

(12) PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 200048247 B2**
(10) Patent No. **781020**

(54) Title
Viral expression vectors

(51) ⁶ International Patent Classification(s)
C12N 015/40 C12N 015/82
C07K 014/08 C12N 015/86
C12N 005/14

(21) Application No: 200048247

(22) Application Date: 2000.05.04

(87) WIPO No: W000/66743

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/132697	1999.05.04	US

(43) Publication Date : 2000.11.17

(43) Publication Journal Date : 2001.01.25

(44) Accepted Journal Date : 2005.04.28

(71) Applicant(s)
Large Scale Biology Corporation

(72) Inventor(s)
Wayne P. Fitzmaurice; Gregory P. Pogue; John A. Lindbo

(74) Agent/Attorney
Fisher Adams Kelly,GPO Box 1413,BRISBANE QLD 4001

(56) Related Art
WO 1995/021248

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/40, C07K 14/08, C12N 15/86, 15/82, 5/14		A2	(11) International Publication Number: WO 00/66743
(21) International Application Number: PCT/US00/12380		(43) International Publication Date: 9 November 2000 (09.11.00)	
(22) International Filing Date: 4 May 2000 (04.05.00)		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, 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, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(30) Priority Data: 60/132,697 4 May 1999 (04.05.99) US			
(71) Applicant (for all designated States except US): LARGE SCALE BIOLOGY CORPORATION [US/US]; 3333 Vaca Valley Parkway, Vacaville, CA 95688 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): FITZMAURICE, Wayne, P. [US/US]; 1218 Las Encinas Court, Vacaville, CA 95688 (US). POGUE, Gregory, P. [US/US]; 419 Trillick Court, Vacaville, CA 95688 (US). LINDBO, John, A. [US/US]; 143 Sundance Drive, Vacaville, CA 95688 (US).			
(74) Agent: HALLUIN, Albert, P.; Howrey Simon Arnold & White, LLP, 1299 Pennsylvania Avenue, N.W., Washington, DC 20004 (US).			
(54) Title: VIRAL EXPRESSION VECTORS			
(57) Abstract			
The present invention provides nucleic acid sequences having an altered viral movement protein and 126/183kDa replicase proteins further characterized in its ability to stabilize a transgene contained in a virus that expresses the altered movement protein. The present invention also provides viral vectors expressing the altered movement protein, cells transformed with the vectors, and host plants infected by the viral vectors.			

VIRAL EXPRESSION VECTORS

CROSS-REFERENCE TO RELATED APPLICATION

5

This application claims the priority benefit of provisional U.S. Patent Application Serial No. 60/132,697, filed May 4, 1999, pending, which is hereby incorporated herein by reference in its entirety.

10

TECHNICAL FIELD

This invention is in the field of plant virology. Specifically, the invention relates to the synthesis of nucleic acid sequences encoding an altered viral movement protein, construction of viral vectors expressing such protein, and generation of host plants infected by the viral vectors. The viral vectors permit rapid local and systemic invasion of a host, and allow stable expression of a transgene of interest.

15

BACKGROUND OF THE INVENTION

20

In the last fifteen years, considerable progress has been made in expressing foreign genes in plants. Foreign proteins are now routinely produced in many plant species for modification of the plant or for production of proteins for use after extraction. Vectors for the genetic manipulation of plants have been derived from several naturally occurring plant viruses. For the production of specific proteins, transient expression of foreign genes in plants using virus-based vectors has several advantages. Products of plant viruses are among the highest produced proteins in plants. Often a viral gene product is the major protein produced in plant cells during virus replication. Many viruses are able to systemically move from an initial infection site to almost all cells of the plant. Because of these reasons, plant viruses have been developed into efficient transient expression vectors for foreign genes in plants. Viruses of multi-cellular plants are relatively small, probably due to the size

25

30

limitation in the pathways that allow viruses to move to adjacent cells in the systemic infection of entire plants. One such plant virus upon which plant expression vectors are based is TMV (tobacco mosaic virus). TMV is the type member of the tobamovirus group. TMV has straight tubular virions of approximately 300 x 18 nm with a 4 nm-diameter hollow canal consisting of approximately 2000 units of a single capsid protein wound helically around a single RNA molecule. Virion particles are 95% protein and 5% RNA by weight. The genome of TMV is composed of a single-stranded RNA of 6395 nucleotides containing five large ORFs. Expression of each gene is regulated independently. The virion RNA serves as the messenger RNA (mRNA) for the 5' genes, encoding the 126 kDa replicase subunit and the overlapping 183 kDa replicase subunit that is produced by read through of an amber stop codon approximately 5% of the time. Expression of the internal genes is controlled by different promoters on the minus-sense RNA that direct synthesis of 3'-coterminial subgenomic mRNAs which are produced during replication. A detailed description of tobamovirus gene expression and life cycle can be found, among other places, in Dawson and Lehto, Advances in Virus Research 38:307-342 (1991).

Thus, it is of scientific and commercial interest to provide new and improved vectors for the genetic manipulation of plants.

SUMMARY OF THE INVENTION

A principal aspect of the present invention is the design of a recombinant viral vector expressing an altered movement protein and altered 126/183 viral proteins to affect stable expression of a transgene in a plant host.

Accordingly, the present invention provides an isolated nucleic acid sequence encoding an altered viral movement protein having the amino acid sequence shown in SEQ ID NOS.: 5 and 6 and altered 126/183 viral proteins. In one aspect, the isolated nucleic acid sequence is essentially identical to the sequence shown in SEQ ID NOS.: 3 and 4, and it contains a Thymine (T) or Uracil (U) residue at position 5212 and Guanine (G) residue at 5303 as shown in Figure 1A. In another aspect, the isolated

nucleic acid sequence is identical to the sequence shown in SEQ ID NOS.: 3 and 4. The alteration of the 30K movement protein and alteration of the 126/183 viral proteins results in an enhanced ability to facilitate stabilization of a transgene contained in a viral vector.

5 In a separate embodiment, the present invention provides a viral vector comprising the nucleic acid sequence encoding an altered viral movement protein having the amino acid sequence shown in SEQ ID NOS.: 5 and 6 and altered 126/183 viral proteins. In one aspect, the viral vector exhibits an enhanced ability compared to a control viral vector to stabilize a transgene contained in the vector. Preferably,
10 the vector is a tobacco mosaic viral vector. A particularly preferred vector is designated BSG1057 (deposited with American Type Culture Collection having accession number 20398, which was deposited on April 28, 1999).

In a separate aspect within this embodiment, the viral vector comprises a transgene of interest. Preferably the transgene is a non-viral gene encoding a protein
15 selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, and a chaperon protein.

The present invention also provides a cell transformed with a subject viral vector. The transformed cell may be animal or plant. Preferably, the cell is a plant cell. The present invention further provides a transgenic plant comprising the viral
20 vector. Preferred transgenic plant may, for example, be *Nicotiana benthamiana* or *Nicotiana tabacum*, but others may be just as readily substituted by one of skill in the art.

BRIEF DESCRIPTION OF THE DRAWINGS

25

Figure 1 depicts a comparison of the nucleotide sequences encoding an altered movement protein contained in the vector BSG1057 (SEQ ID NO.: 4) and the wildtype movement protein contained in the vector BSG1037 (SEQ ID NO.: 3). Sequence identities are indicated by *, and mismatches are indicated by -.

Figure 2 depicts a comparison of the amino acid sequences encoding an altered movement protein contained in the vector BSG1057 (SEQ ID NO.: 6) and the wildtype movement protein contained in the vector BSG1037 (SEQ ID NO.: 5). Sequence identities are indicated by *, and mismatches are indicated by -.

5 Figure 3 is a schematic representation of the restriction sites of the vector BSG1037.

Figure 4 is a schematic representation of the restriction sites of the vector BSG1057.

Figure 5 is the complete sequence of BSG1037 (SEQ ID NO.: 1).

10 Figure 6 is the complete sequence of BSG1057 (SEQ ID NO.: 2).

Figure 7 is a schematic map of locations of mutations in BSG1057.

Figure 8 shows *N. benthamiana* plants at 20 days postinoculation. There are four columns of five plants. The first column on the left shows plants inoculated with first passage BSG1037. Column 2 is seventh passage BSG1037, Column 3 is first passage BSG1057, Column 4 is seventh passage BSG1057.

15

MODES FOR CARRYING OUT THE INVENTION

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains. For example, the general teaching for constructing viral plant vectors and using them to systemically infect plants and express

20

25 heterologous proteins therefrom is disclosed in U.S. Patent Nos. 5,316,931; 5,977,438; 5,889,191; 5,889,190; 5,866,785 and 5,816,653, the entire disclosures of which are hereby incorporated herein by reference.

General Techniques:

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See, *e.g.*, Matthews, PLANT VIROLOGY, 3rd edition (1991); Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

Definitions:

A "plant cell" refers to the structural and physiological unit of plants, consisting of a protoplast and the cell wall.

A "protoplast" is an isolated cell without cell walls, having the potency for regeneration into cell culture, tissue or whole plant.

A "host" encompasses cell, tissue or organism capable of replicating a vector or viral nucleic acid and which is capable of being infected by a virus containing the viral vector or viral nucleic acid. This term is intended to include prokaryotic and eukaryotic cells, organs, tissues, organisms, or *in vitro* extracts thereof, where appropriate. A preferred host cell is a plant cell.

The term "infection" refers to the process of transferring or the ability of a virus to transfer its nucleic acid to a host, wherein the viral nucleic acid is replicated, viral proteins are synthesized, or new viral particles assembled.

The "movement protein" of tobacco mosaic virus is a noncapsid protein required for cell-to-cell movement of the RNA replicons or viruses in plants.

5 The terms "nucleic acid sequence", "polynucleotide", "nucleotides" and "oligonucleotides" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, 10 ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of 15 the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated.

20 As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectively referred to as gene product. If the polynucleotide is derived from 25 genomic DNA, expression may include splicing of the mRNA in an eukaryotic cell.

In the context of polynucleotides, a "linear sequence" or a "sequence" is an order of nucleotides in a polynucleotide in a 5' to 3' direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polynucleotide.

A linear sequence of nucleotides is "essentially identical" to another linear sequence, if both sequences are capable of hybridizing to form a duplex with the same complementary polynucleotide. Sequences that hybridize under conditions of greater stringency are more preferred. It is understood that hybridization reactions can
5 accommodate insertions, deletions, and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. In general, essentially identical sequences of about 40 nucleotides in length will hybridize at about 30°C in 10 x SSC (0.15 M NaCl, 15 mM citrate buffer); preferably, they will hybridize at about 40°C in 6 x SSC; more preferably, they will hybridize at about 50°C in 6 x SSC; even
10 more preferably, they will hybridize at about 60°C in 6 x SSC, or at about 40°C in 0.5 x SSC, or at about 30°C in 6 x SSC containing 50% formamide; still more preferably, they will hybridize at 40°C or higher in 2 x SSC or lower in the presence of 50% or more formamide. It is understood that the rigor of the test is partly a function of the
15 length of the polynucleotide; hence shorter polynucleotides with the same homology should be tested under lower stringency and longer polynucleotides should be tested under higher stringency, adjusting the conditions accordingly. The relationship between hybridization stringency, degree of sequence identity, and polynucleotide length is known in the art and can be calculated by standard formulae; see, e.g.,
20 Meinkoth et al. Sequences that correspond or align more closely to the invention disclosed herein are comparably more preferred. Generally, essentially identical sequences are at least about 50% identical with each other, after alignment of the homologous regions. Preferably, the sequences are at least about 60% identical; more preferably, they are at least about 70% identical; more preferably, they are at least
25 about 80% identical; more preferably, the sequences are at least about 90% identical; even more preferably, they are at least 95% identical; still more preferably, the sequences are 100% identical.

In determining whether polynucleotide sequences are essentially identical, a sequence that preserves the functionality of the polynucleotide with which it is being

compared is particularly preferred. Functionality may be established by different criteria, such as ability to hybridize with a target polynucleotide, and whether the polynucleotide encodes an identical or essentially identical polypeptides. Thus, nucleotide substitutions which cause a non-conservative substitution in the encoded polypeptide are preferred over nucleotide substitutions that create a stop codon; nucleotide substitutions that cause a conservative substitution in the encoded polypeptide are more preferred, and identical nucleotide sequences are even more preferred. Insertions or deletions in the polynucleotide that result in insertions or deletions in the polypeptide are preferred over those that result in the down-stream coding region being rendered out of phase.

