aacacgcaaa 300
 atgccgtgca ttcgcttgca ggtggattgc gatctttaga actggaatat ctgatgatgc 360
 aaattcccta cggatcattg acttatgaca taggcgggaa ttttgcacg catctgttca 420
 agggacgagc atatgtacac tgctgcatgc ccaacctgga cgttcgagac atcatgcggc 480
 acgaaggcca gaaagacagt attgaactat acctttctag gctagagaga ggggggaaaa 540
 cagtcccaa cttccaaaag gaagcatttg acagatacgc agaaattcct gaagacgctg 600
 tctgtcacia tactttccag acaatgcgac atcagccgat gcagcaatca ggcagagtgt 660
 atgccattgc gctacacagc atatatgaca taccagccga tgagttcggg gcggcactct 720
 tgaggaaaaa tgtccatacg tgctatgcg ctttccactt ctctgagaac ctgcttcttg 780
 aagattcata cgtcaatttg gacgaaatca acgcgtgttt ttccgcgcat ggagacaagt 840
 tgaccttttc ttttgcacga gagagtactc ttaattattg tcatagttat tctaataatc 900
 ttaagtatgt gtgcaaaact tacttcccgg cctctaatag agaggtttac atgaaggagt 960

ttttagtcac	cagagttaat	acctggtttt	gtaagttttc	tagaatagat	actttttcttt	1020
tgtacaaagg	tgtggcccat	aaaagtgtag	atagtgagca	gttttatact	gcaatggaag	1080
acgcatggca	ttacaaaaag	actcttgcaa	tgtgcaacag	cgagagaatc	ctccttgagg	1140
attcatcatc	agtcaattac	tggtttccca	aatgagggga	tatggtcac	gtaccattat	1200
tcgacatttc	tttggagact	agtaagagga	cgcgcaagga	agtcttagtg	tccaaggatt	1260
tcgtgtttac	agtgcttaac	cacattcgaa	cataccaggc	gaaagctctt	acatacgcaa	1320
atgttttgtc	ctttgtcgaa	tcgattcgat	cgagggtaat	cattaacggt	gtgacagcga	1380
gggccgaatg	ggatgtggac	aatctttgt	tacaatcctt	gtccatgacg	ttttacctgc	1440
atactaagct	tgccgttcta	aaggatgact	tactgattag	caagttagtg	ctcggttcga	1500
aaacggtgtg	ccagcatgtg	tgggatgaga	tttcgctggc	gtttgggaac	gcatttcctt	1560
ccgtgaaaga	gaggctcttg	aacaggaaac	ttatcagagt	ggcaggcgac	gcattagaga	1620
tcagggtgcc	tgatctatat	gtgaccttcc	acgacagatt	agtgactgag	tacaaggcct	1680
ctgtggacat	gcctgcgctt	gacattagga	agaagatgga	agaaacggaa	gtgatgtaca	1740
atgcactttc	agagttatcg	gtgttaaggg	agtctgacaa	attcgatgtt	gatgtttttt	1800
cccagatgtg	ccaatctttg	gaagttgacc	caatgacggc	agcgaagggt	atagtcgcgg	1860
tcattgagcaa	tgagagcggg	ctgactctca	catttgaacg	acctactgag	gcgaatgttg	1920
cgctagcttt	acaggatcaa	gagaaggctt	cagaagggtc	tttggtagtt	acctcaagag	1980
aagttgaaga	accgtccatg	aagggttcga	tggccagagg	agagttacaa	ttagctgggc	2040
ttgctggaga	tcattccggag	tcgtcctatt	ctaagaacga	ggagatagag	tcttttagagc	2100
agtttcatat	ggcaacggca	gattcgttaa	ttcgtaagca	gatgagctcg	attgtgtaca	2160
cgggtccgat	taaagttcag	caaatgaaaa	actttatcga	tagcctggta	gcatactat	2220
ctgctgcggg	gtcgaatctc	gtcaagatcc	tcaaagatac	agctgctatt	gaccttgaaa	2280
cccgtcaaaa	gtttggagtc	ttggatgttg	catctaggaa	gtggttaatc	aaaccaacgg	2340
ccaagagtca	tgcatggggg	gttggtgaaa	cccacgcgag	gaagtatcat	gtggcgcttt	2400
tggaatatga	tgagcagggt	gtggtgacat	gcgatgattg	gagaagagta	gctgtcagct	2460
ctgagtctgt	tgtttatctc	gacatggcga	aactcagaac	tctgcgcaga	ctgcttcgaa	2520
acggagaacc	gcatgtcagt	agcgcaaagg	ttgttcttgt	ggacggagtt	ccgggctgtg	2580
ggaaaaccaa	agaaattctt	tccagggtta	atthttgatga	agatctaatt	ttagtacctg	2640
ggaagcaagc	cgcggaatg	atcagaagac	gtgcgaattc	ctcagggtatt	attgtggcca	2700
cgaaggacaa	cgtaaaaacc	gttgattctt	tcattgatgaa	ttttgggaaa	agcacacgct	2760

gtcagttcaa gaggttattc attgatgaag ggttgatggt gcatactggt tgtgttaatt	2820
ttcttgtggc gatgtcattg tgcgaaattg catatgttta cggagacaca cagcagattc	2880
catacatcaa tagagtttca ggattcccgt accccgcccc ttttgccaaa ttggaagttg	2940
acgaggtgga gacacgcaga actactctcc gttgtccagc cgatgtcaca cattatctga	3000
acaggagata tgagggtctt gtcattgagca cttcttcggt taaaaagtct gtttcgcagg	3060
agatggtcgg cggagccgcc gtgatcaatc cgatctcaaa acccttgcat ggcaagatcc	3120
tgacttttac ccaatcggat aaagaagctc tgctttcaag agggatttca gatgttcaca	3180
ctgtgcatga agtgcaaggc gagacatact ctgatgtttc actagttagg ttaacccta	3240
caccagtctc catcattgca ggagacagcc cacatgtttt ggtcgcattg tcaaggcaca	3300
cctgttcgct caagtactac actgttggtta tggatccttt agttagtatc attagagatc	3360
tagagaaact tagctcgtag ttgtagata tgtataaggt cgatgcagga acacaatagc	3420
aattacagat tgactcgggtg ttcaaagggt ccaatctttt tggtgcagcg ccaaagactg	3480
gtgatatttc tgatatgcag ttttactatg ataagtgtct cccaggcaac agcaccatga	3540
tgaataattt tgatgctggt accatgaggt tgactgacat ttcattgaat gtcaaagatt	3600
gcatattgga tatgtctaag tctgttgctg cgcctaagga tcaaatacaa ccactaatac	3660
ctatggtacg aacggcggca gaaatgccac gccagactgg actattggaa aatttagtgg	3720
cgatgattaa aaggaacttt aacgcacccg agttgtctgg catcattgat attgaaaata	3780
ctgcatcttt agttgtagat aagttttttg atagttattt gcttaaagaa aaaagaaaac	3840
caaataaaaa tgtttctttg ttcagtagag agtctctcaa tagatgggtta gaaaagcagg	3900
aacaggtaac aataggccag ctgcagatt ttgattttgt agatttgcca gcagttgatc	3960
agtacagaca catgattaa gcacaacca agcaaaaatt ggacacttca atccaaacgg	4020
agtaccggc tttgcagacg attgtgtacc attcaaaaaa gatcaatgca atatttggcc	4080
cgttgtttag tgagcttact aggcaattac tggacagtgt tgattcgagc agatttttgt	4140
ttttcacaag aaagacacca gcgcagattg aggatttctt cggagatctc gacagtcattg	4200
tgccgatgga tgtcttgag ctggatatat caaaatacga caaatctcag aatgaattcc	4260
actgtgcagt agaatacgag atctggcgaa gattgggttt tgaagacttc ttgggagaag	4320
tttggaaca agggcataga aagaccacc tcaaggatta taccgcagg ataaaaactt	4380
gcatctggta tcaaagaaag agcggggacg tcacgacgtt cattggaaac actgtgatca	4440
ttgctgcatg tttggcctcg atgcttccga tggagaaaat aatcaaagga gccttttgcg	4500
gtgacgatag tctgctgtac tttccaaagg gttgtgagtt tccggatgtg caacactccg	4560
cgaatcttat gtggaatttt gaagcaaaac tgtttaaaaa acagtatgga tacttttgcg	4620

gaagatatgt aatacatcac gacagaggat gcatttgtga ttacgatccc ctaaagttga 4680
tctcgaaact tggcgctaaa cacatcaagg attgggaaca cttggaggag ttcagaaggt 4740
ctcttttgtga tgttgctggt tcgttgaaca attgtgctga ttacacacag ttggacgacg 4800
ctgtatggga gggtcataag accgcccctc cagggttcgtt tgtttataaa agtctggtga 4860
agtatttgtc tgataaagtt ctttttagaa gtttgtttat agatggctct agttgttaaa 4920
ggaaaagtga atatcaatga gtttatcgac ctgacaaaaa tggagaagat cttaccgtcg 4980
atgtttaccc ctgtaaagag tgttatgtgt tccaaagttg ataaaataat gggtcatgag 5040
aatgagtcac tgtcagaggt gaaccttctt aaaggagtta agcttattga tagtggatac 5100
gtctgttttag ccggtttggt cgtcacgggc gagtggaaact tgcctgacaa ttgcagagga 5160
ggtgtgagcg tgtgtctggt ggacaaaagg atggaaagag ccgacgaggc cactctcgga 5220
tcttactaca cagcagctgc aaagaaaaga tttcagttca aggtcgttcc caattatgct 5280
ataaccaccc aggacgcgat gaaaaacgtc tggcaagttt tagttaatat tagaaatgtg 5340
aagatgtcag cgggtttctg tccgctttct ctggagtttg tgtcgggtgtg tattgtttat 5400
agaaataata taaaattagg tttgagagag aagattacaa acgtgagaga cggagggccc 5460
atggaactta cagaagaagt cgttgatgag ttcattggaag atgtccctat gtcgatcagg 5520
cttgcaaagt ttcgatctcg aaccggaaaa aagagtgatg tccgcaaagg gaaaaatagt 5580
agtaatgatc ggtcagtgcc gaacaagaac tatagaaatg ttaaggattt tggaggaatg 5640
agtttttaaaa agaataattt aatcgatgat gattcggagg ctactgtcgc cgaatcggat 5700
tcgttttaaaa tagatcttac agtatcacta ctccatctca gttcgtgttc ttgtcattaa 5760
ttaaattgacg cgattatatt ctgtgttctt tcttttgttg gctcttgtag ttgaaccggg 5820
tgtagagacc tggagcaaag aaggccatgt catgacatgt caaattgcgc aggatctggt 5880
ggagccagaa gcagcacatg ctgtaaagat gctgttaccg gactatgcta atggcaactt 5940
atcgtcgtg tgtgtgtggc ctgatcaaat tcgacactgg tacaagtaca ggtggactag 6000
ctctctccat ttcattcgata cacctgatca agcctgttca tttgattacc agagagactg 6060
tcatgatcca catggaggga aggacatgtg tgttgctgga gccattcaaa atttcacatc 6120
tcagcttggga catttccgcc atggaacatc tgatcgtcga tataatatga cagaggcttt 6180
gttattttta tcccacttca tgggagatat tcatcagcct atgcatgttg gatttacaag 6240
tgatatggga ggaaacagta tagatttgcg ctggtttcgc cacaaatcca acctgcacca 6300
tgtttgggat agagagatta ttcttacagc tgcagcagat taccatggta aggatatgca 6360
ctctctccta caagacatac agaggaactt tacagagggt agttgggtgc aagatgttga 6420

atcctggaag gaatgtgatg atatctctac ttgcgccaat aagtatgcta aggagagtat	6480
aaaactagcc tgtaactggg gttacaaaga tgttgaatct ggcgaaactc tgtcagataa	6540
atacttcaac acaagaatgc caattgtcat gaaacggata gctcaggggtg gaatccgttt	6600
atccatgatt ttgaaccgag ttcttggaag ctccgcagat cattctttgg caggagggtca	6660
ccatcaccat caccattgac ctaggccagt agtttggttt aaacccaact gcgaggggta	6720
gtcaagatgc ataataaata acggattgtg tccgtaatca cacgtggtgc gtacgataac	6780
gcatagtgtt tttccctcca cttaaactga agggttgtgt cttggatcgc gcgggtcaaa	6840
tgtatatggg tcatatacat ccgcaggcac gtaataaagc gaggggttcg ggtcgaggtc	6900
ggctgtgaaa ctcgaaaagg ttccggaaaa caaaaaagag atggtaggta atagtgttaa	6960
taataagaaa ataaataata gtggtaagaa aggtttgaaa gttgaggaaa ttgaggataa	7020
tgtaagtgat gacgagtcta tcgcgtcatc gagtacgttt taatcaatat gccttataca	7080
atcaactctc cgagccaatt tgtttactta agttccgctt atgcagatcc tgtgcagctg	7140
atcaatctgt gtacaaatgc attgggtaac cagtttcaaa cgcaacaagc taggacaaca	7200
gtccaacagc aatttgcgga tgcctggaaa cctgtgccta gtatgacagt gagatttcct	7260
gcatcggatt tctatgtgta tagatataat tcgacgcttg atccgttgat cacggcgtta	7320
ttaaatagct tcgatactag aaatagaata atagagggtg ataatacaacc cgcaccgaat	7380
actactgaaa tcgttaacgc gactcagagg gtagacgatg cgactgtagc tataagggtc	7440
tcaatcaata atttggctaa tgaactgggt cgtggaactg gcatgttcaa tcaagcaagc	7500
tttgagactg ctagtggact tgtctggacc acaactccgg ctacttagct attgttgtga	7560
gatttcctaa aataaagtca ctgaagactt aaaattcagg gtggctgata ccaaaatcag	7620
cagtggttgt tcgtccactt aaatataacg attgtcatat ctggatccaa cagttaaacc	7680
atgtgatggg gtatactgtg gtatggcgta aaacaacgga aaagtcgctg aagacttaaa	7740
attcaggggtg gctgatacca aaatcagcag tggttgttcg tccacttaaa aataacgatt	7800
gtcatatctg gatccaacag ttaaaccatg tgatgggtgta tactgtggta tggcgtaaaa	7860
caacggagag gttcgaatcc tcccctaacc gcgggtagcg gccaggtac ccggatgtgt	7920
tttccgggct gatgagtccg tgaggacgaa acccggtatg caagcttggc gtaatcatgg	7980
tcatagctgt ttctgtgtg aaattgttat ccgctcacia ttccacacia catacgagcc	8040
ggaagcataa agtgtaaagc ctgggggtgcc taatgagtga gctaactcac attaattgcg	8100
ttgcgctcac tgcccgttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc	8160
ggccaacgcg cggggagagg cggtttgctg attgggcgct cttccgcttc ctgctcact	8220
gactcgctgc gctcggctcg tcggctgcgg cgagcggtat cagctcactc aaaggcggta	8280

atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag 8340
caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc 8400
cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 8460
taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg 8520
ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc 8580
tcacgctgta ggtatctcag ttcgggtgtag gtcgttcgct ccaagctggg ctgtgtgcac 8640
gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac 8700
ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg 8760
aggatatgtag gcggtgctac agagtctctg aagtgggtggc ctaactacgg ctacactaga 8820
aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 8880
agctcttgat ccggcaaaca aaccaccgct ggtagcgggtg gtttttttgt ttgcaagcag 8940
cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgggggtct 9000
gacgctcagt ggaacgaaaa ctcacgttaa gggatttttg tcatgagatt atcaaaaagg 9060
atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat 9120
gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 9180
tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg 9240
gagggcttac catctggccc cagtgctgca atgataccgc gagaccacg ctcaccggct 9300
ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca 9360
actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg 9420
ccagttaata gtttgcgcaa cgttggtgcc attgctacag gcatcgtggg gtcacgctcg 9480
tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc 9540
cccatgttgt gcaaaaaagc ggtagctcc ttcggtcctc cgatcgttgt cagaagtaag 9600
ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg 9660
ccatccgtaa gatgcttttc tgtgactggg gagtactcaa ccaagtcatt ctgagaatag 9720
tgtatgcggc gaccgagttg ctcttgcccc gcgtcaatac gggataatac cgcgccacat 9780
agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg 9840
atcttaccgc tgttgagatc cagttcgatg taaccactc gtgcacccaa ctgatcttca 9900
gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca 9960
aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat 10020
tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag 10080

```

aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgtctaa 10140
gaaaccatta ttatcatgac attaacctat aaaaataggc gtatcacgag gccctttcgt 10200
ctcgcgcggt tccgtgatga cggtgaaaac ctctgacaca tgcagctccc ggagacggtc 10260
acagcttgtc tgtaagcgga tgccgggagc agacaagccc gtcagggcgc gtcagcgggt 10320
gttggcgggt gtcggggctg gcttaactat gcggcatcag agcagattgt actgagagtg 10380
caccatatgc ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg catcaggcgc 10440
cattegccat tcaggctgcg caactgttgg gaaggcgat cggtgcgggc ctcttcgcta 10500
ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat taagttgggt aacgccaggg 10560
ttttcccagt cacgacgttg taaaacgacg gccagtgaat tcaagcttaa tacgactcac 10620
tata 10624

```

<210> 3
 <211> 891
 <212> DNA
 <213> *Apium graveolens*

```

<400> 3
atgacgcgat tataattctgt gttctttctt ttgttggtc ttgtagttga accgggtgtt 60
agagcctgga gcaaagaagg ccatgtcatg acatgtcaaa ttgcgcagga tctgttggag 120
ccagaagcag cacatgctgt aaagatgctg ttaccggact atgctaattg caacttatcg 180
tcgctgtgtg tgtggcctga tcaaattcga cactggtaca agtacagggtg gactagctct 240
ctccatttca tcgatacacc tgatcaagcc tgttcatttg attaccagag agactgtcat 300
gatccacatg gagggaagga catgtgtgtt gctggagcca ttcaaaattt cacatctcag 360
cttggacatt tccgccatgg aacatctgat cgtcgatata atatgacaga ggctttgtta 420
tttttatccc acttcatggg agatattcat cagcctatgc atgttggatt tacaagtgat 480
atgggaggaa acagtataga tttgcgctgg ttctgccaca aatccaacct gcaccatgtt 540
tgggatagag agattattct tacagctgca gcagattacc atggtaagga tatgcactct 600
ctcctacaag acatacagag gaactttaca gagggtagtt ggttgcaaga tgttgaatcc 660
tggaaggaat gtgatgatat ctctacttgc gccataagt atgctaagga gagtataaaa 720
ctagcctgta actgggggta caaagatgtt gaatctggcg aaactctgtc agataaatac 780
ttcaacacaa gaatgccaat tgtcatgaaa cggatagctc aggggtggaat ccgtttatcc 840
atgatthtga accgagttct tggaagctcc gcagatcatt ctttggcatg a 891

```

<210> 4
 <211> 915
 <212> DNA

<213> *Apium graveolens*

<400> 4

```

atgacgcgat tatattctgt gttctttctt ttgttggtc ttgtagtga accgggtgtt      60
agagcctgga gcaaagaagg ccatgtcatg acatgtcaaa ttgcgcagga tctgttggag    120
ccagaagcag cacatgctgt aaagatgctg ttaccggact atgctaattg caacttatcg    180
tcgctgtgtg tgtggcctga tcaaattcga cactggtaca agtacagggtg gactagctct    240
ctccatttca tcgatacacc tgatcaagcc tgttcatttg attaccagag agactgtcat    300
gatccacatg gagggaagga catgtgtgtt gctggagcca ttcaaaattt cacatctcag    360
cttggacatt tccgccatgg aacatctgat cgtcgatata atatgacaga ggctttgtta    420
tttttatccc acttcatggg agatattcat cagcctatgc atgttggatt tacaagtgat    480
atgggaggaa acagtataga tttgcgctgg tttcgccaca aatccaacct gcaccatgtt    540
tgggatagag agattattct tacagctgca gcagattacc atggtaagga tatgcactct    600
ctcctacaag acatacagag gaactttaca gagggtagtt ggttgcaaga tgttgaatcc    660
tggaaggaat gtgatgatat ctctacttgc gccataagt atgctaagga gagtataaaa    720
ctagcctgta actgggggta caaagatgtt gaatctggcg aaactctgtc agataaatac    780
ttcaacacaa gaatgccaat tgtcatgaaa cggatagctc aggggtggaat ccgtttatcc    840
atgattttga accgagttct tggaagctcc gcagatcatt ctttggcagg aggtcaccat    900
caccatcacc attga                                              915

```

<210> 5

<211> 717

<212> DNA

<213> Artificial Sequence derived from *Aequorea victoria* GFP

<220>

<223> This construct was derived by shuffling in accordance with the methodology of the present invention.

<400> 5

```

atgagtaaag gagaagaact ttctactgga gttgtcccaa ttcttgttga attagatggt      60
gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggatgatgc aacatacggg    120
aaacttacc c ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt    180
gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg    240
catgactttt tcaagagtgc catgcccga ggttatgtac aggaacgcac tatatttttc    300
aaggatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaagggtga tacccttggt    360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa    420
ttggaataca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga    480

```

atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcgttca actagcagac 540
 cattatcaac aaaataactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
 ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt 660
 cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 6

<211> 717

<212> DNA

<213> Artificial Sequence was derived from Aequorea Victoria GFP

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 6

atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggg 60
 gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggtgatgc tacatacgga 120
 aagcttacct ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
 gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg 240
 catgactttt tcaagagtgc catgcccga ggttatgtac aggaacgcac tatacttttc 300
 aaagatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaaggtga tacccttggt 360
 aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa 420
 ctcgagtaca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga 480
 atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcgttca actagcagac 540
 cattatcaac aaaataactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
 ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt 660
 cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 7

<211> 717

<212> DNA

<213> Artificial Sequence was derived from Aequorea Victoria GFP

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 7

atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggg 60
 gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggtgatgc tacatacgga 120
 aagcttacct ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180

```

gtcactactt tctcttatgg tgttcaatgc ttttcccggt atccggatca tatgaaacgg      240
catgactttt tcaagagtgc catgcccgaag ggttatgtac aggaacgcac tatatctttc      300
aaagatgacg ggaactacaa gacgcgtgct gaagtcaagt ttgaagggtga tacccttggt      360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa      420
ttggaataca actataactc acacaatgta tacatcacgg cagacaaaca aaagaatgga      480
atcaaagcta acttcaaaat tcgccacaac attgaagatg gatccgttca actagcagac      540
cattatcaac aaaataactcc aattggcgat ggccctgtcc ttttaccaga caaccattac      600
ctgtcgacac aatctgccct ttcgaaagat cccaacgaaa agcgtgacca catggtcctt      660
cttgagtttg taactgctgc tgggattaca catggcatgg atgaactata caaataa          717

```

<210> 8

<211> 717

<212> DNA

<213> Artificial Sequence was derived from Aequorea Victoria GFP

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention

<400> 8

```

atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggt      60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggatgatgc aacatacggg      120
aaacttacc cttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt      180
gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg      240
catgactttt tcaagagtgc catgcccgaag ggttatgtac aggaagaac tatatctttc      300
aaggatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaagggtga tacccttggt      360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa      420
ctcgagtaca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga      480
atcaaagtta acttcaaaat tcgccacaac attgaagatg gatccgttca actagcagac      540
cattatcaac aaaataactcc aattggcgat ggccctgtcc ttttaccaga caaccattac      600
ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt      660
cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa          717

```

<210> 9

<211> 807

<212> DNA

<213> Tobacco mosaic virus

<400> 9

```

atggctctag ttgttaaagg aaaagtgaat atcaatgagt ttatcgacct gacaaaaatg      60

```

gagaagatct taccgtcgat gtttaccct gtaaagagtg ttatgtgttc caaagttgat 120
 aaaataatgg ttcattgagaa tgagtcattg tcaggggtga accttcttaa aggagttaag 180
 cttattgata gtggatacgt ctgttttagcc ggtttggtcg tcacgggcga gtggaacttg 240
 cctgacaatt gcagaggagg tgtgagcgtg tgtctggtgg acaaaaggat ggaaagagcc 300
 gacgaggcca ctctcggatc ttactacaca gcagctgcaa agaaaagatt tcagttcaag 360
 gtcgttccca attatgctat aaccacccag gacgcgatga aaaacgtctg gcaagtttta 420
 gttaatatta gaaatgtgaa gatgtcagcg ggtttctgtc cgctttctct ggagtttgtg 480
 tcggtgtgta ttgtttatag aaataatata aaattagggt tgagagagaa gattacaaac 540
 gtgagagacg gagggcccat ggaacttaca gaagaagtcg ttgatgagtt catggaagat 600
 gtccctatgt cgatcaggct tgcaaagttt cgatctcgaa ccggaaaaaa gagtgatgtc 660
 cgcaaaggga aaaatagtag tagtgatcgg tcagtgccga acaagaacta tagaaatggt 720
 aaggattttg gaggaatgag ttttaaaaag aataatttaa tcgatgatga ttcggaggct 780
 actgtcgccg aatcggattc gtttttaa 807

<210> 10
 <211> 793
 <212> DNA
 <213> Tomato mosaic virus

<400> 10
 atggctctag ttgttaaagg taaggtaaatt attaagagt ttatcgatct gtcaaagtct 60
 gagaaacttc tcccgatgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
 aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
 cttatagaag gtgggtatgt ttgcttagtt ggtcttggtg tgtccggtga gtggaattta 240
 ccagataatt gccgtggtgg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg 300
 gacgaagcca cactgggggtc atattacact gctgctgcta aaaagcgggt tcagttttaa 360
 gtgggtcccaa attacggtat tactacaaag gatgcagaaa agaacatatg gcaggctcta 420
 gtaaataatta aaaatgtaaa aatgagtgcg ggctactgcc ctttgtcatt agaatttgtg 480
 tctgtgtgta ttgtttataa aaataatata aaattgggtt tgaggagagaa agtaacgagt 540
 gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
 gttccaatgt cggtttagact cgcaaagttt cgaaccaaatt cctcaaaaag aggtccgaaa 660
 aataataata atttaggtta ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat 720
 gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
 tctgattcgt att 793

<210> 11
 <211> 795
 <212> DNA
 <213> Artificial Sequence derived from tobacco mosaic virus & tomato mosaic virus

 <220>
 <223> This sequence was derived by shuffling in accordance with the methodology of the present invention