The term "hybridize" as applied to a polynucleotide refers to the ability of the polynucleotide to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues in a hybridization reaction. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogsteen binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these.

The terms "cytosolic", "nuclear" and "secreted" as applied to cellular proteins specify the extracellular and/or subcellular location in which the cellular protein is mostly localized. Certain proteins are "chaperons", capable of translocating back and forth between the cytosol and the nucleus of a cell.

A "control" is an alternative subject or sample used in an experiment for comparison purpose. For example, where the purpose of the experiment is to ascertain whether a viral vector carrying an altered movement protein possesses an enhanced ability in systemic invasion of a host plant, it is generally preferable to use a control viral vector (e.g. BSG1037 shown in Figures 1-2) expressing the wildtype altered movement protein (e.g. 1037 sequence shown in Figure 2).

A "cell line" or "cell culture" denotes bacterial, plant, insect or higher eukaryotic cells grown or maintained in vitro. The descendants of a cell may not be

completely identical (either morphologically, genotypically, or phenotypically) to the parent cell.

5 A "vector" refers to a recombinant plasmid or virus that comprises a polynucleotide to be delivered into a host cell, either *in vitro* or *in vivo*. The polynucleotide to be delivered may comprise a coding sequence of interest in gene therapy. nucleic acid molecule, preferably self-replicating, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of a DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression
10 vectors that function for transcription and/or translation of the DNA or RNA. Also included are vectors that provide more than one of the above functions.

A "replicon" refers to a polynucleotide comprising an origin of replication (generally referred to as an ori sequence) which allows for replication of the polynucleotide in an appropriate host cell. Examples of replicons include episomes
15 (such as plasmids), as well as chromosomes (such as the nuclear or mitochondrial chromosomes).

A "transcription unit" is a DNA segment capable of directing transcription of a gene or fragment thereof. Typically, a transcription unit comprises a promoter operably linked to a gene or a DNA fragment that is to be transcribed, and optionally
20 regulatory sequences located either upstream or downstream of the initiation site or the termination site of the transcribed gene or fragment.

Nucleic acids of the present invention

25 The present invention encompasses a recombinant viral vector expressing an altered movement protein and altered 126/183 viral proteins to effect stable expression of a transgene in a plant host. Distinguished from the previously described movement protein, the altered protein contains two amino acid substitutions (replacing the threonine residue at position 104 with isoleucine, and replacing the lysine residue at position 134 with arginine, see Figure 2). The altered

viral vector exhibits an enhanced ability to facilitate stabilization of a transgene contained in a virus that expresses the altered movement protein.

In one embodiment, the present invention provides an isolated nucleic acid sequence encoding an altered viral movement protein having the amino acid sequence shown in SEQ ID NOS.: 5 and 6 and altered 126/183 viral proteins. In one aspect within this embodiment, the isolated nucleic acid sequence of the movement protein is essentially identical to the sequence shown in SEQ ID NO. 3, and it contains a Thymine (T) or Uracil (U) residue at position 5212 and Guanine (G) residue at 5303 as shown in Figure 1A. As used herein, a linear sequence of nucleotides is "essentially identical" to another linear sequence, if both sequences are capable of hybridizing to form a duplex with the same complementary polynucleotide.

Hybridization can be performed under conditions of different "stringency". Relevant conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and the washing procedure. Higher stringency conditions are those conditions, such as higher temperature and lower sodium ion concentration, which require higher minimum complementarity between hybridizing elements for a stable hybridization complex to form. In general, a low stringency hybridization reaction is carried out at about 40 °C in 6 x SSC or a solution of equivalent ionic strength/temperature. A moderate stringency hybridization is typically performed at about 50 °C in 6 x SSC, and a high stringency hybridization reaction is generally performed at about 60 °C in 6 x SSC. The essentially identical nucleic acid sequences embodiment in the invention encompass all sequences encoding modified movement proteins containing conservative or non-conservative substitutions that do not significantly affect the claimed structural characteristics (i.e. retain the substitution of isoleucine for threonine₁₀₄, and arginine for lysine₁₃₄). Modification of polypeptides by altering their corresponding nucleic acid sequences is routine practice in the art. Amino acid residues which can be conservatively substituted for one another include but are not limited to: glycine/alanine; valine/isoleucine/leucine; asparagine/glutamine; aspartic acid/glutamic acid; serine/threonine; lysine/arginine; and phenylalanine/tyrosine.

Changes in nucleic acid sequence that do not change the encoded amino acid sequence are generally preferred.

The recombinant viral vector embodiment of this invention comprises a nucleic acid sequence encoding the above-described viral movement protein.

5 In one embodiment, a nucleic acid is introduced into a plant host. Preferably, the nucleic acid may be introduced by way of a viral nucleic acid, using techniques known in the art, and preferably the techniques disclosed in U.S. Patent Nos. 5,316,931; 5,977,438; 5,889,191; 5,889,190; 5,866,785 and 5,816,653, the entire disclosures of which are hereby incorporated herein by reference. Such recombinant
10 viral nucleic acids are stable for the maintenance and transcription of such nonnative sequences in the plant host.

BSG1057 (SEQ ID NO.:2) is a mutant version of BSG1037 (SEQ ID NO.:1). The complete sequences of BSG1057 and BSG1037 are shown in FIG. 5 and FIG. 6. BSG1037 has improved insert retention properties.

15 The difference between these two virus vectors is best demonstrated with the Green Fluorescent Protein (GFP) reporter gene. Both BSG1037 and BSG1057 express GFP which can be visualized under long wave UV light by its green fluorescence. The presence of GFP activity identifies those cells in which the recombinant virus is expressing genes.

20 Nicotiana benthamiana plants inoculated with BSG1037 and BSG1057 were observed under long wave UV light at approximately 4 to 5 days post inoculation. The GFP spots on the leaves of plants inoculated with the BSG1057 virus were noticeably larger than the GFP spots on the leaves of plants inoculated with the BSG1037 virus, indicating the 1057 virus moves cell to cell faster than BSG1037.

25

Sequence comparison between BSG1037 and BSG1057

30 The specific nucleotide changes between 1037 and 1057 are listed in the table below. In those cases where the nucleotide change resulted in an amino acid change, that change is noted (using the single letter code).

	nt position	1037 nt	1057 nt	1037AA	1057 AA
	1138	A	G	E	G
	1268	T	C	No AA changes	
	2382	A	G	K	E
5	3632	G	A	No AA changes	
	5213	C	T	T	I
	5303	A	G	K	R
	5896	C	A	No AA changes	

10

126/183 refers to the 126/183 viral proteins.
MP refers to the movement protein.

15

The transgene transcribed by the vector of present invention can be any gene expressed in a biological entity. The selection of transgene is determined largely by the intended purpose of the vector. Preferably the transgene is a non-viral gene selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, and a chaperon protein.

20

The vectors embodied in this invention can be obtained using recombinant cloning methods and/or by chemical synthesis. A vast number of recombinant cloning techniques such as PCR, restriction endonuclease digestion and ligation are well known in the art, and need not be described in detail herein. One skilled in the art can also use the sequence data provided herein or that in the public or proprietary databases to obtain a desired vector by any synthetic means available in the art.

25

Host cell and transgenic organisms of the present invention:

30

The invention provides host cells transformed with the viral vectors described above. The host cells may be animal or plant, but plant hosts are preferred. The viral vectors containing a transgene of interest can be introduced into a suitable eukaryotic cell by any of a number of appropriate means, including electroporation, transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; and infection (where the vector is coupled to an infectious agent). The choice of introducing vectors will often depend on features of the host cell.

For plant cells, a variety of techniques derived from these general methods is available in the art. See, for example, U.S. Patent Nos. 5,316,931; 5,977,438; 5,889,191; 5,889,190; 5,866,785 and 5,816,653. The host cells may be in the form of whole plants, isolated cells or protoplasts. Illustrative procedures for introducing vectors into plant cells include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs and injection into immature embryos, leaf abrasion, abrasion in solution, high velocity water spray, and other injury of a host as well as imbibing host seeds with water containing the recombinant viral RNA or recombinant plant virus. As is evident to one skilled in the art, each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular plant species may not necessarily be the most effective for another plant species.

Agrobacterium tumefaciens-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated expression vectors to introduce DNA into plant cells is well known in the art. This technique makes use of a common feature of *Agrobacterium* which colonizes plants by transferring a portion of their DNA (the T-DNA) into a host cell, where it becomes integrated into nuclear DNA. The T-DNA is defined by border sequences which are 25 base pairs long, and any DNA between these border sequences is transferred to the plant cells as well. The insertion of a recombinant plant viral nucleic acid between the T-DNA border sequences results in transfer of the recombinant plant viral nucleic acid to the plant cells, where the recombinant plant viral nucleic acid is replicated, and then spreads systemically through the plant. Agro-infection has been accomplished with potato spindle tuber viroid (PSTV) (Gardner *et al.*, *Plant Mol. Biol.* 6:221 (1986); CaV (Grimsley *et al.*, and Lazarowitz, S., *Nucl. Acids Res.* 16:229 (1988)) digitaria streak virus (Donson *et al.*, *Virology* 162:248 (1988)), wheat dwarf virus, and tomato golden mosaic virus (TGMV). Therefore, agro-infection of a susceptible plant could be accomplished with a virion containing a recombinant plant viral nucleic acid based on the

nucleotide sequence of any of the above viruses. Particle bombardment or electroporation or any other methods known in the art may also be used.

Because not all plants are natural hosts for *Agrobacterium*, alternative methods such as transformation of protoplasts may be employed to introduce the subject vectors into the host cells. For certain monocots, transformation of the plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments. See, for example, Potrykus et al., *Mol. Gen. Genet.*, **199**:169-177 (1985); Fromm et al., *Nature*, **319**:791 (1986); and Callis et al., *Genes and Development*, **1**:1183 (1987). Applicability of these techniques to different plant species may depend upon the feasibility to regenerate that particular plant species from protoplasts. A variety of methods for the regeneration of cereals from protoplasts are known in the art.

In addition to protoplast transformation, particle bombardment is an alternative and convenient technique for delivering the invention vectors into a plant host cell. Specifically, the plant cells may be bombarded with microparticles coated with a plurality of the subject vectors. Bombardment with DNA-coated microprojectiles has been successfully used to produce stable transformants in both plants and animals (see, for example, Sanford et al. (1993) *Methods in Enzymology*, **217**:483-509). Microparticles suitable for introducing vectors into a plant cell are typically made of metal, preferably tungsten or gold. These microparticles are available for example, from BioRad (e.g., Bio-Rad's PDS-1000/He). Those skilled in the art will know that the particle bombardment protocol can be optimized for any plant by varying parameters such as He pressure, quantity of coated particles, distance between the macrocarrier and the stopping screen and flying distance from the stopping screen to the target.

Vectors can also be introduced into plants by direct DNA transfer into pollen as described by Zhou et al., *Methods in Enzymology*, **101**:433 (1983). Other techniques for introducing nucleic acids into a plant cell include:

- (a) Hand Inoculations. Hand inoculations are performed using a neutral pH, low molarity phosphate buffer, with the addition of celite or carborundum (usually

about 1%). One to four drops of the preparation is put onto the upper surface of a leaf and gently rubbed.

- (b) Mechanized Inoculations of Plant Beds. Plant bed inoculations are performed by spraying (gas-propelled) the vector solution into a tractor-driven mower while cutting the leaves. Alternatively, the plant bed is mowed and the vector solution sprayed immediately onto the cut leaves.
- (c) High Pressure Spray of Single Leaves. Single plant inoculations can also be performed by spraying the leaves with a narrow, directed spray (50 psi, 6-12 inches from the leaf) containing approximately 1% carborundum in the buffered vector solution.
- (d) Vacuum Infiltration. Inoculations may be accomplished by subjecting a host organism to a substantially vacuum pressure environment in order to facilitate infection.

Once introduced into a suitable host cell, expression of the transgene can be determined using any assay known in the art. For example, the presence of transcribed sense or anti-sense strands of the transgene can be detected and/or quantified by conventional hybridization assays (e.g. Northern blot analysis), amplification procedures (e.g. RT-PCR), SAGE (U.S. Patent No. 5,695,937), and array-based technologies (see e.g. U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934).

Expression of the transgene can also be determined by examining the protein product. A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and PAGE-SDS.

In general, determining the protein level involves (a) providing a biological sample containing polypeptides; and (b) measuring the amount of any immunospecific binding that occurs between an antibody reactive to the transgene

product and a component in the sample, in which the amount of immunospecific binding indicates the level of expressed proteins. Antibodies that specifically recognize and bind to the protein products of the transgene are required for immunoassays. These may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra*. and Sambrook et al. (1989) *supra*. The sample of test proteins can be prepared by homogenizing the eukaryotic transformants (e.g. plant cells) or their progenies made therefrom, and optionally solubilizing the test protein using detergents, preferably non-reducing detergents such as triton and digitonin. The binding reaction in which the test proteins are allowed to interact with the detecting antibodies may be performed in solution, or on a solid tissue sample, for example, using tissue sections or solid support that has been immobilized with the test proteins. The formation of the complex can be detected by a number of techniques known in the art. For example, the antibodies may be supplied with a label and unreacted antibodies may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. Results obtained using any such assay on a sample from a plant transformant or a progeny thereof is compared with those from a non-transformed source as a control.

The eukaryotic host cells of this invention are grown under favorable conditions to effect transcription of the transgene. The host cells may also be employed to generate transgenic organisms such as transgenic plants comprising the recombinant DNA vectors of the present invention. Preferred host cells are those having the propensity to regenerate into tissue or a whole organisms. Examples of these preferred host cells include certain plant cells exemplified herein.

Accordingly, this invention provides transgenic plants carrying the subject vectors. The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, *Methods for Plant Molecular Biology*, A. Weissbach and H. Weissbach, eds., Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of

transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil.

The regeneration of plants containing the subject vector introduced by *Agrobacterium tumefaciens* from leaf explants can be achieved as described by Fraley et al., *Proc. Natl. Acad. Sci. U.S.A.*, **80**:4803 (1983). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed. This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transformant shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil to allow the production of roots. These procedures will vary depending upon the particular plant species employed, as is apparent to one of ordinary skill in the art.

A population of progeny can be produced from the first and second transformants of a plant species by methods well known in the art including cross fertilization and asexual reproduction. Transgenic plants embodied in the present invention are useful for production of desired proteins, and as test systems for analysis of the biological functions of a gene.

EXAMPLES

Tobacco plants inoculated with the viruses BSG1037 or BSG1057 express the reporter gene (GFP) in cells that are infected with either virus. The reporter gene activity (indicative of the presence of virus) is easily observed by illuminating plants with long wave UV light. Viruses that lose expression of the inserted *gfp* gene no longer accumulate the GFP protein and do not exhibit GFP fluorescence under UV illumination.

In order to assess the stability of expression of a foreign gene in the new vector, the *gfp* gene was introduced into the standard vector (giving rise to BSG1037) and the improved vector (giving rise to BSG1057). RNA transcripts of these constructs was generated and used to inoculate *Nicotiana benthamiana* plants. At about 7 days postinoculation, extensive systemic GFP expression was observed. GFP-expressing tissue was harvested, ground in phosphate buffer, the cellular debris removed by low-speed centrifugation, and the resulting "green juice" supernatant solution used to inoculate a new set of *N. benthamiana* plants. Systemic tissue was again harvested at about 7 days and the resulting green juice used to serial passage the virus. The procedure was used to serial passage the viruses a total of 7 times. A comparison was then initiated in which *N. benthamiana* plants were inoculated in parallel with the first passage green juice and the seventh passage green juice for BSG1037 and BSG1057. The first passage virus gave excellent systemic expression of GFP beginning about 4 days post inoculation. The BSG1037 seventh passage virus gave little systemic GFP expression and strong visual TMV mosaic symptoms characteristic of a vector that has lost most or all of the inserted sequence. In contrast, the BSG1057 seventh passage virus still gave excellent systemic GFP expression and the mild visual viral symptoms characteristic of a vector retaining its inserted gene.

At 20 days post inoculation, the plants were cut 2 inches above the soil line and allowed to regrow. Plants were monitored as to the accumulation of GFP protein in the new systemic tissue up to 3 weeks post cutting. Plants containing the BSG 1037 virus (both first and seventh passage) showed very little GFP in regrowth tissues, while showing extensive virus symptoms. This result indicates that the virus population was dominated by viruses that have recombinationally lost the genetic insertion. Plants containing BSG1057 (both first and seventh passage) showed good

systemic invasion in re-growth tissue. This indicates superior genetic stability of the foreign gene insertion, *gfp* gene, in BSG1057 compared with BSG1037.

This increased genetic stability of foreign genes was also seen using two additional gene insertions: interferon gamma from chickens and human alpha galactosidase A. Serial passage experiments of either gene in BSG1037 virus preparations showed variable production of product in plants, while comparable experiments in BSG1057 showed more uniform product accumulation in plants. These experiments indicate that the BSG1057 retains foreign gene insertions through multiple passages to a greater extent.

The invention and the manner and process of making and using it are now described in such full, clear, concise and exact terms as to enable any person skilled in the art to which it pertains, to make and use the same. It is to be understood that the foregoing describes preferred embodiments of the present invention and that modifications may be made therein without departing from the spirit or scope of the present invention as set forth in the claims. To particularly point out and distinctly claim the subject matter regarded as invention, the following claims conclude this specification.

EDITORIAL NOTE

APPLICATION NUMBER 48247/00

The following Sequence Listing pages 1/8 to 8/8 are part of the description. The claims pages follow on pages "20" to "21".

SEQUENCE LISTING

<110> LARGE SCALE BIOLOGY CORPORATION

<120> VIRAL EXPRESSION VECTORS

<130> 008010165USPC01

<150> 60/132,697

<151> 1999-05-04

<160> 6

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 7685

<212> DNA

<213> Nicotiana tabacum

<400> 1

gtatttttac	aacaattlacc	aacaacaaca	aacaacagac	aacattacaa	ttactattta	60
caattacaat	ggcatacaca	cagacagcta	ccacatcagc	tttgctggac	actgtccgag	120
gaaacaactc	cttgggtcaat	gatctagcaa	agcgtcgtct	ttacgacaca	gcggttgaag	180
agttttaacgc	tcgtgaccgc	aggcccaagg	tgaacttttc	aaaagtaata	agcgaggagc	240
agacgcttat	tgctacccgc	gcgtatccag	aattccaaat	tacattttat	aacacgcaaa	300
atgccgtgca	tlcgcttgca	ggtggattgc	gatctttaga	actggaatat	ctgatgatgc	360
aaattcccta	cggatcattg	acttatgaca	taggcgggaa	ttttgcatcg	catctgttca	420
agggacgagc	atatgtacac	tgctgcatgc	ccaacctgga	cgttcgagac	atcatgcggc	480
acgaaggcca	gaaagacagt	attgaactat	acctttctag	gctagagaga	ggggggaaaa	540
cagtcacca	cttccaaaag	gaagcatttg	acagatacgc	agaaattcct	gaagacgctg	600
tctgtccaaa	tactttccag	acatgcgaac	atcagccgat	gcagcaatca	ggcagagtgt	660
atgccattgc	gctacacagc	atatatgaca	taccagccga	tgagtccggg	gcggcactct	720
tgaggaaaaa	tgccatagc	tgctatgccg	ctttccactt	ctccgagaac	ctgcttcttg	780
aagattcatg	cgtcaatttg	gacgaaatca	acgcgtgttt	ttcgcgcgat	ggagacaaagt	840
tgaccttttc	ttttgcatca	gagagtactc	ttaattactg	tcatagttat	tctaataltc	900
tttaagtatgt	gtgcaaaact	tacttcccgc	cctcttaata	agagggtttac	atgaaggagt	960
tttttagtcac	cagagttaat	acctggtttt	gtaagttttc	tagaatagat	acttttcttt	1020
tgtacaaaagg	tgtggcccat	aaaagtgtag	atagttagca	gtttttact	gcaatggaag	1080
acgcattggca	ttacaaaaag	actcttgcaa	tgtgcaacag	cgagagaatc	ctccttgagg	1140
attcatcate	agtcatttac	tggtttccca	aaatgagggg	tatggtcatc	gtaccattat	1200
tcgacatttc	tttgagagact	agtaagagga	cgcgcaagga	agtcttagtg	tccaaggatt	1260
tcgtgtttac	agtgcttaac	cacattcgaa	cataccaggc	gaaagctctt	acatacgcaa	1320
atgttttgtc	cttcgtcgaa	tcgattcgat	cgagggtaat	cattaacggg	gtgacagcga	1380
ggtcggaatg	ggaatgtggac	aaatctttgt	tacaatcctt	gtccatgacg	ttttacctgc	1440
atactaagct	tgccgttcta	aaggatgact	tactgattag	caagttagt	ctcggttcga	1500
aaacggtgtg	ccagcatgtg	tgggatgaga	tttcgctggc	gtttgggaac	gcatttccct	1560
ccgtgaaaaga	gaggtctctg	aacaggaaac	ttatcagagt	ggcaggcgac	gcattagaga	1620
tcagggtgccc	tgatctatat	gtgaccttcc	acgacagatt	agtgactgag	tacaaggcct	1680
ctgtggacat	gcctgcgctt	gacattagga	agaagatgga	agaaacggaa	gtgatgtaca	1740
atgcactttc	agaattatcg	gtgttaaggg	agtctgacaa	attcgatgtt	gatgtttttt	1800
cccagatgtg	ccaatctttg	gaagttgacc	caatgacggc	agcgaagggt	atagtcgagg	1860
tcattgagcaa	tgagagcggg	ctgactctca	catttgaacg	acctactgag	gcgaatgttg	1920
cgctagcttt	acaggatcaa	gagaaggctt	cagaagggtc	attggtagtt	acctcaagag	1980
aagttgaaga	accgtccatg	aagggttcga	tggccagagg	agagttacaa	ttagctsggtc	2040