 <400> 11
 atggctctag ttgttaaagg taaggtaaatt attaatagagt ctatcgatct gtcaaagtct 60
 gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
 aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
 cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtga gtggaattta 240
 ccagataatt gccgtggtgg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg 300
 gacgaagcca cactgggggc atattacact gctgctgcta aaaagcgggt tcagttcaag 360
 gtcgttccca attatgctat aaccacccag gatgcagaaa agaacatatg gcaggtctta 420
 gtaaatatta aaaatgtaaa aatgagtgcg ggctactacc ctttgtcatt agaatttgtg 480
 tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt 540
 gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
 gttccaatgt cgatcaggct tgcaaagttt cgaaccaaatt cctcaaaaag aggtccgaaa 660
 aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaag aagttttgat 720
 gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
 tctgattcgt attaa 795

<210> 12
 <211> 795
 <212> DNA
 <213> Artificial Sequence derived from tobacco mosaic virus & tomato mosaic virus

 <220>
 <223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

 <400> 12
 atggctctag ttgttaaagg taaggtaaatt attaatagagt ttatcgatct gtcaaagtct 60
 gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
 aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
 cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtgt gtggaattta 240

ccagataatt	gccgtggtgg	tgtgagtgtc	tgcatggttg	acaagagaat	ggaaagagcg	300
gacgaggcca	cactcggatc	ttactacact	gctgctgcta	aaaagcggtt	tcagttcaag	360
gtcgttccca	attatgctat	aaccaccag	gatgcagaaa	agaacatatg	gcagggtctta	420
gtaaatatta	aaaatgtaaa	aatgagtgcg	ggctactgcc	ctttgtcatt	agaatttgtg	480
tctgtgtgta	ttgtttataa	aaataatata	aaattgggtt	tgagggagaa	agtaacgagt	540
gtgaacgatg	gaggacccat	ggaactttca	gaagaagttg	ttgatgagtt	catggagaa	600
gttccaatgt	cggttagact	cgcaaagttt	cgaaccaa	cctcaaaaag	aggtccgaaa	660
aataataata	atttaggtaa	ggggcgttca	ggcgggaaggc	ctaaaccaa	aagttttgat	720
gaagttggaa	aagagtttga	taatttgatt	gaagatgaag	ccgagacgtc	ggtcgcggat	780
tctgattcgt	attaa					795

<210> 13

<211> 795

<212> DNA

<213> Artificial Sequence derived from tobacco mosaic virus & tomato
mosaic virus

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 13

atgggtcttag	ttgttaaagg	taaggtaa	attaatgagt	ttatcgatct	gtcaaagtct	60
gagaaacttc	tcccgtcgat	gttcacgcct	gtaaggagtg	ttatggtttc	aaaggttgat	120
aagattatgg	tccatgaaaa	tgaatcattg	tctgaagtaa	atctcttaaa	aggtgtaaaa	180
cttatagaag	gtgggtatgt	ttgcttagtt	ggctcttggtg	tgtccggtga	gtggaattta	240
ccagataatt	gccgtgggtg	tgtgagtgtc	tgcattgggtg	acaagagaat	ggaaagagcg	300
gacgaagcca	cactgggggtc	atattacact	gctgctgcta	aaaagcgggt	tcagttttaa	360
gtgggcccaa	attacggtat	tactaccag	gacgcgatga	aaaacgtctg	gcaggctcta	420
gtaaatatta	aaaatgtaaa	aatgagtgcg	ggctactgcc	ctttgtcatt	agaatttggtg	480
tctgtgtgta	ttgtttataa	aaataatata	aaattgggtt	tgagggagaa	agtaacgagt	540
gtgaacgatg	gaggacccat	ggaactttca	gaagaagttg	ttgatgagtt	catggagaat	600
gttccaatgt	cgatcagact	cgcaaagttt	cgaaccaa	cctcaaaaag	aggtccgaaa	660
aataataata	atttaggtaa	ggggcgttca	ggcggaaggc	ctaaaccaa	aagttttgat	720
gaagttgaaa	aagagtttga	taatttgatt	gaagatgaag	ccgagacgtc	ggtcgcggat	780
tctgattcgt	attaa					795

<210> 14
 <211> 796
 <212> DNA
 <213> Artificial Sequence derived from tobacco mosaic virus & tomato mosaic virus
 <220>
 <223> This sequence was derived by shuffling in accordance with the methodology of the present invention.
 <400> 14
 atggctctag ttgttaaagg taaggtaaata attaatgagt ttatcgatct gtcaaagtct 60
 gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaagggtgat 120
 aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgttaag 180
 cttattgata gtggatacgt ctgttttagcc ggtttggtcg tcacgggcga gtggaattta 240
 ccagataatt gccgtggtgg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg 300
 gacgaagcca cactgggggtc atattacact gctgctgcta aaaagcgggt tcagttcaag 360
 gtcgttccca aattacggta ttactacca ggatgcagaa aagaacatat ggcaggctct 420
 agtaaataatt aaaaatgtaa aaatgagtg cggctactgc ccgctttctc tggagtttgt 480
 gtctgtgtgt attgtttata aaaataatat aaaattgggt ttgagggaga aagtaacgag 540
 tgtgaacgat ggaggacca tggaactttc agaagaagtt gttgatgagt tcatggagaa 600
 tgttccaatg tcggttagac tcgcaaagtt tcgaaccaa tcctcaaaaa gaggtccgaa 660
 aaataataat aatttaggta aggggcgttc aggcggaagg cctaaaccaa aaagttttga 720
 tgaagttgaa aaagagtttg ataatttgat tgaggatgat tcggaggcta ctgtcgccga 780
 ttctgattcg tattaata 796

<210> 15
 <211> 795
 <212> DNA
 <213> Artificial Sequence derived from tobacco mosaic virus & tomato mosaic virus
 <220>
 <223> This sequence was derived by shuffling in accordance with the methodology of the present invention.
 <400> 15
 atggctctag ttgttaaagg aaaagtgaat attaatgagt ttatcgatct gtcaaagtct 60
 gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaagggtgat 120
 aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgttaaaa 180
 cttatagaag gtgggtatgt ttgcttagtt ggtcttggtg tgtccggcga gtggaattta 240
 ccagataatt gccgtggtgg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg 300

gacgaagcca cactgggggtc atattacact gctgctgcaa agaaaagatt tcagttcaag 360
 gtcgttccca attatgctat aaccacccag gatgcagaaa agaacatatg gcgggtctta 420
 gtaaataatta aaaatgtaaa aatgagtgcg ggctactgcc cgctttctct ggagtttggtg 480
 tctgtgtgta ttgtttataa aaataatata aaattggggtt tgagggagaa agtaacgagt 540
 gtgaacgatg aaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
 gttccaatgt cgatcaggct cgcaaagttt cgaaccaaat cctcaaaaag aggtccgaaa 660
 aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat 720
 gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
 tctgattcgt actaa 795

<210> 16
 <211> 888
 <212> DNA
 <213> Selaginella lepidophylla

<400> 16
 atggcaacga ccaagacgag cgggatggcg ctggctttgc tcctcgtcgc cgccctggcc 60
 gtgggagctg cggcctgggg gaaagagggc catcgctca cttgtatggt cgccgagccc 120
 tttctaagct ctgaatccaa gcaagctgtg gaggagcttc tctctggaag agatctcccg 180
 gacttgtgtt catgggccga tcagattcga agatcgtata agtttagatg gactggtcct 240
 ttgcactaca tcgatactcc agacaacctc tgcacctatg actatgatcg tgactgccac 300
 gattcccatg ggaagaagga cgtgtgtgtc gctgggtggga tcaacaatta ctcgtcgcag 360
 ctggaaacgt ttctagattc agagagctcg tcgtataact tgaccgagggc gctgctcttc 420
 ctggctcact ttgtcgggga tatacaccag cccttgacg tagcatttac gagtgatgcc 480
 ggaggcaatg gcgtgcacgt ccgctgggtt ggacgaaagg ccaacttgca tcacgtctgg 540
 gatacagaat ttatttctag agccaatcgt gtgtactacc acgacatttc caagatgctc 600
 cggaacatta ccaggagcat aactaagaag aatttcaata gttggagcag atgtaagact 660
 gatccggcgg cttgtattga tagttatgcg acagaaagta tagatgcttc ttgcaactgg 720
 gcatacaaag acgcacccga cggaagctct ctagatgatg attacttctc ttcacgcctt 780
 ccaattgttg agcagcgtct tgctcaaggg ggcgtcaggc tggcgtcaat actcaacagg 840
 atttttggag gagcaaagtc gaacaggctc agtcgctcaa gcatgtag 888

<210> 17
 <211> 3637
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Encodes cycle 3 GFP

<400> 17

gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt	60
caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa	120
ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt	180
gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt	240
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt	300
ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg	360
tattatcccc tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga	420
atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa	480
gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga	540
caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa	600
ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca	660
ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta	720
ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac	780
ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc	840
gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag	900
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgtgaga	960
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt	1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata	1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag	1140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa	1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt	1260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc	1320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa	1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa	1440
gacgatagtt accggataag gcgcagcggg cgggctgaac ggggggttcg tgcacacagc	1500
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa	1560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa	1620
caggagagcg cacgaggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg	1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc	1740

tatggaaaaa cgccagcaac gcggcctttt tacgggttctt ggcctttttgc tggcctttttg 1800
ctcacatgtt ctttcctgcg ttatcccttg attctgtgga taaccgtatt accgcctttg 1860
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg 1920
aagcggaaga gcgccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat 1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg 2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt 2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg 2160
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct gggtagcgat gagtaaagga 2220
gaagaacttt tcaactggagt tgtcccaatt cttgttgaat tagatgggtga tgttaatggg 2280
caciaatttt ctgtcagtgg agaggggtgaa ggtgatgcta catacggaaa gcttaccctt 2340
aaatttattt gcactactgg aaaactacct gttccatggc caacacttgt cactactttc 2400
tcttatgggtg ttcaatgctt ttcccgttat ccggatcata tgaaacggca tgactttttc 2460
aagagtgcc a tgcgcgaagg ttatgtacag gaacgcacta tatctttcaa agatgacggg 2520
aactacaaga cgcgtgctga agtcaagttt gaaggtgata cccttggtta tcgtatcgag 2580
ttaaaaggta ttgattttta agaagatgga aacattctcg gacacaaact cgagtacaac 2640
tataactcac acaatgtata catcacggca gacaaacaaa agaattggaat caaagctaac 2700
ttcaaaattc gccacaacat tgaagatgga tccgttcaac tagcagacca ttatcaacaa 2760
aatactccaa ttggcgatgg ccctgtcctt ttaccagaca accattacct gtcgacacaa 2820
tctgcccttt cgaaagatcc caacgaaaag cgtgaccaca tggtccttct tgagtttgta 2880
actgctgctg ggattacaca tggcatggat gaactataca aataagaatt cctgcagccc 2940
gggggatcca ctagttctag agcggccgcc accgcggtgg agctccaatt cgccctatag 3000
tgagtcgtat tacgcgcgct cactggccgt cgtttttacaa cgtcgtgact gggaaaaccc 3060
tggcgttacc caacttaatc gccttgcagc acatccccct ttcgccagct ggcgtaatag 3120
cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggga 3180
cgcgccctgt agcggcgcat taagcgcggc ggggtgtggtg gttacgcgca gcgtgaccgc 3240
tacacttgcc agcgccttag cgcccgctcc tttcgttttc ttcccttctt ttctcgccac 3300
gttcgcgggc tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag 3360
tgctttacgg cacctcgacc ccaaaaaact tgattagggt gatggttcac gtagtgggcc 3420
atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg 3480
actcttgctc caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata 3540

agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac aaaaatttaa 3600

cgcgaaatttt aacaaaatat taacgcttac aatttag 3637

<210> 18

<211> 796

<212> DNA

<213> Tobamovirus Cg

<400> 18

atgtcttacg agcctaaagt gagcgacttc cttgctctta cgaaaaagga ggaaatttta 60

cccaaggctc ttacgaggtt aaagactgtc tctattagta ctaaggatgt tatatctgtt 120

aaggattctg agtcctctgt tgatatagat ttactagtta atgtgccatt agataagtat 180

agatatgtgg gtgttttagg tgttgttttt accggtgagt ggtagtgcc ggatttcggt 240

aaaggtggag taacagtgag cgtgattgac aaacggcttg agaactccaa agagtgcata 300

attggtacgt acagagctgc tgcgaaagac aaaaggttcc agttcaagct gggtccaaat 360

tacttcgtgt ctgttgacaga tgccaagcga aaaccgtggc aagttcatgt gcgtattcaa 420

aatttaagga ttgaagctgg atggcaacct ctggccttag aggtgggttc tgttgctatg 480

gtcactaata acgtgggtgt taagggtttg agagaaaagg tcatcgcagt gaatgatccg 540

aatgtcgaag gtttcgaagg cgtgggttgac gatttcgtcg attcggtcgc agcattcaag 600

gcgggtgaca ctttcagaaa gaaaaagaaa aggattggag gaaaggatgt aaataataat 660

aagtttagat atagaccgga gagatacgcc ggtcaggatt cgttaaatta taaagaagaa 720

aacgtcttac aacatcacga actcgaatca gtaccagtat ttcgcagcga cgtgggcaga 780

gccacagcg atgctt 796

<210> 19

<211> 823

<212> DNA

<213> Tobamovirus Ob

<400> 19

atgtcaaagg ctattgtcaa gatcgatgaa ttcattaaat tatccaagtc tgaagagggt 60

ttaccttctg cattcacaag aatgaagtcg gtcagagtct caacagtgga taagataatg 120

gccaaagaga atgacaatat ttccgaagta gatttactta agggtggtta gttagttaaa 180

aatggttatg tttgttttagt aggtcttgtg gtgtcaggag agtggaattt acccgacaac 240

tgcagagggt gtgtaagtat ctgtctgata gacaaacgta tgcaacgtca taacgaagct 300

actttagggt cgtacactac caaagccagc aagaaaaact ttcggttcaa gcttataccg 360

aattactcga taacctctca agatgctgaa aggcgtcctt gggaagttat ggtaaatatt 420

cgtgggtgtg ctatgtccga aggttgggtg ccattatcct tagagttcgt ttctgtttgt 480

attgttcata aaaacaatgt tagaaagggc ctaagagaga aggtgactgc cgtgtccgaa 540
gacgacgcta tagaactcac agaagagggt gttgatgagt ttatagaagc cgtaccgatg 600
gcgcgacgtt tgcagaactt gagaaaaccc aagtacaaca aagaaaaaga aaataaaaat 660
ttgaataata aaaatagtat aggagtttcc aaacctgtcg gtttggaaag aaataaagta 720
aggagtgtag ttagaaaagg ggtaggagt gatagtagtt taggtgtgac tgatatgagt 780
caggacggta gctcaagcga gatatcatcc gattcgttta ttt 823