ttgtctggaga	tcatccggaa	tcgtcctatt	ctaagaacga	ggagatagag	tcttttagagc	2100
agtttccatat	ggcgacggca	gattcgttaa	ttcgttaagca	gatgagctcg	atttgtgtaca	2160
cgggtccgat	taaaagttcag	caaatgaaaa	actttatcga	tagcctcgta	gcatacctat	2220
ctgtctcggt	gtcgaatctc	gtcaagatcc	tcaaagatac	agctgctatt	gaccttgaaa	2280
cccgtcaaaa	gtttggagtc	ttggatgttg	catctaggaa	gtggtttaac	aaaccaacgg	2340
ccaagagtca	tgcattgggtg	gttgttgaaa	cccacgcgag	gaagtatcat	gtggcgcttt	2400
tggaatatga	tgagcaggggt	gtgggtgacat	gcgatgattg	gagaagagta	gctgttagct	2460
ctgagtcgtg	tgtttattcc	gacatggcga	aaactcagaac	tctgcgcaga	ctgcttcgaa	2520
acgggagaacc	gcattgtcagt	agcgcaaaagg	ttgttcttgt	ggacggagtt	ccgggctgtg	2580
gaaaaaccaa	agaaattctt	tccagggtta	atthtgatga	agatctaatt	ttagtacctg	2640
ggaagcaagc	cgcggaatg	atcagaagac	gtgcgaattc	ctcagggatt	attgtggcca	2700
cgaaggacaa	cgttaaaacc	gttgattctt	tcatgatgaa	ttttgggaaa	agcacacgct	2760
gtcagttcaa	gagggttattc	attgatgaag	ggttgatgtt	gcatactggg	tgtgttaatt	2820
ttcttctggc	gatgtcattg	tgcaaaattg	catatgttta	cggagacaca	cagcagattc	2880
catacatcaa	tagagtttca	ggattcccg	accccgccca	ttttgcca	ttggaaagtgtg	2940
acgaggttga	gacacgcaga	actactctcc	gttgtccagc	cgatgtcaca	cattatctga	3000
acaggagata	tgagggcttt	gtcatgagca	cttcttcggt	taaaaagtct	gtttcgacgg	3060
agatggctcg	cggagccgccc	gtgatcaatc	cgatctcaaa	acccttgcat	ggcaagatct	3120
tgacttttac	ccaatcggat	aaagaagctc	tgctttcaag	agggtattca	gatgttcaca	3180
ctgtgcata	agtgcaaggc	gagacatact	ctgatgttcc	actagttagg	ttaaacccta	3240
caccgggtcc	catcatttga	ggagacagcc	cacatgtttt	ggtcgcattg	tcaaggcaca	3300
cctgttctgc	caagtactac	actgttgta	tggtatcttt	agttagtatc	attagagatc	3360
tagagaaact	tagtctgtac	ttgttagata	tgtataaggt	cgatgcagga	acacaaatgc	3420
aattacacat	tgactcgggt	ttcaaaaggt	ccaatctttt	tggtgcagcg	ccaaagactg	3480
gtgatatctc	tgatatgcag	ttttactatg	ataagtgtct	cccaggcaac	agcaccatga	3540
tgaataatth	tgatgctgtt	accatgaggt	tgactgacat	ttcatgaa	gtcaaaagatt	3600
gcataattga	tatgtctaa	tctgttgctg	cgcctaagga	tcaaatcaaa	gcactaatac	3660
ctatgggtcg	aacggcgcca	gaaatgccac	gccagactgg	actattggaa	aattttagtgg	3720
cgatgattaa	aagaaaactt	aacgcacccc	agttgtctgg	catcattgat	attgaaaata	3780
ctgcatcttt	gggtgttagat	aagttttttg	atagttatth	gcttaaaaga	aaaagaaaac	3840
caaatcaaaa	tgttcttttg	ttcagtagag	agtctctcaa	tagatgtgta	gaaaagcagg	3900
aacaggtaac	aataggccag	ctcgcagatt	ttgatthtgg	ggatttgcca	gcagttgac	3960
agtacagaca	catgattaaa	gcacaaccca	aacaaaagtt	ggacacttca	atccaaaacgg	4020
agtaaccggc	tttgcagacg	atttgtgtacc	attcaaaaaa	galcaatgca	atattcggcc	4080
cgtttgttag	tgagcttact	aggcaattac	tggacagtgt	tgattcgagc	agatttttgt	4140
ttttcacaag	aaagacacca	gcgcagattg	aggatttctt	cggagatctc	gacagtcacg	4200
tgccgatgga	tgtcttggag	ctggatatat	caaaatacga	caaattctag	aatgaattcc	4260
actgtgcagt	agaatacgag	atctggcgaa	gattgggttt	cgaagacttc	ttgggagaag	4320
tttggaaaac	agggcataga	aagaccaccc	tcaaggatta	taccgcaggt	ataaaaaact	4380
gcactctgga	tcaaaagaaag	agcggggacg	tcacgacggt	cattggaaac	actgtgatca	4440
ttgtctgcat	tttggcctcg	atgcttccga	tggagaaaat	aatcaaaagga	gccttttgcg	4500
gtgacgatag	tctgtgttac	tttccaaaagg	gttgtgagtt	tccggatgtg	caacactccg	4560
cgaatcttat	gtggaatttt	gaagcaaaac	tgthtaaaaa	acagtatgga	tacttttgcg	4620
gaagatatgt	aatacatcac	gacagaggat	gcatttgtga	ttacgatccc	ctaaagttga	4680
ttctcgaact	tggtgtctaaa	cacatcaagg	attgggaaca	cttggaggag	ttcagaaggt	4740
ctcttttgtga	tgttgtgtgt	tcgttgaaca	attgtgcgta	ttacacacag	ttggacgacg	4800
ctgtatggga	ggttcataag	accgccccct	caggttcgtt	tgthtaaaa	agttctggtga	4860
agttatttgc	tgataaagtt	cttttttagaa	gtttgtttat	agatggctct	agttgttaaa	4920
ggaaaagtga	atatcaatga	gtttatcgac	ctgacaaaaa	tggagaagat	cttaccgtcg	4980
atgttttacc	ctgtaaagag	tgthtatgtg	tccaaagtgt	ataaaaataat	ggtttcatgag	5040
aatgagtcac	tgtcaggggt	gaaccttctt	aaaggagtta	agcttattga	tagtggatca	5100
gtctgtttag	ccgggtttgt	cgtcacgggc	gagtggaact	tgcttgacaa	ttgcagagga	5160
gggtgtgagcg	tgtgtctggt	ggacaaaagg	atggaaagag	cgcagagggc	cactctcgga	5220
tcttactaca	cagcagctgc	aaagaaaaga	tttcagttca	aggtcgttcc	caattatgct	5280
ataaccaccc	aggacgcgat	gaaaaacgtc	tggcaagttt	tagttaatat	tagaaaatgtg	5340
aaagtgtcag	cgggtttctg	tccgctttct	ctggagtttg	tgtcgggtgtg	tattgtttat	5400
agaaaataata	taaaatttgg	tttgagagag	aagattacaa	acgtgagaga	cggaggggccc	5460


```

atggaactha cagaagaagt cgttgatgag ttcatggaag atgtccctat gtcgatcagg 5520
cttgcaaaagt ttcgatctcg aaccggaaaa aagagtgatg tccgcaaaagg gaaaaatagt 5580
agtagtgatc ggtagtgcc gaacaagaac tatagaatg ttaaggattt tggagggaatg 5640
agttttaaaa agaataattt aatcgatgat gattcggagg ctactgtcgc cgaatcggat 5700
tcgtttttaa tagatcttac agtatcacta ctcactctca gttcgtgttc ttgtcattaa 5760
ttaaatggct agcaaaaggag aagaactttt cactggagtt gtcccaattc ttgttgaatt 5820
agatggtgat gttaatgggc acaaattttc tgtcagtga gagggtgaag gtgatgttac 5880
atcaggaagg actaccctta aatttatttg cactactgga aaactacctg ttccatggcc 5940
aacacttgtc actactttct cttatgggtt tcaatgcttt tcccgttatt cggatccatat 6000
gaaacggcat gactttttca agagtgccat gccggaagg ttagttacagg aacgcactat 6060
atctttcaaa gatgacggga actacaagac gcgtgctgaa gtcaagtttg aagggtgatac 6120
ccttgtaaat cgtatcgagt taaaagggtat tgatttttaa gaagatggaa acattctcgg 6180
acacaaactc gagtacaact ataactcaca caatgtatac atcacggcag acaaacaaaa 6240
gaatggaatc aaagctaact tcaaaattcg ccacaacatt gaagatggat ccgttcaact 6300
agcagaccat tatcaacaaa atactccaat tggcgatggc cctgtccttt taccagacaa 6360
ccattaccct tcgacacaaat ctgccctttc gaaagatccc aacgaaaagc gtgaccacat 6420
gggcctctct gagtgtgtta ctgctgctgg gattacacat ggcattggat agctctacaa 6480
ataatgacac tcgaggggta gtcaagatgc ataataata acggattgtg tccgtaatac 6540
cacgtggtgc tgacgataac gcatagtgtt ttccctcca cttaaatcga aggggtgtgt 6600
cttggtatgc gcgggtcaaa tgtatatggt tcatatacat ccgaggcac gtaataaagc 6660
gaggggttgc ggtcggagtc ggctgtgaaa ctcgaaaagg ttccggaaaa caaaaaagag 6720
agtggtaggt aatagtgtta ataataagaa aataataat agtggtgaag aagggttgaa 6780
agltgaggaa attgaggata atgtaagtga tgacgagctc atcgcgtcat cgaatagctt 6840
ttaatcaata tgcccttatac aatcaactct ccgagccaat ttgtttactt aagttccgct 6900
tatgcagatc ctgtgcagct gatcaatctg tgtacaaatg cattgggtaa ccagtttcaa 6960
acgcaacaag ctaggacaac agtccaacag caatttgagg atgcctggaa acctgtgcct 7020
agtatgacag tgagatttcc tgcacggat ttctatgtgt atagatataa ttcgacgctt 7080
gatccgttga tcacggcgtt attaaatagc ttcgatacta gaaatagaat aatagagggt 7140
gataatcaac ccgcaccgaa tactactgaa atcgtaacg cgaactcagag ggtagacgat 7200
gcgactgtag ctataagggc ttcaatcaat aatttggtta atgaactggg tctgtggaact 7260
ggcatgttca atcaagcaag ctttgagact gctagtggac ttgtctggac cacaactccg 7320
gctacttagc tattgtgttg agatttccta aaataaagtc actgaagact taaaattcag 7380
ggtggctgat accaaaaatc gcagtggttg ttctgctact taaatataac gattgtcata 7440
tctggatcca acagttaaac catgtgatgg tgtatactgt ggtatggcgt aaaaacaacg 7500
aaaagtgcgt gaagacttaa aattcagggt ggctgatacc aaaaacagca gtggtgtgtc 7560
gtccacttaa aaataacgat tgtcatatct ggatccaaca gttaaaccat gtgatgggtg 7620
atactgtggt atggcgtaaa caacggagag gttcgaatcc tcccctaacc gcgggtagcg 7680
gcca

```

<210> 2

<211> 7686

<212> DNA

<213> *Nicotiana tabacum*

<400> 2

```

gtatttttac aacaattacc aacaacaaca aacaacagac aacattacaa ttactattta 60
caattacaat ggcatacaca cagacagcta ccacatcagc tttgctggac actgtccgag 120
gaaacaactc cttggtcaat gatctagcaa agcgtcgtct ttacgacaca gcggttgaag 180
agtttaacgc tcgtgaccgc aggcccaagg tgaacttttc aaaagtaata agcgaaggagc 240
agacgcttat tgctaccggc gcgtatccag aattccaaat tacattttat aacacgcaaa 300
atgcgctgca ttgcgttgca ggtggattgc gatctttaga actggaatat ctgatgatgc 360
aaattcccta cggatcattg acttatgaca taggcgggaa ttttgcatcg catctgttca 420
agggacgagc atatgtacac tgtgtcatgc ccaacctgga cgttcgagac atcatcgggc 480
acgaaggcca gaaagacagt attgaactat acctttctag gctagagaga ggggggaaaa 540
cagcccccaa cttccaaaag gaagcatttg acagatacgc agaaattcct gaagacgctg 600

```

tctgtcaca	tactttccag	acatgcgaac	atcagccgat	gcagcaatca	ggcagagtgt	660
atgccattgc	gtcacacagc	atatatgaca	taccagccga	tgagttcggg	gcggcactct	720
tgaggaaaa	tgccatagc	tgetatgccg	ctttccactt	ctccgagaac	ctgcttcttg	780
aagattcattg	cgccaatttg	gacgaatca	acgcgtgttt	ttccgcgcgat	ggagacaagt	840
tgaccttttc	ttttgcatca	gagagtactc	tttaattactg	tcatagttat	tctaataatc	900
tttaagtattg	gtgcaaaact	tacttcccgg	cctcctaatag	agagggtttac	atgaaggagt	960
tttttagtcac	cagagttaat	acctggtttt	gtaagttttc	tagaatagat	acttttcttt	1020
tgtaaaaagg	tggtggccat	aaaagtgtag	atagttagca	gttttatact	gcaatggaag	1080
acgcattggca	ttacaaaaag	actcttgcaa	tggtgcaacag	cgagagaatc	ctccttgggg	1140
attcatcacc	agtcatttac	tggtttccca	aaatgagggg	tatggtcatc	gtaccattat	1200
tcgacatttc	tttgagact	agtaagagga	cgcgcaagg	agtcttagtg	tccaaggatt	1260
tcgtgttcac	agtgcttaac	cacattcgaa	cataccaggc	gaaagctctt	acatacgcaa	1320
atgtttttg	cttcgtcgaa	tcgattcgat	cgagggtaat	cattaacggt	gtgacagcga	1380
ggtccgaatg	ggatgtggac	aaatctttgt	tacaatcctt	gtccatgacg	ttttacctgc	1440
ataactaagct	tgccgttcta	aaggatgact	tactgattag	caagtttagt	ctcgggtcga	1500
aaacgggtgtg	ccagcatgtg	tggtgatgaga	tttcgctggc	gtttgggaac	gcatttccct	1560
ccgtaaaaga	gaggtctctg	aacaggaaac	ttatcagagt	ggcaggcgac	gcattagaga	1620
tcagggtgcc	tgatctatat	gtgaccttcc	acgacagatt	agtgactgag	tacaaggcct	1680
ctgtggacat	gcctgcgctt	gacattagga	agaagatgga	agaaacggaa	gtgatgtaca	1740
atgcactttc	agaattatcg	gtgttaagg	agtcctgaca	attcgatgtt	gatgtttttt	1800
cccagatgtg	ccaatctttg	gaagttgacc	caatgacggc	agcgaagggt	atagtcggg	1860
tcattgagcaa	tgagagcggg	ctgactctca	catttgaaacg	acctactgag	gcgaatgttg	1920
cgctagcttt	acaggatcaa	gagaaggctt	cagaagggtc	attggtagtt	acctcaagag	1980
aaagttgaaga	accgtccatg	aagggttcga	tggtccagag	agagttacaa	ttagctgggtc	2040
ttgtctggaga	tcattccgaa	tcgtccctat	ctaagaacga	ggagatagag	tccttagagc	2100
agtttcatat	ggcgacggca	gattcgttaa	ttcgtaagca	gatgagctcg	attgtgtaca	2160
cgggtcccgat	taaagtccag	caaatgaaaa	actttatcga	tagcctggta	gcatacctat	2220
ctgctcgcggt	gtcgaaatctc	gtcaaatctc	tcaaatgatac	agctgctatt	gaccttgaaa	2280
cccgctcaaaa	gtttggagtc	ttggatgttg	catctaggaa	gtggttaatc	aaaccaacgg	2340
ccaagagtc	tgcatggggg	gttggtgaaa	cccacgcgag	ggagtatcat	gtggcgcttt	2400
tggaatatga	tgagcagggt	gtggtgacat	gcgatgattg	gagaagagta	gctgttagct	2460
ctgagttcgt	tggttattcc	gacatggcga	aactcagaac	ttcgcgcaga	ctgcttcgaa	2520
acggagaacc	gcattgtcagt	agcgcaaaag	ttgttcttgt	ggacggaggt	ccgggctgtg	2580
gaaaaaccaa	agaaattctt	tccagggtta	atthttagta	agatctaat	ttagtacctg	2640
ggaaagcaagc	cgcgaaatg	atcagaagac	gtgcgaattc	ctcagggtat	attgtggcca	2700
cgaaggacaa	cgttaaaacc	gttgattctt	tcattgatgaa	ttttgggaaa	agcacacgct	2760
gtcagttcaa	gaggttattc	attgatgaag	ggttgatgtt	gcatactggt	tgtgttaatt	2820
ttcttgtggc	gatgtcattg	tgcgaaattg	catatgttta	cggagacaca	cagcagattc	2880
catatcatcaa	tagagtttca	ggattccccg	accccgccca	ttttgccaaa	tgggaagtgtg	2940
acgaggtgga	gacacgcaga	actactctcc	gttggtccagc	cgatgtcaca	cattatctga	3000
acaggagata	tgagggtctt	gtcatgagca	cttcttcggg	taaaaagtct	gtttcgagg	3060
agatgggtcg	cgagagccgc	gtgatcaatc	cgatctcaaa	acccttgcat	ggcaagatcc	3120
tgactttttac	ccaatcggat	aaagaagctc	tgctttcaag	agggatttca	gatgttcaca	3180
ctgtgcattga	agtgaaggc	gagacatact	ctgattgttc	actagttagg	ttacccctta	3240
caccgggtctc	catcattgca	ggagacagcc	cacatgtttt	ggtcgcattg	tcaaggcaca	3300
cctgtttcgt	caagtactac	actgttgta	tggatccttt	agttagtatc	attagagatc	3360
tagagaaact	tagctcgtac	ttgttagata	tgataaagg	cgatgcagga	acacaatagc	3420
aattacagat	tgactcggg	ttcaaaaggt	ccaatctttt	tggtgcagcg	ccaaagactg	3480
tgatatttct	tgatatgcag	ttttactatg	ataagtgtct	cccaggcaac	agcaccatga	3540
tgaaataattt	tgatgctgtt	accatgaggt	tgactgacat	ttcattgaat	gtcaagatt	3600
gcataattgga	tatgtctaa	tctgttgctg	cacctaaagg	tcaaatcaaa	ccactaatac	3660
ctatgggtacg	aacggcgcca	gaaatgccac	gccagactgg	actattggaa	aatttagtgg	3720
cgatgattaa	aagaaacttt	aacgcacccg	agttgtctgg	catcattgat	attgaaaaata	3780
ctgcatcttt	gggtgtgat	aagttttttg	atagttattt	gcttaaaaga	aaaagaaaac	3840
caaataaaaa	tgtttctttg	ttcagttag	agtcctctaa	tagatgggtta	gaaaagcagg	3900
aacaggtaac	aataggccag	ctcgagatt	ttgattttgt	ggatttgcca	gcagttgatc	3960
agtacagaca	catgatttaa	gcacaacca	aacaaaagtt	ggacacttca	atccaaacgg	4020

agtacccggc	tttggagagc	attgtgtacc	attcaaaaaa	gatcaatgca	atattcggcc	4080
cgttgttttag	tgagcttact	aggcaattac	tggacagtgt	tgatttcgagc	agatttttgt	4140
ttttcacaa	aaagacacca	gogcagattg	aggatttctt	cggagatctc	gacagtcacg	4200
tgccgatgga	tgctctggag	ctggatatat	caaaatcaga	caaattctcag	aatgaattcc	4260
actgtgcagt	agaaatcagag	atctggcgaa	gattgggttt	cgaagacttc	ttgggagaag	4320
tttggaaaca	agggcataga	aagaccaccc	tcaaggatta	taccgcaggt	ataaaaactt	4380
gcctctggta	tcaaaagaa	agcggggacg	tcacgacgtt	cattggaaac	actgtgatca	4440
ttgtctgatg	tttggcctcg	atgcttccga	tgagaaaaat	aatcaaaagga	gccttttgcg	4500
gtgacgatag	tctgtctgtac	tttccaaaag	gttgtgagtt	tccggatgtg	caacactccg	4560
cgaatcttat	gtggaaatttt	gaagcaaaac	tgtttaaaaa	acaglatgga	tacttttgcg	4620
gaagatatgt	aatacatcac	gacagaggat	gcatttgtga	ttacgatccc	ctaaagtga	4680
tctcgaaact	tggtgctaaa	cacatcaagg	attgggaaca	cttggaggag	ttcagaaggt	4740
ctctttgtga	tggtgtctgt	tcgttgaaca	attgtgcgta	ttacacacag	ttggacgacg	4800
ctgtatggga	ggttcataag	accgcccttc	caggttctgt	tgtttataaa	agtctggtga	4860
agtatattgtc	tgataaagtt	cttttttaga	gtttgtttat	agatggctct	agttgttaaa	4920
ggaaaaagtga	atatcaatga	gtttatcgac	ctgacaaaaa	tggagaagat	cttaccgtcg	4980
atgtttaccc	ctgtaaaagag	tgttatgtgt	tccaaagtgt	ataaaataat	ggttcatgag	5040
aatgagtcac	gttcaggggt	gaaccttctt	aaaggagtta	agcttatgga	tagtggtac	5100
gtctgtttag	cgggttttgt	cgtcacgggc	gagtggaaact	tgcttgacaa	ttgcagagga	5160
gggtgtgagcg	tggtgtctgt	ggacaaaagg	atggaaaagag	cgcagcaggg	cattctcggg	5220
tcttactaca	cagcagctgc	aaagaaaaga	tttcagtlca	aggtcgttcc	caattatgct	5280
ataaccaccc	aggacgcgat	gagaaacgtc	tggaagttt	tagttaatat	tagaaatgtg	5340
aagatgtcag	cgggtttctg	tccgctttct	ctggagtttg	tgctgggtgtg	tattgtttat	5400
agaaataata	taaaataggg	tttgagagag	aagattacaa	acgtgagaga	cggaggggccc	5460
atggaaactta	cagaagaagt	cgttgatgag	ttcatggaa	atgtccctat	gtcgatcagg	5520
cttgcaaaagt	ttcgatctcg	aaccggaaaa	aagagtgatg	tccgcaaaag	gaaaaatagt	5580
agtagtgatc	ggtcagtgcc	gaacaagaac	tatagaaatg	ttaaggatgt	tggagggaatg	5640
agtttttaaaa	agaaataatt	aatcgatgat	gattcggagg	ctactgtcgc	cgaatcggat	5700
tcgttttaaaa	tagatcttac	agtatcaacta	ctccatctca	gttcgtgttc	ttgtcattaa	5760
tlaaaatggct	agcaaaaggag	aagaactttt	cactggagtt	gtcccaattc	ttgttgaaat	5820
agatgggtgat	gttaaatgggc	acaaaatttc	tgctcagtgga	gagggtgaag	gtgatgtcac	5880
atcaggaaag	cttcacacta	aattttattg	cactactgga	aaactacctg	ttccatggcc	5940
aacacttgct	actactttct	cttatgggtg	tcaatgcttt	tcccggtatc	cggatcatat	6000
gaaacggcat	gactttttca	agagtgccat	gcccgaaagt	tatgtacagg	aacgcactat	6060
atctttcaaaa	gatgacggga	actacaagac	gcgtgctgaa	gtcaagtttg	aaggtgatac	6120
ccttggttaat	cgtatcgagt	taaaagggtat	tgatttttaaa	gaagatggaa	acattctcgg	6180
acacaaaactc	gagtacaact	ataactcaca	caatgtatac	atcacggcag	acaaaacaaa	6240
gaatgggaatc	aaagctaaact	tcaaaaattcg	ccacaacatt	gaagatggat	cogttcaact	6300
agcagaccat	tatcaacaaa	atactccaat	tggcgatggc	cctgtccttt	taccagacaa	6360
ccattacctg	tcgacacaaat	ctgccctttc	gaaagatccc	aacgaaaagc	gtgaccacat	6420
ggctcctctt	gagttttgaa	ctgctgctgg	gattacacat	ggcatggatg	agctctacaa	6480
ataatgacac	tcgaggggta	gtcaagatgc	ataataaata	acggatgtgt	tccgtaatca	6540
cacgttggtgc	gtacgataac	gcatagtgtt	tttccctcca	cttaaatcga	agggtttgtgt	6600
cttggtatgc	gcgggtcaaa	tgtatatggt	tcataatcat	cgcaggccac	gtataaaagc	6660
gaggggttcg	ggtcgaggtc	ggctgtgaaa	ctcgaaaagg	ttccggaaaa	caaaaaagag	6720
agtggttaggt	aatagtgtta	ataataagaa	aataaataat	agtggttaaga	aaggtttgaa	6780
agttgaggaa	attgaggata	atgtaagtga	tgacgagttc	atcgctcat	cagatcagtt	6840
ttaatacaata	tgccctatac	aatcaactct	cgcagccaat	ttgttlaact	aagttccgct	6900
tatgcagatc	ctgtgcagct	gatcaatctg	tgtaacaaatg	cattgggttaa	ccagtttcaa	6960
acgcaacaag	ctaggacaac	agtccaacag	caatttgcgg	atgcctggaa	acctgtgcct	7020
agtatgacag	tgagatttcc	tgcatcggat	ttctatgtgt	atagatataa	ttcgacgctt	7080
gatccgttga	tcacggcggt	attaaatagc	ttcgatacta	gaaatagaat	aatagaggtt	7140
gataatacaac	ccgcaccgaa	tactactgaa	atcggttaacg	cgactcagag	ggtagacgat	7200
gcgactgtgag	ctataagggc	ttcaatcaat	aatttggtta	atgaactggt	tcgtggaaact	7260
ggcatgttca	atcaagcaag	ctttgagact	gctagtggac	ttgtctggac	cacaactccg	7320
gctactctagc	tattgttgtg	agatttctca	aaataaagtc	actgaagact	taaaattcag	7380
gggtggctgat	acccaaaatca	gcagtggttg	ttcgtccact	taaatataac	gattgtcata	7440

```

tctggatcca acagtttaaa catgtgatgg tgtatactgt qgtatggcgt aaaacaacgg 7500
aaaagtcgct gaagacttaa aattcagggg ggctgatacc aaatcagca gtggttggtc 7560
gtccacttaa aaataacgat tgtcatatct ggatccaaca gttaaacat gtgatggtgt 7620
atactgtggt atggcgtaaa acaacggaga gggtcgaatc ctcccctaac cgcgggtagc 7680
ggccca 7686