<210> 20
<211> 769
<212> DNA
<213> Tobacco mosaic virus-U2

<400> 20
atggctgtta gtctcagaga tactgtcaaa attagcgagt tcattgatct ttcgaaacag 60
gatgagatac ttccggcatt catgactaag gtcaagagcg tcagaatata gactgtggac 120
aagattatgg ctgttaagaa tgatagtctt tctgatgtag atttacttaa aggtgttaag 180
ttagttaaga atgggtacgt gtgcttagct ggtttggtag tgtctgggga gtggaatctc 240
ccggacaact gccgtggtgg tgtcagtgtt tgtattgtag ataagagaat gaaaaggagt 300
aaggaggcaa cgctgggtgc gtatcacgcc cctgcttgca aaaagaattt ttcctttaag 360
ctaattcccta attattcaat aacatccgag gatgctgaga agcacccatg gcaagtatta 420
gtgaatatca aaggagtggc tatggaagaa ggatactgtc ctttatcttt ggagttcgtt 480
tcaatttgtg tagtacataa aaataatgta agaaaagggt tgagggaacg tattttgaga 540
gtaacagacg gctcgccaat tgaactcact gaaaaagttg ttgaggagtt catagatgaa 600
gtaccaatgg ctgtgaaact cgaaagggtc cggaaaacaa aaaagagagt ggtaggtaat 660
agtgttaata ataagaaaat aaataatagt ggtaagaaag gtttgaaagt tgaggaaatt 720
gaggataatg taagtgatga cgagtctatc gcgtcatcga gtacgtttt 769

<210> 21
<211> 808
<212> DNA
<213> Artificial Sequence derived from tomato mosaic virus and TMV-Cg

<220>
<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 21
atggctctag ttgttaaagg taaggtaaatt attaagagt ttatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120

```

aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
cttatagaag gtgggtatgt ttgcttagtt ggtcttggtg tgtccggtga gtggaattta 240
ccagataatt gccgtggttg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg 300
gacgaagcca cactgggggc atattacact gctgctgcta aaaagcgggt tcagttttaa 360
gtggtcccaa attacggtat tactacaaag gatgcagaaa agaacatatg gcaagttcat 420
gtgcgtattc aaaatttaag gattgaagct ggatggcaac ctctggcctt agaggtgggt 480
tctggtgcta tggctactaa taacgtgggt gttaagggtt tgagagaaaa ggcatcgcga 540
gtgaatgatc cgaatgtcga aggtttcgaa ggcgtggttg acgatttcgt cgattcggtc 600
gcagcattca aggcggttga cactttcaga aagaaaaaga aaaggattgg aggaaaggat 660
gtaaataata ataagtttag atatagaccg gagagatacg ccggtcagga ttcgttaaata 720
tataaagaag aaaacgtctt acaacatcac gaactcgaat cagtaccagt atttcgcagc 780
gacgtgggca gagccacag cgatgctt 808

```

<210> 22

<211> 799

<212> DNA

<213> Artificial Sequence derived from tomato mosaic virus and TMV-Cg

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 22

```

atgtcttacg agcctaaagt gagcgacttc cttgctctta cgaaaaagga ggaaatttta 60
cccaaggctc ttacgagggt aaagactgtc tctattagta ctaaggatgt tatactctgtt 120
aaggattctg agtccctgtg tgatatagat ttactagtta atgtgccatt agataagtat 180
agatatgtgg gtgttttagg tgttggtttt accggtgagt ggaatttacc agataattgc 240
cgtggtgggt tgagtgtctg catggttgac aagagaatgg aaagagcgga cgaagccaca 300
ctgggggtcat attacactgc tgctgcgaaa gacaaaaggt tccagttcaa gctgggtcca 360
aattacttcg tgtctgttgc agatgccaaag cgaaaaccgt ggcaagttca tgtgcgtatt 420
caaaatttaa ggattgaagc tggatggcaa cctctggcct tagaggtgggt ttctgttgct 480
atggtcacta ataacgtggg tggttaagggt ttgagagaaa aggtcatcgc agtgaatgat 540
ccgaatgtcg aaggtttcga aggcgtgggt gacgatttcg tcgattcggg cgcagcattc 600
aaggcgggtg acactttcag aaagaaaaag aaaaggattg gaggaaagga tgtaaataat 660
aataagttta gatatagacc ggagagatac gccggtcagg attcgttaaa ttataaagaa 720
gaaaacgtct tacaacatca cgaactcgaa tcagtaccag tatttcgcag cgacgtgggc 780

```

agagcccaca gcgatgctt

799

<210> 23

<211> 823

<212> DNA

<213> Artificial Sequence derived from ToMV & TMV-Ob

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 23

```

aaataaacga atcggatgat atctcgcttg agctaccgctc ctgactcata tcagtcacac      60
ctaaactact atcactccta accccttttc taactacact ccttacttta tttctttcca      120
aaccgacagg tttggaaact cctatactat ttttattatt caaattttta ttttcttttt      180
ctttgttgta cttgggtttt ctcaagttct gcaaacgctc cgccatcggt acggcttcta      240
taaactcatc aacaacctct tctgtgagtt ctatagcgct gtcttcggac acggcagtca      300
ccttctctct tagacccttt ctaacattgt ttttatgaac aatacaaa gaaacgaact      360
ctaaggataa tggacaccaa ccttcggaca tagccacacc acgaatattt accataactt      420
cccaaggacg cctttcagca tcttgagagg ttatcgagta attcggtata agcttgaacg      480
aaaagttttt cttgctggct ttggtagtgt acgaacctaa agtagcttcg ttatgacggt      540
gcatacgttt gtctatcaga cagatactta caccacctct gcagttgtcg ggtaaattcc      600
actctcctga caccacaaga cctactaaac aaacataacc accttctata agttttacac      660
cttttaagag atttacttca gacaatgatt cattctcttt ggccattatc ttatccactg      720
ttgagactct gaccgacttc attcttgtga atgcagaagg taaaacctct tcagacttgg      780
ataatttaat gaattcatcg atcttgacaa tagcctttga cat                        823

```

<210> 24

<211> 792

<212> DNA

<213> Artificial Sequence derived from ToMV & TMV-Ob

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 24

```

aatacgaatc agaatccgcg accgacgtct cggcttcac tccaatcaaa ttatcaaact      60
ctttttcaac ttcataaaaa ctttttggtt taggccttcc gcctgaacgc cccttaccta      120
aattattatt atttttcgga cctctttttg aggatttgg tcgaaacttt gcgagtctaa      180
ccgacattgg aacattctcc atgaactcat caacaacctc ttctgtgagt tctatagcgt      240
cgtcttcgga cacggcagtc accttctctc ttagaccctt tctaacattg tttttatgaa      300

```

caatacaaac agaaacgaac tctaatagaca aagggcagta gcccgcactc attttttacat 360
 ttttaatat tactaagacc tgccatatgt tcttttctgc atcctttgta gtaataaccgt 420
 aatttgggac cactttaaac tgaaaccgct ttttagcagc agcagtgtaa tatgacccca 480
 gtgtggcttc gtccgctctt tccattctct tgtcaacat gcagacactc acaccaccac 540
 ggcaattatc tggtaaattc cactctcttg acaccacaag acctactaaa caaacataac 600
 catttttaac taacttaaca cccttaagag atttacttcg gacaatgatt cattttcatg 660
 gaccataatc ttatcaacct ttgaaacat aacactcttt acaggcgtga atgcagaagg 720
 taaaacctct tcagactttg acagatcgat aaactcatta atattttacct tacctttaac 780
 aactagagcc at 792

<210> 25

<211> 769

<212> DNA

<213> Artificial Sequence derived from ToMV & TMV-U2

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 25

aatacgaatc agaatccgcg atagactcgt catcacttac attatcctca atttcctcaa 60
 ctttcaaacc tttcttacca ctattattta ttttcttatt attaacacta ttacctacca 120
 ctctcttttt tgttttccgg aacctttcga gtttcacagc cattgggtact tcatctatga 180
 actcatcaac aacttcttct gaaagtcca tgggtcctcc atcgttcaca ctcgttactt 240
 tctccctcaa acccaatttt atattatttt tataaacaat acacacagac acaaattcta 300
 aagataaagg gcagtatcct tcttccatag ccactccttt gatattcact aatacttgcc 360
 atgggtgctt ttctgcatcc tcggatgtta ttgaataatt aggaccact ttaaactgaa 420
 accgcttttt agcagcaggg gcgtgatacg caccagcgt tgcctcctta ctctttcca 480
 ttctcttgtc aaccatgcag aactcacac caccacggca gttgtccggg agattccact 540
 caccggacac aacaagacca actaagcaaa catacccacc ttctataagt ttacacctt 600
 ttaagagatt tacttcagac aatgattcat tttcatggac cataatctta tcaaccttg 660
 aaaccataac actctttaca ggcgtgaaca tcgacgggag aagtttctca gactttgaca 720
 gatcgataaa ctcatataa ttaccttac ctttaacaac tagagccat 769

<210> 26

<211> 772

<212> DNA

<213> Artificial Sequence derived from ToMV & TMV-U2

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 26

```

aatacgaatc agaatccgcg accgacgtct cggcttcact tacattatcc tcaatttcct      60
caactttcaa aactttctta ccactattat ttattttctt attattaaca ctattaccta     120
ccactctctt ttttgttttc cggaaccttt cgagtttcac agccattggg acttcatcta     180
tgaactcatc aacaactttt tcagtgagtt caattggcga gccgtctggt actctcaaaa     240
tacgttccct caaacccaat tttatattat ttttataaac aatacacaca gacacaaatt     300
ctaatacaca agggcagtag cccgcactca tttttacatt tttaatatct actaagacct     360
gccatgggtg cttctcagca tcctcggatg ttattgaata attagggatt agcttaaagg     420
aaaaattctt tttgcaagca ggggcgtgat acgcaccag tgtggcttcg tccgctcttt     480
ccattctctt gtcaaccatg cagacactca caccaccag gcagttgtcc gggagattcc     540
actcaccgga cacaacaaga ccaactaagc acacgtaccc attcttaact aacttaacac     600
ctttaagtaa atctacatca gacaatgatt cattttcatg gaccataatc ttatcaacct     660
ttgaaaccat aacactcttt acaggcgtga acatcgacgg gagaagtttc tcagactttg     720
acagatcgat aaactcgcta attttgacag tatctctgag actaacagcc at              772

```

<210> 27

<211> 805

<212> DNA

<213> Artificial Sequence derived from ToMV & TMV-U1

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 27

```

atggctctag ttgttaaagg aaaagtgaat attaatgagt ttatcgatct gtcaaagtct      60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat     120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa     180
cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtga gtggaattta     240
ccagataatt gccgtggtgg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg     300
gacgaagcca ctctcggatc ttactacaca gcagctgcaa agaaaagatt tcagttcaag     360
gtcgttccca attatgctat aaccacccag gacgcgatga aaaacgtctg gcaagtttta     420
gttaatatta gaaatgtgaa gatgtcagcg ggtttctgtc cgctttctct ggagtttgtg     480
tctgtgtgta ttgtttataa aaataatata aaattggggt tgagggagaa agtaacgagt     540

```

gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggaagat 600
 gtcccaatgt cggtagact cgcaaagttt cgatctcgaa ccggaaaaaa gagtgatgtc 660
 cgcaaaggga aaaatagtag tagtgatcgg tcagtgccga acaagaacta tagaaatggt 720
 aaggattttg gaggaatgag ttttaaaaag aataatttaa tcgatgatga ttcggagacg 780
 tcggtcgcgg attctgattc gtatt 805

<210> 28

<211> 804

<212> DNA

<213> Artificial Sequence derived from ToMV & TMV-U1

<220>

<223> This sequence was derived by shuffling in accordance with the methodology of the present invention.