```

<210> 3
 <211> 857
 <212> DNA
 <213> Nicotiana tabacum

```

<400> 3
atggctctag ttgttaaagg aaaagtgaat atcaatgagt ttatcgacct gacaaaaatg 60
gagaagatct taccgtcgat gtttaccctt gttaaagagt ttatgtgttc caaagtgtat 120
aaaataatgg ttcatgagaa tgagtcattg tcagggggtga accttcttaa aggagttaag 180
cttattgata gtggatacgt ctgtttagcc ggtttggtcg tcacggggcga gtggaacttg 240
cctgacaatt gcagaggagg tgtgagcgtg tgtctgggtg acaaaaggat ggaaagagcc 300
gacgaggcca ctctcggatc ttactacaca gcagctgcaa agaaaagatt tcagttcaag 360
gtcgttccca attatgctat aaccacccag gacgcgatga aaaacgtctg gcaagtttta 420
gttaataatta gaaatgtgaa gatgtcagcg aaaaacgtctg gcaagtttta gttaataatta 480
gaaatgtgaa gatgtcagcg ggtttctgtc cgctttctct ggagtttgtg tcggttgtga 540
ttgtttatag aaataatata aaattaggtt tgagagagaa gattacaaac gtgagagacg 600
gagggcccat ggaacttaca gaagaagtcg ttgatgagtt catggaagat gtccctatgt 660
cgatcaggct tgcaaaagtt cgatctcgaa ccggaaaaaa gagtgtgttc cgcaaaagga 720
aaaatagtag tagtgatcgg tcagtgcgga acaagaacta tagaaatgtt aaggattttg 780
gaggaatgag ttttaaaaag aataatttaa tcgatgatga ttcgagggct actgtcgcgc 840
aatcggatcc gtttttaa 857

```

<210> 4
 <211> 807
 <212> DNA
 <213> Nicotiana tabacum

```

<400> 4
atggctctag ttgttaaagg aaaagtgaat atcaatgagt ttatcgacct gacaaaaatg 60
gagaagatct taccgtcgat gtttaccctt gttaaagagt ttatgtgttc caaagtgtat 120
aaaataatgg ttcatgagaa tgagtcattg tcagggggtga accttcttaa aggagttaag 180
cttattgata gtggatacgt ctgtttagcc ggtttggtcg tcacggggcga gtggaacttg 240
cctgacaatt gcagaggagg tgtgagcgtg tgtctgggtg acaaaaggat ggaaagagcc 300
gacgaggcca ttctcggatc ttactacaca gcagctgcaa agaaaagatt tcagttcaag 360
gtcgttccca attatgctat aaccacccag gacgcgatga gaaacgtctg gcaagtttta 420
gttaataatta gaaatgtgaa gatgtcagcg ggtttctgtc cgctttctct ggagtttgtg 480
tcggttgtga ttgtttatag aaataatata aaattaggtt tgagagagaa gattacaaac 540
gtgagagacg gagggcccat ggaacttaca gaagaagtcg ttgatgagtt catggaagat 600
gtccctatgt cgatcaggct tgcaaaagtt cgatctcgaa ccggaaaaaa gagtgtgttc 660
cgcaaaagga aaaatagtag tagtgatcgg tcagtgcgga acaagaacta tagaaatgtt 720
aaggattttg gaggaatgag ttttaaaaag aataatttaa tcgatgatga ttcgagggct 780
actgtcgcgc aatcggatcc gtttttaa 807

```

<210> 5
 <211> 267
 <212> PRT
 <213> Nicotiana tabacum

<400> 5

```

Met Ala Leu Val Val Lys Gly Lys Val Asn Ile Asn Glu Phe Ile Asp
 1          5          10          15
Leu Thr Lys Met Glu Lys Ile Leu Pro Ser Met Phe Thr Pro Val Lys
          20          25          30
Ser Val Met Cys Ser Lys Val Lys Ile Met Val His Glu Asn Glu Ser
          35          40          45
Leu Ser Gly Val Asn Leu Leu Lys Gly Val Lys Leu Ile Asp Ser Gly
          50          55          60
Tyr Val Cys Leu Ala Gly Leu Val Val Thr Gly Glu Trp Asn Leu Pro
          65          70          75          80
Asp Asn Cys Arg Gly Gly Val Ser Val Cys Leu Val Asp Lys Arg Met
          85          90          95
Glu Arg Ala Asp Glu Ala Thr Leu Gly Ser Tyr Tyr Thr Ala Ala Ala
          100          105          110
Lys Lys Arg Phe Gln Phe Lys Val Val Pro Asn Tyr Ala Ile Thr Thr
          115          120          125
Gln Asp Ala Met Lys Asn Val Trp Gln Val Leu Val Asn Ile Arg Asn
          130          135          140
Val Lys Met Ser Ala Gly Phe Cys Pro Leu Ser Leu Glu Phe Val Ser
          145          150          155          160
Val Cys Ile Val Tyr Arg Asn Asn Ile Lys Leu Gly Leu Arg Glu Lys
          165          170          175
Ile Thr Asn Val Arg Asp Gly Gly Pro Met Glu Leu Thr Glu Glu Val
          180          185          190
Val Asp Glu Phe Met Glu Asp Val Pro Met Ser Ile Arg Leu Ala Lys
          195          200          205
Phe Arg Ser Arg Thr Gly Lys Lys Ser Asp Val Arg Lys Gly Lys Asn
          210          215          220
Ser Ser Ser Asp Arg Ser Val Pro Asn Lys Asn Tyr Arg Asn Val Lys
          225          230          235          240
Asp Phe Gly Gly Met Ser Phe Lys Lys Asn Asn Leu Ile Asp Asp Asp
          245          250          255
Ser Glu Ala Thr Val Ala Glu Ser Asp Ser Phe
          260          265

```

```

<210> 6
<211> 268
<212> PRT
<213> Nicotiana tabacum

```

```

<400> 6
Met Ala Leu Val Val Lys Gly Lys Val Asn Ile Asn Glu Phe Ile Asp
 1          5          10          15
Leu Thr Lys Met Glu Lys Ile Leu Pro Ser Met Glu Thr Pro Val Lys
          20          25          30
Ser Val Met Cys Ser Lys Val Asp Lys Ile Met Val His Glu Asn Glu
          35          40          45
Ser Leu Ser Gly Val Asn Leu Leu Lys Gly Val Lys Leu Ile Asp Ser
          50          55          60
Gly Tyr Val Cys Leu Ala Gly Leu Val Val Thr Gly Glu Trp Asn Leu
          65          70          75          80
Pro Asp Asn Cys Arg Gly Gly Val Ser Val Cys Leu Val Asp Lys Arg
          85          90          95
Met Glu Arg Ala Asp Glu Ala Thr Leu Gly Ser Tyr Tyr Thr Ala Ala
          100          105          110
Ala Lys Lys Arg Phe Gln Phe Lys Val Val Pro Asn Tyr Ala Ile Thr
          115          120          125

```

```

Thr Gln Asp Ala Met Lys Asn Val Trp Gln Val Leu Val Asn Ile Arg
 130          135          140
Asn Val Lys Met Ser Ala Gly Phe Cys Pro Leu Ser Leu Glu Phe Val
145          150          155          160
Ser Val Cys Ile Val Tyr Arg Asn Asn Ile Lys Leu Gly Leu Arg Glu
          165          170          175
Lys Ile Thr Asn Val Arg Asp Gly Gly Pro Met Glu Leu Thr Glu Glu
          180          185          190
Val Val Asp Glu Phe Met Glu Asp Val Pro Met Ser Ile Arg Leu Ala
          195          200          205
Lys Phe Arg Ser Arg Thr Gly Lys Lys Ser Asp Val Arg Lys Gly Lys
          210          215          220
Asn Ser Ser Ser Asp Arg Ser Val Pro Asn Lys Asn Tyr Arg Asn Val
225          230          235          240
Lys Asp Phe Gly Gly Met Ser Phe Lys Lys Asn Asn Leu Ile Asp Asp
          245          250          255
Asp Ser Glu Ala Thr Val Ala Glu Ser Asp Ser Phe
          260          265

```

What is claimed is:

1. An isolated nucleic acid comprising: (a) a nucleotide sequence encoding an altered viral movement protein having the amino acid sequence shown in SEQ ID NO: 6; and (b) a nucleotide sequence encoding an altered
5 126/183 replicase complex, wherein the nucleotide sequence has a nucleotide alteration at positions 1138, 1268, 2382, and 3632 as shown in SEQ ID NO: 2.

2. The isolated nucleic acid of claim 1 wherein the nucleotide sequence in (a) comprises a Thymine (T) or Uracil (U) residue at position
10 5213 and Guanine (G) residue at 5303 as shown in Figure 1A, and wherein said nucleotide sequence and a nucleotide sequence as set forth in SEQ ID NO: 4 are capable of hybridizing to form a duplex with the same complementary polynucleotide.

3. The isolated nucleic acid of claim 1 wherein the nucleotide
15 sequence in (a) is identical to the sequence shown in SEQ ID NO: 4.

4. The isolated nucleic acid according to claim 1, wherein the altered 126/183 replicase complex enhances the stabilization of a transgene contained in a virus comprising the isolated nucleic acid that expresses the altered replicase complex.

5. The isolated nucleic acid sequence of claim 1, wherein the
20 altered movement protein enhances the ability to facilitate stabilization of a transgene contained in a virus comprising the isolated nucleic acid that expresses the altered movement protein.

6. A viral vector comprising: (a) a nucleotide sequence encoding

an altered viral movement protein having the amino acid sequence shown in SEQ ID NO: 6; and (b) a nucleotide sequence encoding an altered 126/183 replicase complex, wherein the nucleotide sequence has a nucleotide alteration at positions 1138, 1268, 2382, and 3632 as shown in SEQ ID NO:

5 2.

7. The viral vector of claim 6, which is capable of exhibiting an enhanced ability compared to a control viral vector to stabilize a transgene contained in the vector.

8. The viral vector of claim 6, wherein the vector is a tobacco
10 mosaic viral vector.

9. The viral vector of claim 7, wherein the transgene is a non-viral gene.

10. The viral vector of claim 9, wherein the non-viral transgene encodes a protein selected from the group consisting of a membrane protein, a
15 cytosolic protein, a secreted protein, a nuclear protein, and a chaperon protein.

11. The viral vector of claim 6 that is designated BSG1057 deposited with American Type Culture Collection accession number 203981.

12. A cell transformed with the viral vector of claim 6.

20 13. The cell of claim 12 that is a plant cell.

14. A transgenic plant comprising the viral vector of claim 6.

15. An isolated nucleic acid comprising a nucleotide sequence set forth in SEQ ID NO: 2.

Comparison of sequences from BSG1037 (SEQ ID NO.:3) and BSG1057 (SEQ ID NO.:4) in the 30kDa movement protein coding region (nts 4903-5709). Non-identities are indicated by _ and identities are indicated by *.

```

1037      ATGGCTCTAGTTGTTAAAGGAAAAGTGAATATCAATGAGTTATCGACCT
1057      ATGGCTCTAGTTGTTAAAGGAAAAGTGAATATCAATGAGTTATCGACCT
*****

1037      GACAAAAATGGAGAAGATCTTACCGTCGATGTTTACCCCTGTAAGAGTG
1057      GACAAAAATGGAGAAGATCTTACCGTCGATGTTTACCCCTGTAAGAGTG
*****

1037      TTATGTGTTCCAAAGTTGATAAAATAATGGTTCATGAGAATGAGTCATTG
1057      TTATGTGTTCCAAAGTTGATAAAATAATGGTTCATGAGAATGAGTCATTG
*****

1037      TCAGGGGTGAACCTTCTTAAGGAGTTAAGCTTATTGATAGTGGATACGT
1057      TCAGGGGTGAACCTTCTTAAGGAGTTAAGCTTATTGATAGTGGATACGT
*****

1037      CTGTTTAGCCCGTTTGGTCGTACCGGCGAGTGGAACTTGCTGACAATT
1057      CTGTTTAGCCCGTTTGGTCGTACCGGCGAGTGGAACTTGCTGACAATT
*****

1037      GCAGAGGAGGTGTGAGCGTGTGTCTGTTGGACAAAAGGATGGAAGAGCC
1057      GCAGAGGAGGTGTGAGCGTGTGTCTGTTGGACAAAAGGATGGAAGAGCC
*****

1037      GACGAGGCCACTCTCGGATCTTACTACACAGCAGCTGCAAGAAAAGATT
1057      GACGAGGCCACTCTCGGATCTTACTACACAGCAGCTGCAAGAAAAGATT
*****

1037      TCAGTTCAAGGTGTTCCCAATTATGCTATAACCCAGGACGCGATGA
1057      TCAGTTCAAGGTGTTCCCAATTATGCTATAACCCAGGACGCGATGA
*****

1037      AAAACGTCTGGCAAGTTTGTGTTAATATTAGAAATGTGAAGATGTCAGCG
1057      GAAACGTCTGGCAAGTTTGTGTTAATATTAGAAATGTGAAGATGTCAGCG
*****

1037      GGTTCGTGTCGCTTCTCTGGAGTTTGTGTCGGTGTGATTGTTTATAG
1057      GGTTCGTGTCGCTTCTCTGGAGTTTGTGTCGGTGTGATTGTTTATAG
*****

1037      AAATAATATAAAATTAGGTTTGAGAGAGAAGATTACAAACGTGAGAGACG
1057      AAATAATATAAAATTAGGTTTGAGAGAGAAGATTACAAACGTGAGAGACG
*****

1037      GAGGGCCCATGGAACTTACAGAAGAAGTCGTTGATGAGTTCATGGAAGAT
1057      GAGGGCCCATGGAACTTACAGAAGAAGTCGTTGATGAGTTCATGGAAGAT
*****

1037      GTCCTATGTCGATCAGGCTTGCAAAGTTTCGATCTCGAACCAGAAAAAA
1057      GTCCTATGTCGATCAGGCTTGCAAAGTTTCGATCTCGAACCAGAAAAAA
*****

```

Figure 1a

```
1057      GAGTGATGTCCGCARAGGGAAAAATAGTAGTAGTGATCGGTCAGTGCCGA
      GAGTGATGTCCGCARAGGGAAAAATAGTAGTAGTGATCGGTCAGTGCCGA
      *****
1037      ACAAGAACTATAGAAATGTTAAGGATTTGGAGGAATGAGTTTTAAAAAG
1057      ACAAGAACTATAGAAATGTTAAGGATTTGGAGGAATGAGTTTTAAAAAG
      *****
1037      AATAATTTAATCGATGATGATTCGGAGGCTACTGTCGCCGAATCGGATTC
1057      AATAATTTAATCGATGATGATTCGGAGGCTACTGTCGCCGAATCGGATTC
      *****
1037      GTTTAA
1057      GTTTAA
      *****
```

Fig. 1b
2/17

Comparison of sequences from BSG1037 (SEQ ID NO.:5) and BSG1057 (SEQ ID NO.:6) in the 30kDa movement protein (aal-268).
Non-identities are indicated by _ and identities are indicated by *.

```

1037      MALVVGKGVNINEFIDLTMEKILPSMETPVKSVMCCKVOKIMVHENESL
1057      MALVVGKGVNINEFIDLTMEKILPSMETPVKSVMCCKVOKIMVHENESL
*****

1037      SGVNLLKGVKLIIDSGYVCLAGLVVTGEWNLPDNCRGGVSVCLVDKRMERA
1057      SGVNLLKGVKLIIDSGYVCLAGLVVTGEWNLPDNCRGGVSVCLVDKRMERA
*****

1037      DEATLGSYYTAAAKRFQFKVVPNYAITTQDAMKNVWQVLNIRNVKMSA
1057      DEAILGSYYTAAAKRFQFKVVPNYAITTQDAMKNVWQVLNIRNVKMSA
*** _ *****

1037      GFCPLSLEFVSVCIYRNNIKLGLREKITNVRDGGPMELTEEVVDEFMED
1057      GFCPLSLEFVSVCIYRNNIKLGLREKITNVRDGGPMELTEEVVDEFMED
*****

1037      VPMSIRLAKFRSRTGKKS DVRKGKNSSSDRSVPKNYRNVDI GGMSFKK
1057      VPMSIRLAKFRSRTGKKS DVRKGKNSSSDRSVPKNYRNVDI GGMSFKK
*****

1037      NNLIDDDEATVAESDSF
1057      NNLIDDDEATVAESDSF
*****

```

Figure 2

Sequence 10403 b.p. GTATTTTACAA ... CGACTCACTATA circular
 126/183 reading frame begins at 69, 3417 is suppressible stop codon, and ends at 4919.30K

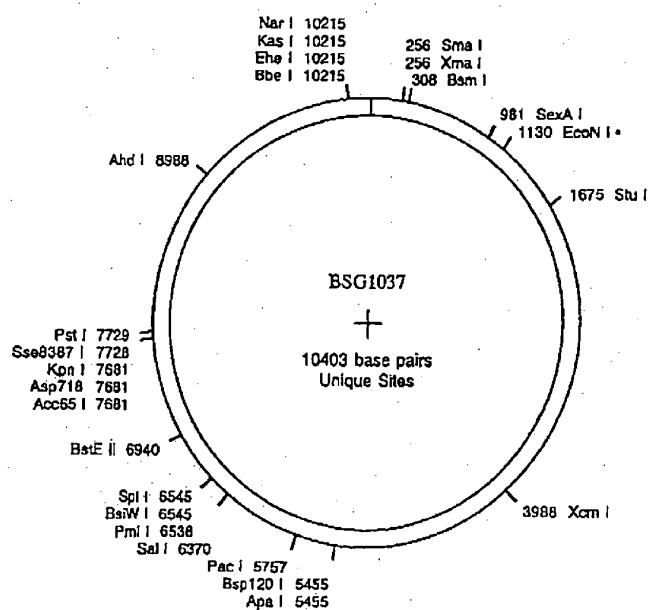


Figure 3

... sequence 10403 b.p. GTATTTTACAA ... CGACTCACTATA circular

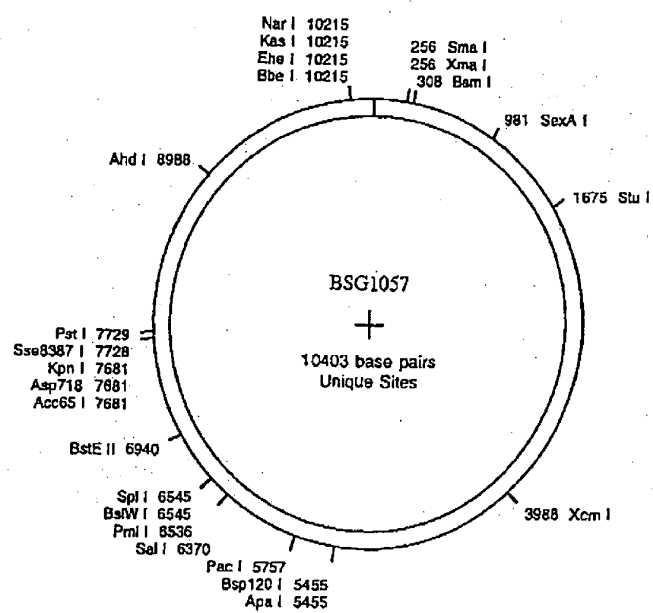


Figure 4

Complete sequence of BSG 1037 (SEQ ID NO.: 1)

GTATTTTACAAACAATTACCAACAACAACAAACAACAGACAACATTACAATTACTAT
TTACAATTACAATGGCATAACACA
CAGACAGCTACCACATCAGCTTTGCTGGACACTGTCCGAGGAAACAACCTCCTTGGTC
AATGATCTAGCAAAGCGTCGTCT
TTACGACACAGCGTTGAAGAGTTTAACGCTCGTGACCGCAGGCCCAAGGTGAACT
TTTCAAAAGTAATAAGCGAGGAGC
AGACGCTTAITGCTACCCGGCGGTATCCAGAATTCCAAATTACATTTTATAACACGC
AAAAATGCCGTGCATTGCTTGCA
GGTGGATTGCGATCTTTAgAACTGGAATATCTGATGATGCAAATTCCTACGGATCA
TTGACTTATGACATAGGCGGGAA
TTTTGCATCGCATCTGTTCAAGGGACGAGCATATGTACACTGCTGCATGCCAACCT
GGACGTTCCGAGACATCATGCCGC
ACGAAGGCCAGAAAGACAGTATTGAACTATACCTTTCTAGGCTAGAGAGAGGGGGG
AAAACAGTCCCCAACTTCCAAAAG
GAAGCATTTGACAGATACGCAGAAATTCCTGAAGACGCTGTCTGTCCACAATACTTTC
CAGACATGCGAACATCAGCCGAT
GCAGCAATCAGGCAGAGTGTATGCCATTGCGCTACACAGCATATATGACATACCAG
CCGATGAGTTCCGGGCGGCACTCT
TGAGGAAAAATGTCCATACGTGCTATGCCGCTTCCACTTCTCCGAGAACCTGCTTC
TTGAAGATTTCATGCGTCAATTTG
GACGAAATCAACGCGTGTTTTCGCGCGATGGAGACAAGTTGACCTTTTCTTTTGA
TCAGAGAGTACTCTTAATTAATCTG
TCATAGTTATTTCTAATATTCTTAAGTATGTGTGCAAAACTTACTTCCCGCCTCTAAT
AGAGAGGTTTACATGAAGGAGT
TTTAGTCACCAGAGTTAATACCTGGTTTGTAAAGTTTCTAGAATAGATACTTTTCT
TTTGTACAAAGGTGTGGCCCAT
AAAAGTGTAGATAGTGAGCAGTTTATACTGCAATGGAAGACGCATGGCATTACAA
AAAGACTCTTGCAATGTGCAACAG
CGAGAGAATCCTCCTTGAGGATTCATCATCAGTCAATTACTGGTTTCCCAAAATGAG
GGATATGGTCATCGTACCATTAT
TCGACATTTCTTTGGAGACTAGTAAGAGGACGCGCAAGGAAGTCTTAGTGTCCAAGG
ATTCGTGTTTACAGTGCTTAAC
CACATTGCAACATACCAGGCGAAAGCTTTACATACGCAAAATGTTTGTCTTCGTC
GAATCGATTGATCGAGGGTAAT
CATTAAACGGTGTGACAGCGAGGTCCGAATGGGATGTGGACAAATCTTTGTTACAATC
CTTGTCATGACGTTTTACCTGC
ATACTAAGCTTGCCGTTCTAAAGGATGACTTACTGATTAGCAAGTTTAGTCTCGGT
CGAAAACGGTGTGCCAGCATGTG
TGGGATGAGATTTGCTGGCGTTTGGGAACGCATTTCCCTCCGTGAAAGAGAGGCTC
TTGAACAGGAAACTTATCAGAGT
GGCAGGCGACGCATTAGAGATCAGGGTGCCTGATCTATATGTGACCTTCCACGACA
GATTAGTGACTGAGTACAAGGCCT
CTGTGGACATGCCGTGCGCTTGACATTAGGAAGAAGATGGAAGAAACGGAAGTGATG
TACAATGCACTTTCAGAATTATCG