<400> 28

atggctctag ttgttaaagg aaaagtgaat atcaatgagt ttatcgacct gacaaagtct 60
 gagaaacttc tcccgatgat gtttaccct gtaaagagtg ttatgggtcc aaagttgata 120
 agattatggt tcatgagaat gagtcattgt caggggtgaa ccttcttaaa ggagttaagc 180
 ttattgatag tggatacgtc tgtttagccg gtttggtcgt cacgggagag tggaacttgc 240
 ctgacaattg ccgtgggtgg gtgagcgtgt gtctgggtga caagagaatg gaaagagcgg 300
 acgaagccac actggggtca tattacactg ctgctgctaa aaagcggttt cagttcaagg 360
 tcgttcccaa ttatgctata accaccagc atgcagaaaa gaacatatgg caggtcttag 420
 taaatattaa aaatgtgaag atgagtgcgg gctactgcc tttgtcatta gaatttgtgt 480
 cgggtgtgat tgtttataga aataatataa aattgggttt gagagagaaa gtaacgagtg 540
 tgaacgatgg agggcccatg gaacttacag aagaagtcgt tgatgagttc atggaagatg 600
 tccctatgtc gatcaggctt gcaaagtttc gatctcgaa cctcaaaaag agtgatgtcc 660
 gcaaaggga aaatagtagt agtgatcggc cagtgccga caagaactat agaaatgtta 720
 aggattttgg aggaatgagt tttaaaaaga ataatttaac cgatgatgat tcggaggcta 780
 ctgtcgcgga ttctgattcg tttt 804

<210> 29

<211> 717

<212> DNA

<213> Aequorea victoria GFP ORF

<400> 29

atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggt 60
 gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggatgatg aacatacggg 120
 aaacttacc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180

```

gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg      240
catgactttt tcaagagtgc catgccccgaa ggttatgtac aggaaagaac tatatctttc      300
aaggatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaagggtga tacccttggt      360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa      420
ttggaataca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga      480
atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcggttca actagcagac      540
cattatcaac aaaatactcc aattggcgat ggcctgtcc ttttaccaga caaccattac      600
ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt      660
cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa          717

```

<210> 30

<211> 717

<212> DNA

<213> Aequorea victoria GFP Cycle 3 ORF

<400> 30

```

atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggg      60
gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggatgatgc tacatacggg      120
aagcttacct ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt      180
gtcactactt tctcttatgg tgttcaatgc ttttcccgtt atccggatca tatgaaacgg      240
catgactttt tcaagagtgc catgccccgaa ggttatgtac aggaacgcac tatatctttc      300
aaagatgacg ggaactacaa gacgcgtgct gaagtcaagt ttgaagggtga tacccttggt      360
aatcgtatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa      420
ctcgagtaca actataactc acacaatgta tacatcacgg cagacaaaca aaagaatgga      480
atcaaagcta acttcaaaat tcgccacaac attgaagatg gatccggttca actagcagac      540
cattatcaac aaaatactcc aattggcgat ggcctgtcc ttttaccaga caaccattac      600
ctgtcgacac aatctgccct ttcgaaagat cccaacgaaa agcgtgacca catggtcctt      660
cttgagtttg taactgctgc tgggattaca catggcatgg atgaactata caaataa          717

```

<210> 31

<211> 3637

<212> DNA

<213> Artificial Sequence

<220>

<223> Plasmid encoding wild type Aequorea victoria GFP Cycle 3 ORF

<400> 31

```

gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatctttc taaatacatt      60

```

caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa	120
ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt	180
gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt	240
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt	300
ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcg	360
tattatcccc tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga	420
atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa	480
gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac ttactttctga	540
caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa	600
ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca	660
ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta	720
ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac	780
ttctgcgctc ggcccttccg gctggctggg ttattgctga taaatctgga gccggtgagc	840
gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag	900
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga	960
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt	1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata	1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag	1140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa	1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt	1260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc	1320
cgtagttagg ccaccacttc aagaactctg tagcacccgc tacatacctc gctctgctaa	1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa	1440
gacgatagtt accggataag gcgcagcggg cgggctgaac ggggggttcg tgcacacagc	1500
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa	1560
gcgccacgct tccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa	1620
caggagagcg cacgaggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg	1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc	1740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg	1800
ctcacatggt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg	1860

agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg	1920
aagcggaaga gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat	1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg	2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt	2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	2160
ccaagcgcg c aattaaccct cactaaaggg aacaaaagct gggtaaccgat gagtaaagga	2220
gaagaacttt t cactggagt tgtcccaatt cttgttgaat tagatgggtga tgttaatggg	2280
cacaaatttt ctgtcagtgg agaggggtgaa ggtgatgcaa catacggaaa acttaccctt	2340
aaatttattt gcactactgg aaaactacct gttccatggc caacacttgt cactactttc	2400
tcttatgggtg ttcaatgctt ttcaagatac ccagatcata tgaaacggca tgactttttc	2460
aagagtgcc a tgcccgaagg ttatgtacag gaaagaacta tatttttcaa ggatgacggg	2520
aactacaaga cacgtgctga agtcaagttt gaagggtgata cccttggtta tagaatcgag	2580
ttaaaaggta ttgattttta agaagatgga aacattcttg gacacaaatt ggaatacaac	2640
tataactcac acaatgtata catcatggca gacaaacaaa agaatggaat caaagttaac	2700
ttcaaaatta gacacaacat tgaagatgga agcgttcaac tagcagacca ttatcaacaa	2760
aatactccaa ttggcgatgg ccctgtcctt ttaccagaca accattacct gtccacacaa	2820
tctgcccttt cgaaagatcc caacgaaaag agagaccaca tggtccttct tgagtttgta	2880
acagctgctg ggattacaca tggcatggat gaactataca aataagaatt cctgcagccc	2940
gggggatcca ctagttctag agcggccgcc accgcggtgg agctccaatt cgccctatag	3000
tgagtcgtat tacgcgcgct cactggccgt cgtttttacaa cgtcgtgact gggaaaaccc	3060
tggcgttacc caacttaatc gccttgcagc acatccccct ttcgccagct ggcgtaatag	3120
cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggga	3180
cgcgccctgt agcggcgcac taagcgcggc ggggtgtggtg gttacgcgca gcgtgaccgc	3240
tacacttgcc agcgccctag cgcccgctcc tttcgctttc ttccttccct ttctcgccac	3300
gttcgccggc tttccccgtc aagctctaaa tcggggggctc cctttagggg tccgatttag	3360
tgctttacgg cacctcgacc ccaaaaaact tgattagggg gatgggtcac gtagtgggcc	3420
atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg	3480
actcttggtc caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata	3540
agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac aaaaatttaa	3600
cgcgaatttt aacaaaatat taacgcttac aatttag	3637

<210> 32
 <211> 3637
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Encodes Cycle 3 BFP gene

<400> 32
 gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt 60
 caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa 120
 ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt 180
 gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt 240
 tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt 300
 ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg 360
 tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga 420
 atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 480
 gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga 540
 caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa 600
 ctgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 660
 ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaaactggc gaactactta 720
 ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 780
 ttctgcgctc ggcccttccg gctggctggg ttattgctga taaatctgga gccggtgagc 840
 gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag 900
 ttatctacac gacggggagt caggcaacta tggatgaacg aatagacag atcgctgaga 960
 taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt 1020
 agattgattht aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 1080
 atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 1140
 aaaagatcaa aggatcttct tgagatcctt tttttctgcy cgtaatctgc tgcttgcaaa 1200
 caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 1260
 ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 1320
 cgtagttagg ccaccacttc agaactctg tagcaccgcc tacatacctc gctctgctaa 1380
 tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa 1440
 gacgatagtt accggataag gcgcagcggg cgggctgaac ggggggttcg tgcacacagc 1500
 ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 1560

gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa	1620
caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg	1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc	1740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg	1800
ctcacatggt ctttcctgcg ttatccctcg attctgtgga taaccgtatt accgcctttg	1860
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg	1920
aagcgggaaga gcgccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat	1980
gcagctggca cgacagggtt cccgactgga aagcgggcag tgagcgcaac gcaattaatg	2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt	2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	2160
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct gggtagcgat gagtaaagga	2220
gaagaacttt tcaactggagt tgtcccaatt cttgttgaat tagatggtga tgttaatggg	2280
cacaaatfff ctgtcagtgg agaggggtgaa ggtgatgcta catacggaaa gcttacactt	2340
aaatftatft gcactactgg aaaactacct gttccatggc caacacttgt cactactttc	2400
tctcatggtg ttcaatgctt ttctcgttat ccggatcata tgaaacggca tgactttttc	2460
aagagtgcc a tgcccgaagg ttatgtacag gaacgcacta tatctttcaa agatgacggg	2520
aactacaaga cgcgtgctga agtcaagttt gaagggtgata cccttgttaa tcgtatcgag	2580
ttaaaaggta ttgattttta agaagatgga aacattctcg gacacaaact cgagtacaac	2640
tttaactcac acaatgtata catcacggca gacaaacaaa agaattggaat caaagctaac	2700
ttcaaaattc gccacaacat tgaagatgga tccgttcaac tagcagacca ttatcaacaa	2760
aatactccaa ttggcgatgg ccctgtcctt ttaccagaca accattacct gtcgacacaa	2820
tctgcccttt cgaaagatcc caacgaaaag cgtgaccaca tggtccttct tgagtttgta	2880
actgctgctg ggattacaca tggcatggat gaactataca aataagaatt cctgcagccc	2940
gggggatcca ctagttctag agcggccgcc accgcggtgg agtccaatt cgccctatag	3000
tgagtcgtat tacgcgcgct cactggccgt cgttttacia cgtcgtgact gggaaaaccc	3060
tggcgttacc caacttaatc gccttgcagc acatccccct ttcgccagct ggcgtaatag	3120
cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggga	3180
cgcgccctgt agcggcgc at taagcgcggc ggggtgtggtg gttacgcgca gcgtgaccgc	3240
tacacttgcc agcgccttag cgcccgctcc tttcgctttc ttcccttctt ttctcgccac	3300
gttcgcgggc tttccccgtc aagctctaaa tcggggggctc cctttagggt tccgatttag	3360
tgctttacgg cacctcgacc ccaaaaaact tgattagggt gatggttcac gtagtgggcc	3420

atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg 3480
 actcttggtc caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata 3540
 agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac aaaaatttaa 3600
 cgcgaaatatt aacaaaatat taacgcttac aatttag 3637

<210> 33
 <211> 717
 <212> DNA
 <213> Aequorea victoria BFP Cycle 3 ORF

<400> 33
 atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggt 60
 gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggatgatgc tacatacggg 120
 aagcttacac ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
 gtcactactt tctctcatgg tgttcaatgc ttttctcgtt atccggatca tatgaaacgg 240
 catgactttt tcaagagtgc catgcccga ggttatgtac aggaacgcac tatatctttc 300
 aaagatgacg ggaactacaa gacgcgtgct gaagtcaagt ttgaaggatga tacccttggt 360
 aatcgtatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa 420
 ctcgagtaca actttaactc acacaatgta tacatcacgg cagacaaaca aaagaatgga 480
 atcaaagcta acttcaaaat tcgccacaac attgaagatg gatccgttca actagcagac 540
 cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
 ctgtcgacac aatctgccct ttcgaaagat cccaacgaaa agcgtgacca catggtcctt 660
 cttgagtttg taactgctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 34
 <211> 295
 <212> PRT
 <213> Selaginella lepidophylla

<400> 34

Met Ala Thr Thr Lys Thr Ser Gly Met Ala Leu Ala Leu Leu Leu Val
 1 5 10 15

Ala Ala Leu Ala Val Gly Ala Ala Ala Trp Gly Lys Glu Gly His Arg
 20 25 30

Leu Thr Cys Met Val Ala Glu Pro Phe Leu Ser Ser Glu Ser Lys Gln
 35 40 45

Ala Val Glu Glu Leu Leu Ser Gly Arg Asp Leu Pro Asp Leu Cys Ser

<210> 35

<211> 19

<212> PRT

<213> Apium graveolens fragment of Cel I expressed by TMV

<400> 35

Asp	Met	Cys	Val	Ala	Gly	Ala	Ile	Gln	Asn	Phe	Thr	Ser	Gln	Leu	Gly
1				5				10						15	

His Phe Arg

Claims:

1. An *in vitro* method of obtaining a peptide sequence having a desired functional property, comprising:

- a) preparing at least one heteroduplex polynucleotide sequence wherein said heteroduplex polynucleotide sequence has at least two non-complementary nucleotide base pairs;
 - b) mixing copies of the heteroduplex polynucleotide sequence with an effective amount of a mismatch endonuclease RES I having the sequence of SEQ ID NO: 34 or a combination of mismatch endonucleases including a mismatch endonuclease RES I having the sequence of SEQ ID NO: 34, a proofreading enzyme, dNTPs, and a ligase enzyme;
 - c) allowing sufficient time for non-complementary nucleotide base pairs to be converted to complementary base pairs, wherein a population of polynucleotide sequence variants result;
 - d) expressing polynucleotide sequence variants to generate peptide sequence variants; and
 - e) screening or selecting the polypeptide sequence variants for the desired functional property.
2. A method according to claim 1 wherein the heteroduplex polynucleotide sequence encodes endonuclease activity.
3. A method according to claim 1 or 2 wherein expressing polynucleotide variants comprises transfecting a host organism with a recombinant plasmid or a recombinant viral vector that encodes one of the polynucleotide variants.
4. A method according to claim 1 or 2 wherein expressing polynucleotide variants comprises transforming a host organism with one of the polynucleotide variants.

5. A method of making a mismatch endonuclease RES I having the sequence of SEQ ID NO: 34 comprising:
 - a) transfecting a host plant, or host animal, or host yeast or host fungus or host bacterium with a recombinant viral vector that encodes a polynucleotide sequence for said RES I;
 - b) growing the host plant, or host animal, or host yeast or host fungus or host bacterium wherein the polynucleotide sequence for said RES I is expressed; and
 - c) extracting said RES I from the host.

6. A method for determining a mutation in a target sequence of a single stranded polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including said target sequence, wherein said polynucleotides are amplified, labeled with a detectable marker, hybridized to one another, subjected to the activity of an endonuclease and analyzed for the presence of said mutation, the improvement comprising the use of a mismatch endonuclease RES I having the sequence of SEQ ID NO: 34 or a sequence variant thereof of *Selaginella lepidophylla* origin, the activity of said mismatch endonuclease enzyme of *Selaginella lepidophylla* origin comprising:
 - a. detection of mismatches whether known or unknown between said hybridized polynucleotides;
 - b. catalytic formation of a substantially single-stranded nick at a target sequence containing a mismatch; and
 - c. recognition of a mutation in a target polynucleotide sequence, said recognition being substantially unaffected by flanking polynucleotide sequences.

7. A method according to claim 6, wherein the activity of said mismatch endonuclease enzyme further comprises d) recognition of polynucleotide loops and insertions between said hybridized polynucleotides.
8. An *in vitro* method of obtaining a polynucleotide sequence encoding a desired functional property, comprising:
 - a) preparing at least one heteroduplex polynucleotide sequence wherein said heteroduplex polynucleotide sequence has at least two non-complementary nucleotide base pairs;
 - b) mixing copies of the heteroduplex polynucleotide sequence with an effective amount of a mismatch endonuclease RES I having the sequence of SEQ ID NO: 34, or a combination of mismatch endonucleases including said RES I, a proofreading enzyme, dNTPs, and a ligase enzyme;
 - c) allowing sufficient time for non-complementary nucleotide base pairs to be converted to complementary base pairs, wherein a population of polynucleotide sequence variants result; and
 - d) screening or selecting said polynucleotide sequence variants for the desired functional property.
9. A nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 16.
10. A polypeptide comprising the amino acid sequence of SEQ ID NO: 34.

FIG. 1

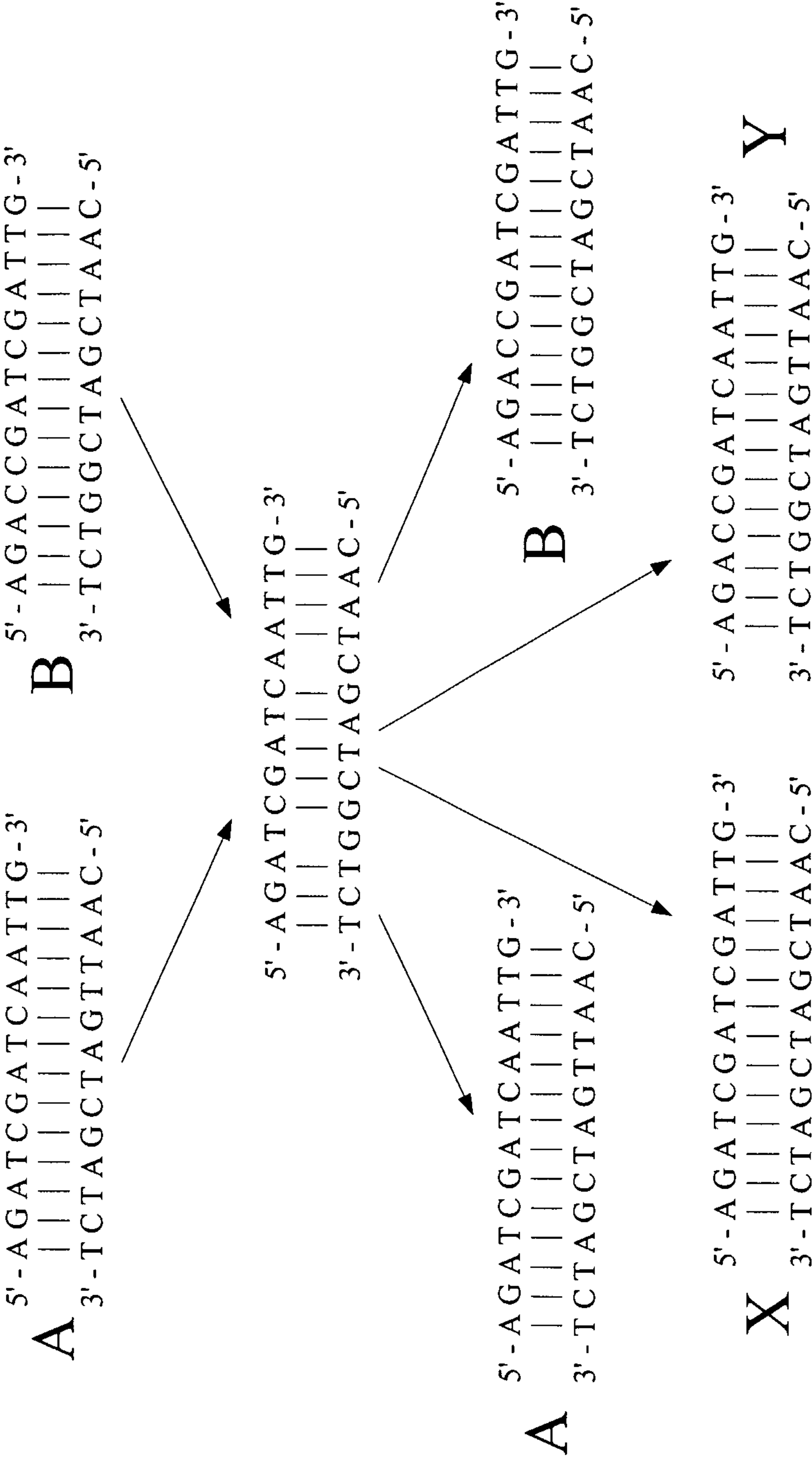


FIG. 2A



FIG. 2B

Possible Strand Combinations	Possible +/- Strand Combinations	Partially Complementary Populations
1+/2-		
1+/3+		
1+/4-		
2-/3+		
2-/4-		
3+/4-		

3/18

FIG. 3

```
atggcaacga ccaagacgag cgggatggcg ctggctttgc tcctcgtcgc cgccctggcc
gtgggagctg cggcctgggg gaaagagggc catcgctca cttgtatggt cgccgagccc
tttctaagct ctgaatccaa gcaagctgtg gaggagcttc tctctggaag agatctccc
gacttgtgtt catgggccga tcagattcga agatcgtata agtttagatg gactggtcct
ttgcactaca tcgatactcc agacaacctc tgcacctatg actatgatcg tgactgccac
gattcccatg ggaagaagga cgtgtgtgtc gctggtggga tcaacaatta ctcgtcgcag
ctggaaacgt ttctagattc agagagctcg tcgtataact tgaccgaggc gctgctcttc
ctggctcact ttgtcgggga tatacaccag cccttgcacg tagcatttac gagtgatgcc
ggaggcaatg gcgtgcacgt ccgctggttt ggacgaaagg ccaacttgca tcacgtctgg
gatacagaat ttatttctag agccaatcgt gtgtactacc acgacatttc caagatgctc
cggaacatta ccaggagcat aactaagaag aatttcaata gttggagcag atgtaagact
gatccggcgg cttgtattga tagttatgcg acagaaagta tagatgcttc ttgcaactgg
gcatacaaag acgcaccgga cggaagctct ctagatgatg attacttctc ttcacgcctt
ccaattgttg agcagcgtct tgctcaaggg ggcgtcaggc tggcgtcaat actcaacagg
atTTTTGGAG gagcaaagtc gaacagggtcc agtcgctcaa gcatgtag
```

4/18

FIG. 4

Met Ala Thr Thr Lys Thr Ser Gly Met Ala Leu Ala Leu Leu Val
Ala Ala Leu Ala Val Gly Ala Ala Ala Trp Gly Lys Glu Gly His Arg
Leu Thr Cys Met Val Ala Glu Pro Phe Leu Ser Ser Glu Ser Lys Gln
Ala Val Glu Glu Leu Leu Ser Gly Arg Asp Leu Pro Asp Leu Cys Ser
Trp Ala Asp Gln Ile Arg Arg Ser Tyr Lys Phe Arg Trp Thr Gly Pro
Leu His Tyr Ile Asp Thr Pro Asp Asn Leu Cys Thr Tyr Asp Tyr Asp
Arg Asp Cys His Asp Ser His Gly Lys Lys Asp Val Cys Val Ala Gly
Gly Ile Asn Asn Tyr Ser Ser Gln Leu Glu Thr Phe Leu Asp Ser Glu
Ser Ser Ser Tyr Asn Leu Thr Glu Ala Leu Leu Phe Leu Ala His Phe
Val Gly Asp Ile His Gln Pro Leu His Val Ala Phe Thr Ser Asp Ala
Gly Gly Asn Gly Val His Val Arg Trp Phe Gly Arg Lys Ala Asn Leu
His His Val Trp Asp Thr Glu Phe Ile Ser Arg Ala Asn Arg Val Tyr
Tyr His Asp Ile Ser Lys Met Leu Arg Asn Ile Thr Arg Ser Ile Thr
Lys Lys Asn Phe Asn Ser Trp Ser Arg Cys Lys Thr Asp Pro Ala Ala
Cys Ile Asp Ser Tyr Ala Thr Glu Ser Ile Asp Ala Ser Cys Asn Trp
Ala Tyr Lys Asp Ala Pro Asp Gly Ser Ser Leu Asp Asp Asp Tyr Phe
Ser Ser Arg Leu Pro Ile Val Glu Gln Arg Leu Ala Gln Gly Gly Val
Arg Leu Ala Ser Ile Leu Asn Arg Ile Phe Gly Gly Ala Lys Ser Asn
Arg Ser Ser Arg Ser Ser Met

5/18

FIG. 5

gtggcacttt tgggggaaat gtgcgcgga cccctatttg tttatttttc taaatacatt caaatatgta
tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa
catttccgtg tggcccttat tccctttttt gcggcatttt gccttcctgt ttttgctcac ccagaaacgc
tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag
cggtaagatc cttgagagtt ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta
tgtggcgcggt tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga
atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg
cagtgtgcc ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaag
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga tcgttgggaa ccggagctga
atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaact
attaactggc gaactactta ctctagcttc cgggcaacaa ttaatagact ggatggaggc ggataaagt
gcaggaccac ttctgcgctc ggcccttccg gctggctggg ttattgctga taaatctgga gccggtgagc
gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac
gacggggagt caggcaacta tggatgaacg aaatagacag atcgtgaga taggtgcctc actgattaag
cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat ttttaattta
aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca
ctgagcgtca gaccccgtag aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc
tgcttgcaaa caaaaaaacc accgctacca gcggtgggtt gtttgccgga tcaagagcta ccaactcttt
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg
ccaccacttc aagaactctg tagcaccgcc tacatacttc gctctgctaa tcctgttacc agtggtgct
gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag gcgcagcgg
cgggctgaac ggggggttcg tgcacacagc ccagcttggg gcgaacgacc tacaccgaac tgagatacct
acagcgtgag ctatgagaaa ggcgccagct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc
agggtcggaa caggagagcg cagcagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg
ggtttcgcca cctctgactt gagcgtcgat ttttgatgat ctgctcaggg gggcggagcc tatggaaaaa
cgccagcaac gcggcctttt tacggttcct ggcccttttg tggccttttg ctacatgtt ctttctgcg
ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc cgcagccgaa
cgaccgagcg cagcagatca gtgagcgagg aagcgggaaga gcgccaata cgcaaaccgc ctctccccgc
gcgttggccg attcattaat gcagctggca cgacagggtt cccgactgga aagcgggcag tgagcgcaac
gcaattaatg tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg ccaagcgcgc
aattaaccct cactaaaggg aacaaaagct ggggtaccgat gagtaaagga gaagaacttt tcaactggag
tgtccaatt cttgttgaat tagatgggtga tgttaatggg cacaattttt ctgtcagtgg agagggtgaa
ggtgatgcta catacgaaa gcttaccctt aaatttattt gcactactgg aaaactacct gttccatggc
caacacttgt cactactttc tcttatgggt ttcaatgctt ttcccgttat ccggatcata tgaaacggca
tgactttttc aagagtgcc aagagtgcca agtcaagttt gaaggtgata cccttggttaa tcgtatcgag ttaaaaggta
aactacaaga cgcgtgctga agcattctcg gacacaaaact cgagtacaac tataactcac acaatgtata
ttgattttta agaagatgga aacattctcg caaagctaac ttcaaaattc gccacaacat tgaagatgga
catcacggca gacaaacaaa agaattggaat caaagctaac ttcaaaattc gccacaacat tgaagatgga
tccgttcaac tagcagacca ttatcaacaa aatactccaa ttggcgatgg cctgtcctt ttaccagaca
accattacct gtcgacacaa tctgcccttt cgaaagatcc caacgaaaag cgtgaccaca tggccttct
tgagtttgta actgctgctg ggattacaca tggcatggat gaactataca aataagaatt cctgcagccc
gggggatcca ctagtcttag agcggccgcc accgcgggtg agctccaatt cgccctatag tgagtctgat
tacgcgcgct cactggcgtt cgtttttaca cgtcgtgact gggaaaaccc tggcgttacc caacttaac
gccttgtagc acatccccct ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca
acagttgcgc agcctgaatg gcgaatggga cgcgcctgt agcggcgcat taagcgcggc ggggtgtggtg
gttacgcgca gctgaccgc tacacttgcc agcgccttag cgcccgctcc tttcgctttc ttccttctc
ttctcgccac gttcgccggc tttccccgct aagctctaaa tcgggggctc cctttagggt tccgatttag
tgctttacgg cacctcgacc ccaaaaaact tgattagggg gatgggtcac gtagtgggcc atcgccctga
tagacgggtt ttcgcccttt gacgttggag tccacgttct ttaatagtg actcttgctc caaactggaa
caacactcaa ccctatctcg gtctattctt ttgatttata agggattttg ccgatttcgg cctattgggt
aaaaatgag ctgatttaac aaaaatttaa cgcaattttt aacaaaatat taacgcttac aatttag

6/18

FIG. 6

```
atgtcttacg agcctaaagt gagcgacttc cttgctctta cgaaaaagga ggaaatttta
cccaaggctc ttacgagggt aaagactgtc tctattagta ctaaggatgt tatatctgtt
aaggattctg agtccctgtg tgatatagat ttactagtta atgtgccatt agataagtat
agatatgtgg gtgttttagg tgttggtttt accggtgagt ggtagtgcc ggatttcggt
aaaggtggag taacagtgag cgtgattgac aaacggcttg agaactcaa agagtgcata
attggtacgt acagagctgc tgcgaaagac aaaagggtcc agttcaagct gggtccaaat
tacttcgtgt ctgttgcaga tgccaagcga aaaccgtggc aagttcatgt gcgtattcaa
aatttaagga ttgaagctgg atggcaacct ctggccttag aggtgggttc tgttgctatg
gtcactaata acgtgggttg taagggtttg agagaaaagg tcatcgcagt gaatgatccg
aatgtcgaag gtttcgaagg cgtggttgac gatttcgctg attcggtcgc agcattcaag
gcggttgaca ctttcagaaa gaaaaagaaa aggattggag gaaaggatgt aaataataat
aagtttagat atagaccgga gagatacgcc ggtcaggatt cgttaaatta taaagaagaa
aacgtcttac aacatcacga actcgaatca gtaccagtat ttcgcagcga cgtgggcaga
gcccacagcg atgctt
```

7/18

FIG. 7

atgtcaaagg	ctattgtcaa	gatcgatgaa	ttcattaaat	tatccaagtc	tgaagagggt
ttaccttctg	cattcacaag	aatgaagtcg	gtcagagtct	caacagtgga	taagataatg
gccaaagaga	atgacaatat	ttccgaagta	gatttactta	aggggtgtta	gttagttaa
aatggttatg	tttgtttagt	aggtcttggt	gtgtcaggag	agtggaattt	acccgacaac
tgacagagggt	gtgtaagtat	ctgtctgata	gacaaacgta	tgcaacgtca	taacgaagct
actttagggt	cgtacactac	caaagccagc	aagaaaaact	tttcgttcaa	gcttataaccg
aattactcga	taacctctca	agatgctgaa	aggcgtcctt	gggaagttat	ggtaaatatt
cgtgggtgtg	ctatgtccga	aggttggtgt	ccattatcct	tagagtctgt	ttctgtttgt
attgttcata	aaaacaatgt	tagaaagggt	ctaagagaga	aggtgactgc	cgtgtccgaa
gacgacgcta	tagaactcac	agaagagggt	gttgatgagt	ttatagaagc	cgtaccgatg
gcgcgacggt	tgcaagaactt	gagaaaaccc	aagtacaaca	aagaaaaaga	aaataaaaat
ttgaataata	aaaatagtat	aggagtttcc	aaacctgtcg	gtttggaaag	aaataaagta
aggagtgtag	ttagaaaagg	ggttaggagt	gatagtagtt	taggtgtgac	tgatatgagt
caggacggta	gctcaagcga	gatatcatcc	gattcgttta	ttt	

8/18

FIG. 8

atggctgtta	gtctcagaga	tactgtcaaa	attagcgagt	tcattgatct	ttcgaaacag
gatgagatac	ttccggcatt	catgactaag	gtcaagagcg	tcagaatatc	gactgtggac
aagattatgg	ctgttaagaa	tgatagtctt	tctgatgtag	atttacttaa	aggtgttaag
ttagttaaga	atgggtacgt	gtgcttagct	ggtttggtag	tgtctgggga	gtggaatctc
ccggacaact	gccgtggtgg	tgtcagtgtt	tgtattgtag	ataagagaat	gaaaaggagt
aaggaggcaa	cgctgggtgc	gtatcacgcc	cctgcttgca	aaaagaattt	ttcctttaag
ctaatecccta	attattcaat	aacatccgag	gatgctgaga	agcacccatg	gcaagtatta
gtgaatatca	aaggagtggc	tatggaagaa	ggatactgtc	ctttatcttt	ggagttcggt
tcaattttgtg	tagtacataa	aaataatgta	agaaaagggt	tgagggaacg	tattttgaga
gtaacagacg	gctcgccaat	tgaactcact	gaaaaagttg	ttgaggagtt	catagatgaa
gtaccaatgg	ctgtgaaact	cgaaagggtc	cggaaaacaa	aaaagagagt	ggtaggtaat
agtgttaata	ataagaaaat	aaataatagt	ggtaagaaag	gtttgaaagt	tgaggaaatt
gaggataatg	taagtgatga	cgagtctatc	gcgtcatcga	gtacgtttt	

9/18

FIG. 9

atggctctag	ttgttaaagg	taaggtaa	attaatgagt	ttatcgatct	gtcaaagtct
gagaaacttc	tcccgtcgat	gttcacgcct	gtaaagagtg	ttatgggttc	aaagggtgat
aagattatgg	tccatgaaaa	tgaatcattg	tctgaagtaa	atctcttaaa	aggtgtaaaa
cttatagaag	gtgggtatgt	ttgcttagtt	ggctctgttg	tgtccggtga	gtggaattta
ccagataatt	gccgtggtgg	tgtgagtgtc	tgcattggtg	acaagagaat	ggaaagagcg
gacgaagcca	cactggggtc	atattacact	gctgctgcta	aaaagcggtt	tcagtttaaa
gtggtcccaa	attacggtat	tactacaaag	gatgcagaaa	agaacatatg	gcaagttcat
gtgcgtattc	aaaattttaag	gattgaagct	ggatggcaac	ctctggcctt	agaggtggtt
tctgttgcta	tggtcactaa	taacgtgggt	gttaagggtt	tgagagaaaa	ggcatcgcga
gtgaatgatc	cgaatgtcga	aggtttcgaa	ggcgtggttg	acgatttcgt	cgattcggtc
gcagcattca	aggcgggtga	cactttcaga	aagaaaaaga	aaaggattgg	aggaaaggat
gtaaataata	ataagtttag	atatagaccg	gagagatacg	ccggtcagga	ttcgttaaat
tataaagaag	aaaacgtctt	acaacatcac	gaactcgaat	cagtaccagt	atttcgcagc
gacgtgggca	gagcccacag	cgatgctt			

10/18

FIG. 10

atgtcttacg	agcctaaagt	gagcgacttc	cttgctctta	cgaaaaagga	ggaaatttta
cccaaggctc	ttacgagggt	aaagactgtc	tctattagta	ctaaggatgt	tatatctgtt
aaggattctg	agtcctctgt	tgatatagat	ttactagtta	atgtgccatt	agataagtat
agatatgtgg	gtgttttagg	tgttggtttt	accggtgagt	ggaatttacc	agataattgc
cgtgggtggt	tgagtgtctg	catgggtgac	aagagaatgg	aaagagcgga	cgaagccaca
ctgggggtcat	attacactgc	tgctgcgaaa	gacaaaaggt	tccagttcaa	gctgggtcca
aattacttcg	tgtctgttgc	agatgccaag	cgaaaaccgt	ggcaagttca	tgtgcgtatt
caaaatttaa	ggattgaagc	tggatggcaa	cctctggcct	tagaggtggt	ttctgttgct
atggtcacta	ataacgtggt	tgtaagggt	ttgagagaaa	aggtcatcgc	agtgaatgat
ccgaatgtcg	aaggtttcga	aggcgtgggt	gacgatttcg	tcgattcggt	cgcagcattc
aaggcgggtg	acactttcag	aaagaaaaag	aaaaggattg	gaggaaagga	tgtaaataat
aataagttta	gatatagacc	ggagagatac	gccggtcagg	attcgttaaa	ttataaagaa
gaaaacgtct	tacaacatca	cgaactcgaa	tcagtaccag	tatttcgcag	cgacgtgggc
agagcccaca	gcgatgctt				

11/18

FIG. 11

```
aaataaacga atcggatgat atctcgcttg agctaccgtc ctgactcata tcagtcacac
ctaaactact atcactccta accccttttc taactacact ccttacttta tttctttcca
aaccgacagg tttggaaact cctatactat ttttattatt caaatTTTTa ttttcttttt
ctttgttgta cttgggtttt ctcaagttct gcaaacgtcg cgccatcggt acggcttcta
taaactcatc aacaacctct tctgtgagtt ctatagcgtc gtcttcggac acggcagtca
ccttctctct tagacccttt ctaacattgt ttttatgaac aatacaaaca gaaacgaact
ctaaggataa tggacaccaa ccttcggaca tagccacacc acgaatattt accataactt
cccaaggacg cctttcagca tcttgagagg ttatcgagta attcgggtata agcttgaacg
aaaagttttt cttgctggct ttggtagtgt acgaacctaa agtagcttcg ttatgacgtt
gcatacgttt gtctatcaga cagatactta caccacctct gcagttgtcg ggtaaattcc
actctcctga caccacaaga cctactaaac aaacataacc accttctata agttttacac
cttttaagag atttacttca gacaatgatt cattctcttt ggccattatc ttatccactg
ttgagactct gaccgacttc attcttgatga atgcagaagg taaaacctct tcagacttgg
ataatttaat gaattcatcg atcttgacaa tagcctttga cat
```

12/18

FIG. 12

aatacgaatc	agaatccgcg	accgacgtct	cggcttcac	ttcaatcaaa	ttatcaaact
ctttttcaac	ttcatcaaaa	ctttttgggt	taggccttcc	gcctgaacgc	cccttaccta
aattattatt	atttttcgga	cctctttttg	aggatttggt	tcgaaacttt	gcgagtctaa
cgcacattgg	aacattctcc	atgaactcat	caacaacctc	ttctgtgagt	tctatagcgt
cgtcttcgga	cacggcagtc	accttctctc	ttagaccctt	tctaacattg	tttttatgaa
caatacaaac	agaaacgaac	tctaatagaca	aagggcagta	gcccgcactc	atttttacat
ttttaatat	tactaagacc	tgccatatgt	tcttttctgc	atcctttgta	gtaataaccgt
aatttgggac	cactttaaac	tgaaaccgct	tttttagcagc	agcagtgtaa	tatgacccca
gtgtggcttc	gtccgctctt	tccattctct	tgtcaaccat	gcagacactc	acaccaccac
ggcaattatc	tggtaaatc	cactctcctg	acaccacaag	acctactaaa	caaacataac
cattttttaac	taacttaaca	cccttaagag	atttacttcg	gacaatgatt	cattttcatg
gaccataatc	ttatcaacct	ttgaaaccat	aacactcttt	acaggcgtga	atgcagaagg
taaaacctct	tcagactttg	acagatcgat	aaactcatta	atatttacct	tacctttaac
aactagagcc	at				

13/18

FIG. 13

aatacgaatc	agaatccgcg	atagactcgt	catcacttac	attatcctca	atttcctcaa
ctttcaaacc	tttcttacca	ctattattta	ttttcttatt	attaacacta	ttacctacca
ctctcttttt	tgttttccgg	aacctttcga	gtttcacagc	cattggtact	tcctctatga
actcatcaac	aacttcttct	gaaagttcca	tgggtcctcc	atcgttcaca	ctcgttactt
tctccctcaa	acccaatttt	atattatttt	tataaacaat	acacacagac	acaaattcta
aagataaagg	gcagtatcct	tcttccatag	ccactccttt	gatattcact	aatacttgcc
atgggtgctt	ttctgcatcc	tcggatgtta	ttgaataatt	agggaccact	ttaaactgaa
accgcttttt	agcagcaggg	gcgtgatacg	caccacgcgt	tgcctcctta	ctcctttcca
ttctcttgtc	aaccatgcag	acactcacac	caccacggca	gttgtcgggg	agattccact
caccggacac	aacaagacca	actaagcaaa	catacccacc	ttctataagt	tttacacctt
ttaagagatt	tacttcagac	aatgattcat	tttcatggac	cataatctta	tcaacctttg
aaaccataac	actctttaca	ggcgtgaaca	tcgacgggag	aagtttctca	gactttgaca
gatcgataaa	ctcattaata	tttaccttac	ctttaacaac	tagagccat	

14/18

FIG. 14

```
aatacgaatc agaatccgcg accgacgtct cggcttcact tacattatcc tcaatttcct
caactttcaa aactttctta ccactattat ttattttctt attattaaca ctattaccta
ccactctctt ttttgttttc cggaaccttt cgagtttcac agccattggg acttcatcta
tgaactcatc aacaactttt tcagtgaagt caattggcga gccgtctgtt actctcaaaa
tacgttccct caaacccaat tttatattat ttttataaac aatacacaca gacacaaatt
ctaatacaca agggcagtag cccgcactca tttttacatt tttaatatat actaagacct
gccatgggtg cttctcagca tcctcggatg ttattgaata attagggatt agcttaaagg
aaaaattctt ttgcaagca ggggcgtgat acgcacccag tgtggcttcg tccgctcttt
ccattctctt gtcaaccatg cagacactca caccaccacg gcagttgtcc gggagattcc
actcaccgga cacaacaaga ccaactaagc acacgtacc attcttaact aacttaacac
ctttaagtaa atctacatca gacaatgatt catthttcat gaccataatc ttatcaacct
ttgaaacct aacactcttt acaggcgtga acatcgacgg gagaagtttc tcagactttg
acagatcgat aaactcgcta attttgacag tatctctgag actaacagcc at
```

15/18

FIG. 15

atggctctag	ttgttaaagg	aaaagtgaat	attaatgagt	ttatcgatct	gtcaaagtct
gagaaacttc	tcccgtcgat	gttcacgcct	gtaaagagtg	ttatggtttc	aaaggttgat
aagattatgg	tccatgaaaa	tgaatcattg	tctgaagtaa	atctcttaaa	aggtgtaaaa
cttatagaag	gtgggtatgt	ttgcttagtt	ggtcttggtg	tgtccggtga	gtggaattta
ccagataatt	gccgtggtgg	tgtgagtgtc	tgcatgggtg	acaagagaa	ggaaagagcg
gacgaagcca	ctctcggatc	ttactacaca	gcagctgcaa	agaaaagatt	tcagttcaag
gtcgttccca	attatgctat	aaccacccag	gacgcgatga	aaaacgtctg	gcaagtttta
gttaatatta	gaaatgtgaa	gatgtcagcg	ggtttctgtc	cgctttctct	ggagtttggtg
tctgtgtgta	ttgtttataa	aaataatata	aaattgggtt	tgaggagaa	agtaacgagt
gtgaacgatg	gaggacccat	ggaactttca	gaagaagttg	ttgatgagtt	catggaagat
gtcccaatgt	cggtagact	cgcaaagttt	cgatctcgaa	ccggaaaaaa	gagtgatgtc
cgcaaaggga	aaaatagtag	tagtgatcgg	tcagtgccga	acaagaacta	tagaaatggt
aaggattttg	gaggaatgag	ttttaaaaag	aataatttaa	tcgatgatga	ttcggagacg
tcggtcgcgg	attctgattc	gtatt			

16/18

FIG. 16

atggctctag	ttgttaaagg	aaaagtgaat	atcaatgagt	ttatcgacct	gacaaagtct
gagaaacttc	tcccgtcgat	gtttaccct	gtaaagagt	ttatggttcc	aaagttgata
agattatggg	tcatgagaat	gagtcattgt	caggggtgaa	ccttcttaaa	ggagttaagc
ttattgatag	tggtacgtc	tgtttagccg	gtttgggtcgt	cacgggagag	tggaacttgc
ctgacaattg	ccgtgggtgg	gtgagcgtgt	gtctgggtgga	caagagaatg	gaaagagcgg
acgaagccac	actggggtca	tattacactg	ctgctgctaa	aaagcgggtt	cagttcaagg
tcgttcccaa	ttatgctata	accacccagg	atgcagaaaa	gaacatatgg	caggtcttag
taaatattaa	aaatgtgaag	atgagtgcgg	gctactgccc	tttgctatta	gaatttgtgt
cgggtgtgtat	tgtttataga	aataatataa	aattgggttt	gagagagaaa	gtaacgagtg
tgaacgatgg	agggcccatg	gaacttacag	aagaagtcgt	tgatgagttc	atggaagatg
tccttatgtc	gatcaggctt	gcaaagtttc	gatctcgaat	cctcaaaaag	agtgatgtcc
gcaaagggaa	aaatagtagt	agtgatcggg	cagtgccgaa	caagaactat	agaaatgtta
aggattttgg	aggaatgagt	tttaaaaaga	ataatttaat	cgatgatgat	tcggaggcta
ctgtcgcgga	ttctgattcg	tttt			

17/18

FIG. 17

```
atggctctag ttgttaaagg aaaagtgaat atcaatgagt ttatcgacct gacaaaaatg
gagaagatct taccgtcgat gtttaccctt gtaaagagtg ttatgtgttc caaagttgat
aaaataatgg ttcattgagaa tgagtcattg tcaggggtga accttcttaa aggagttaag
cttattgata gtggatacgt ctgttttagcc ggtttggtcg tcacgggcga gtggaacttg
cctgacaatt gcagaggagg tgtgagcgtg tgtctggtgg acaaaaggat ggaaagagcc
gacgaggcca ctctcggatc ttactacaca gcagctgcaa agaaaagatt tcagttcaag
gtcgttccca attatgctat aaccacccag gacgcgatga aaaacgtctg gcaagtttta
gttaataatta gaaatgtgaa gatgtcagcg ggtttctgtc cgctttctct ggagtttgtg
tcggtgtgta ttgtttatag aaataatata aaattagggt tgagagagaa gattacaaac
gtgagagacg gagggcccat ggaacttaca gaagaagtcg ttgatgagtt catggaagat
gtccctatgt cgatcaggct tgcaaagttt cgatctcgaa ccggaaaaaa gagtgatgtc
cgcaaaggga aaaatagtag tagtgatcgg tcagtgccga acaagaacta tagaaatgtt
aaggattttg gaggaatgag ttttaaaaag aataatttaa tcgatgatga ttcggaggct
actgtcgccg aatcggattc gtttttaa
```

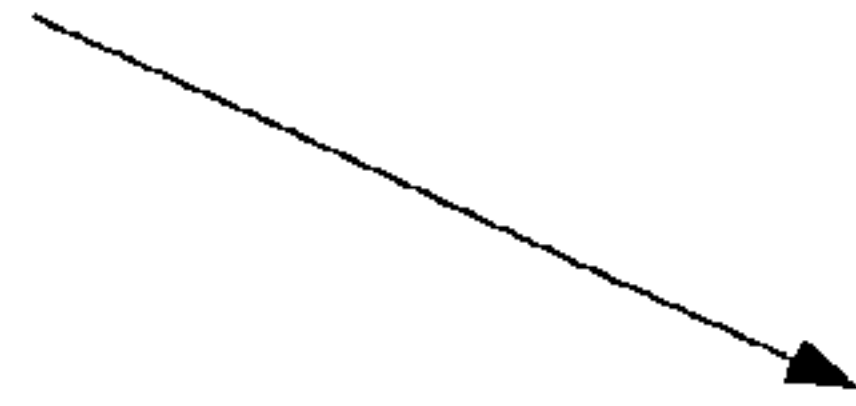
18/18

FIG. 18

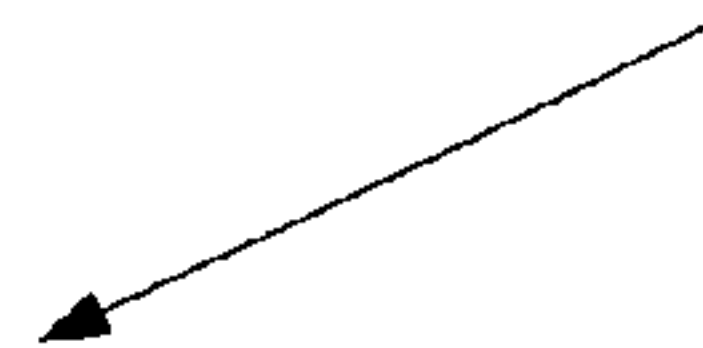
```
atggctctag ttgttaaagg taaggtaaatt attaatgagt ttatcgatct gtcaaagtct
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa
cttatagaag gtgggtatgt ttgcttagtt ggtcttggtg tgtccggtga gtggaattta
ccagataatt gccgtggtgg tgtgagtgtc tgcattggtg acaagagaat ggaaagagcg
gacgaagcca cactgggggc atattacact gctgctgcta aaaagcgggt tcagtttaaa
gtggtcccaa attacggtat tactacaaag gatgcagaaa agaacatatg gcaggtctta
gtaaatatta aaaatgtaaa aatgagtgcg ggctactgcc ctttgtcatt agaatttgtg
tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt
gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat
gttccaatgt cggtttagact cgcaaagttt cgaaccaaatt cctcaaaaag aggtccgaaa
aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat
gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat
tctgattcgt att
```

A 5' - A G A T C G A T C A A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T A G C T A G T T A A C - 5'

B 5' - A G A C C G A T C G A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T G G C T A G C T A A C - 5'

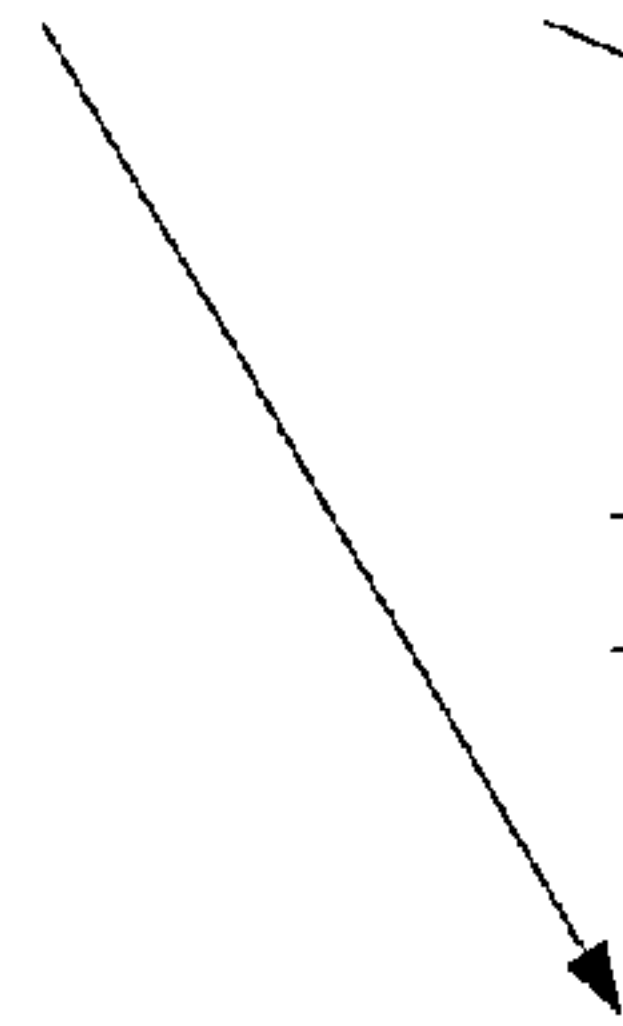
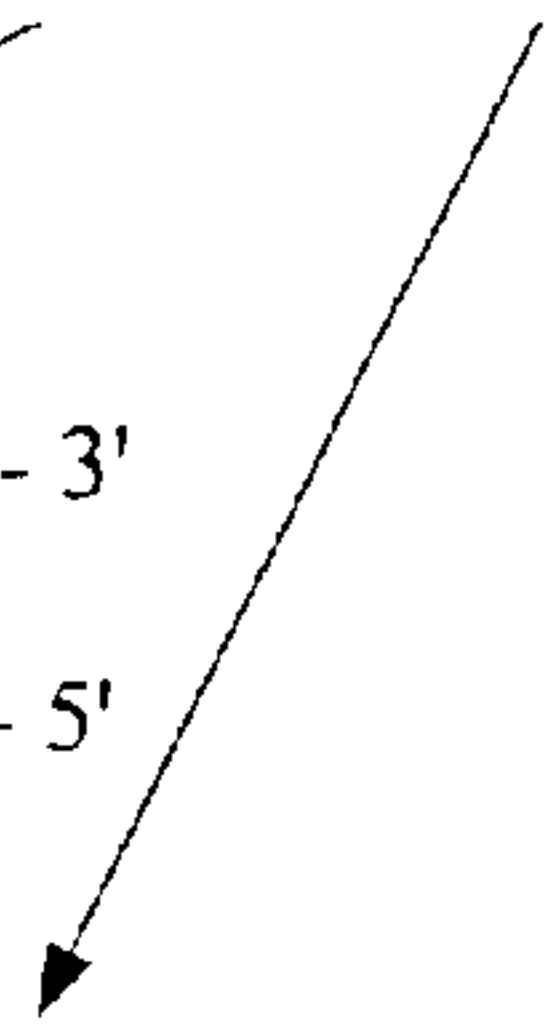


5' - A G A T C G A T C A A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T G G C T A G C T A A C - 5'



A 5' - A G A T C G A T C A A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T A G C T A G T T A A C - 5'

B 5' - A G A C C G A T C G A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T G G C T A G C T A A C - 5'



X 5' - A G A T C G A T C G A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T A G C T A G C T A A C - 5'

5' - A G A C C G A T C A A T T G - 3'
| | | | | | | | | | | | | | | |
3' - T C T G G C T A G T T A A C - 5'

Y