Figure 5a

6/17

GTGTTAAGGGAGTCTGACAAATTCGATGTTGATGTTTTTCCCAGATGTGCCAATCTT
TGGAAGTTGACCCAATGACGGC
AGCGAAGGTTATAGTCGCGGTCTAGCAATGAGAGCGGTCTGACTCTCACATTTGA
ACGACCTACTGAGGCGAATGTTG
CGCTAGCTTTACAGGATCAAGAGAAGGCTTCAGAAGGTGCATTGGTAGTTACCTCAA
GAGAAGTTGAAGAACCGTCCATG
AAGGGTTCGATGGCCAGAGGAGATTACAATTAGCTGGTCTTGCTGGAGATCATCCG
GAATCGTCCTATTCTAAGAACGA
GGAGATAGAGTCTTTAGAGCAGTTTCATATGGCGACGGCAGATTTCGTTAATTCGTAA
GCAGATGAGCTCGATTGTGTACA
CGGGTCCGATTAAAGTTTCAGCAAATGAAAACTTTATCGATAGCCTGGTAGCATCAC
TATCTGCTGCGGTGTCTGAATCTC
GTCAAGATCCTCAAAGATACAGCTGCTATTGACCTTGAAACCCGTCAAAAGTTTGA
GTCTTGATGTTGCATCTAGGAA
GTGGTTAATCAAACCAACGGCCAAGAGTCATGCATGGGGTGTGTTGTAACCCACG
CGAGGAAGTATCATGTGGCGCTT
TGGAATATGATGAGCAGGGTGTGGTGACATGCGATGATTGGAGAAGAGTAGCTGTT
AGCTCTGAGTCTGTTGTTTATCC
GACATGGCGAAACTCAGAACTCTGCGCAGACTGCTTCGAAACGGAGAACCGCATGT
CAGTAGCGCAAAGGTTGTTCTTGT
GGACGGAGTTCGGGGCTGTGGAAAAACCAAAGAAATCTTTCCAGGGTTAATTTTGA
TGAAGATCTAATTTTAGTACCTG
GGAAGCAAGCCGCGGAAATGATCAGAAGACGTGCGAATTCCTCAGGGATTATTGTG
GCCACGAAGGACAACGTTAAACCC
GTTGATTCTTTTCATGATGAATTTTGGGAAAAGCACACGCTGTCTAGTTCAAGAGGTTA
TTCATTGATGAAGGGTTGATGTT
GCATACTGGTTGTGTTAATTTCTTGTGGCGATGTCATTGTGCGAAATTGCATATGTT
TACGGAGACACACAGCAGATTC
CATACATCAATAGAGTTTCAGGATTCCCGTACCCCGCCCATTTTGCCAAATTGGAAG
TTGACGAGGTGGAGACACGCAGA
ACTACTCTCCGTTGTCCAGCCGATGTCACACATTATCTGAACAGGAGATATGAGGGC
TTTGTCTAGAGCACTTCTTCGGT
TAAAAAGTCTGTTTCGCAGGAGATGGTCGGCGGAGCCgCCGTGATCAATCCGATCTC
AAAACCCCTTGCATGGCAAGATCT
TGACTTTTACCCAATCGGATAAAGAAGCTCTGCTTTCAAGAGGGTATTCAGATGTTT
ACACTGTGCATGAAGTGCAAGGC
GAGACATACTCTGATGTTTCACTAGTTAGGTTAACCCTACACCGGTCTCCATCATTG
CAGGAGACAGCCACATGTTTT
GGTCGCTTGTCAAGGCACACCTGTTGCTCAAGTACTACACTGTTGTTATGGATCC
TTTAGTTAGTATCATTAGAGATC
TAGAGAACTTAGCTCGTACTGTTAGATATGTATAAGGTGCGATGCAGGAACACAAT
AGCAATTACAGATTGACTCGGTG
TTCAAAGGTTCCAATCTTTTGTGTCAGCGCCAAAGACTGGTGATATTCTGATATGC
AGTTTACTATGATAAGTGTCT
CCCAGGCAACAGCACCATGATGAATAATTTTGATGCTGTTACCATGAGGTTGACTGA
CATTTCAATTGAATGTCAAAGATT

Figure 5b

GCATATTGGATATGTCTAAGTCTGTTGCTGCGCCTAAGGATCAAATCAAACCACTAA
TACCTATGGTACGAACGGCGGCA
GAAATGCCACGCCAGACTGGACTATTGGAAAATTTAGTGGCGATGATTAAGAA
CTTTAACGCACCCGAGTTGTCTGG
CATCATTGATATTGAAAATACTGCATCTTTGGTTGTAGATAAGTTTTTTGATAGTTAT
TTGCTTAAAGAAAAAGAAAAAC
CAAATAAAAAATGTTTCTTTGTTCAGTAGAGAGTCTCTCAATAGATGGTTAGAAAAGC
AGGAACAGGTAACAATAGGCCAG
CTCGCAGATTTTGATTTTGTGGATTGCCAGCAGTTGATCAGTACAGACACATGATT
AAAGCACAACCCAAACAAAAGTT
GGACACTTCAATCCAAACGGAGTACCCGGCTTTGCAGACGATTGTGTACCATTCAAA
AAAGATCAATGCAATATTCGGCC
CGTTGTTTGTAGTGTACTAGGCAATTACTGGACAGTGTGATTGAGCAGATTTTT
GTTTTTACAAGAAAGACACCA
GCGCAGATTGAGGATTTCTTCGGAGATCTCGACAGTCATGTGCCGATGGATGTCTTG
GAGCTGGATATATCAAAATACGA
CAAATCTCAGAATGAATCCACTGTGCAGTAGAATACGAGATCTGGCGAAGATTGG
GTTTCGAAGACTTCTTGGGAGAAG
TTTGAAACAAGGCGATAGAAAGACCACCTCAAGGATTATACCGCAGGTATAAAA
ACTTGCATCTGGTATCAAAGAAAG
AGCGGGGACGTCACGACGTTTATTGAAACACTGTGATCATTGCTGCATGTTTGGCC
TCGATGCTTCCGATGGAGAAAAAT
AATCAAAGGAGCCTTTTGGCGTGACGATAGTCTGCTGTACTTTCAAAGGGTTGTGA
GTTTCCGGATGTGCAACACTCCG
CGAATCTTATGTGAATTTGAAGCAAACTGTTTAAAAAACAGTATGGATACTTTT
GCGGAAGATATGTAATACATCAC
GACAGAGGATGCATTGTGTATTACGATCCCCTAAAGTTGATCTCGAAACTTGGTGCT
AAACACATCAAGGATTGGGAACA
CTTGGAGGAGTTCAGAAGGTCTCTTTGTGATGTTGCTGTTTCGTTGAACAATTGTGCG
TATTACACACAGTTGGACGACG
CTGTATGGGAGGTTTATAAGACCGCCCTCCAGGTTGCTTTGTTTATAAAAGTCTGG
TGAAGTATTTGTCTGATAAAGTT
CTTTTTAGAAAGTTTGTATAGATGGCTCTAGTGTAAAGGAAAAGTGAATATCAA
TGAGTTTATCGACCTGACAAAAA
TGGAGAAGATCTTACCGTCGATGTTTACCCCTGTAAAGAGTGTATGTGTTCCAAAG
TTGATAAAATAATGGTTCATGAG
AATGAGTCATTGTCAGGGGTGAACCTTCTTAAAGGAGTTAAGCTTATTGATAGTGA
TACGTCTGTTTAGCCGGTTTGGT
CGTCACGGGCGAGTGGAACCTTGCCTGACAATTGCAGAGGAGGTGTGAGCGTGTGTC
TGGTGGACAAAAGGATGGAAAGAG
CCGACGAGGCCACTCTCGGATCTTACTACACAGCAGCTGCAAAGAAAAGATTTAG
TTCAAGGTCGTTCCCAATTATGCT
ATAACCACCCAGGACGCGATGAAAAACGTCTGGCAAGTTTTAGTTAATATTAGAAAT
GTGAAGATGTCAGCGGTTTCTG
TCCGCTTCTCTGGAGTTTGTGTCGGTGTGATTGTTTATAGAAATAATATAAAATTA
GGTTTGAGAGAGAAGATTACAA

Figure 5c

ACGTGAGAGACGGAGGGGCCATGGAACCTACAGAAGAAGTCGTTGATGAGTTCATG
GAAGATGTCCCTATGTCGATCAGG
CTTGCAAAGTTTCGATCTCGAACCAGGAAAAAAGAGTGATGTCCGCAAAGGGAAAAA
TAGTAGTAGTGATCGGTCAGTGCC
GAACAAGAAGCTATAGAAATGTTAAGGATTTTGGAGGAATGAGTTTTAAAAAGAATA
ATTTAATCGATGATGATTTCGGAGG
CTACTGTGCGCGAATCGGATTCGTTTTAAATAGATCTTACAGTATCACTACTCCATCT
CAGTTCGTGTTCTTGTCATTAA
TTAAATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGA
ATTAGATGGTGATGTTAATGGGC
ACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGCTTACCC
TTAAATTTATTGCACTACTGGA
AAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCT
TTTCCCGTTATCCGGATCATAT
GAAACGGCATGACTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCA
CTATATCTTTCAAAGATGACGGGA
ACTACAAGACGCGTGCTGAAGTCAAGTTGAAGGTGATACCTTGTTAATCGTATCG
AGTTAAAAGGTATTGATTTTAAA
GAAGATGGAACATTCTCGGACACAACTCGAGTACAATACTACACAATGT
ATACATCACGGCAGACAAACAAAA
GAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTT
AACTAGCAGACCATATCAACAAA
ATACTCCAATTGGCGATGGCCCTGTCTTTTACCAGACAACCATTACCTGTGACAC
AATCTGCCCTTTCGAAAGATCCC
AACGAAAAGCGTGACCACATGGGCCTTCTTGAGTTGTAACTGCTGCTGGGATTACA
CATGGCATGGATGAGCTCTACAA
ATAATGACACTCGAGGGGTAGTCAAGATGCATAATAAATAACGGATTGTGTCCGTA
ATCACACGTGGTGCGTACGATAAC
GCATAGTGTTTTTCCCTCCACTTAAATCGAAGGGTTGTGTCTGGATCGCGCGGGTC
AAATGTATATGGTTCATATACAT
CCGACGGCACGTAATAAAGCGAGGGGTTCCGGTCGAGGTCGGCTGTGAAACTCGAA
AAGGTTCCGGAACAAAAAGAG
AGTGGTAGGTAATAGTGTTAATAATAAGAAAATAAATAATAGTGGTAAGAAAAGTT
TGAAAGTTGAGGAAATTGAGGATA
ATGTAAGTGATGACGAGTCTATCGCGTCATCGAGTACGTTTTAATCAATATGCCTTA
TACAATCAACTCTCCGAGCCAAT
TTGTTTACTTAAGTTCCGCTTATGCAGATCCTGTGCAGCTGATCAATCTGTGTACAAA
TGCATTGGGTAACCAAGTTTCAA
ACGCAACAAGCTAGGACAACAGTCCAACAGCAATTTGCGGATGCCTGGAAACCTGT
GCCTAGTATGACAGTGAGATTTCC
TGCATCGGATTTCTATGTGTATAGATATAATTTCGACGCTTGATCCGTTGATCACGGC
GTTATTAAATAGCTTCGATACTA
GAAATAGAATAATAGAGGTTGATAATCAACCCGCACCGAATACTACTGAAATCGTT
AACGCGACTCAGAGGGTAGACGAT
GCGACTGTAGCTATAAGGGCTTCAATCAATAATTGGCTAATGAACTGGTTCGTGGA
ACTGGCATGTTCAATCAAGCAAG

Figure 5d

9/17

CTTTGAGACTGCTAGTGGACTTGTCTGGACCACAACCTCCGGCTACTTAGCTATTGTTG
TGAGATTTCTAAAATAAAGTC
ACTGAAGACTTAAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTTGTTTCGTCC
ACTTAAATATAACGATTGTCATA
TCTGGATCCAACAGTTAAACCATGTGATGGTGTATACTGTGGTATGGCGTAAAAACA
CGGAAAAGTCGCTGAAGACTTAA
AATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTTGTTTCGTCCACTTAAAAATAAC
GATTGTCATATCTGGATCCAACA
GTIAAACCATGTGATGGTGTATACTGTGGTATGGCGTAAACAACGGAGAGGTTCGA
ATCCTCCCCTAACCGCGGtagcg
gccca

Figure 5e

10/17

Complete Sequence of BSG 1057 (SEQ ID NO.: 2):

GTATTTTACAACAATTACCAACAACAACAACAACAGACAACATTACAATT
ACTATTTACAATTACAATGGCATAACACA
CAGACAGCTACCACATCAGCTTTGCTGGACACTGTCCGAGGAAACAACCTCT
TGGTCAATGATCTAGCAAAGCGTCGTCT
TTACGACACAGCGGTTGAAGAGTTTAACGCTCGTGACCGCAGGCCCAAGGTG
AACTTTTCAAAGTAATAAGCGAGGAGC
AGACGCTTATTGCTACCCGGGCGTATCCAGAATTCCAAATTACATTTTATAAC
ACGCAAAATGCCGTGCATTGCGTTGCA
GGTGGATTGCGATCTTTAGAACTGGAATATCTGATGATGCAAATTCCTACGG
ATCATTGACTTATGACATAGGCGGGAA
TTTTGCATCGCATCTGTTCAAGGGACGAGCATATGTACACTGCTGCATGCCCA
ACCTGGACGTTTCGAGACATCATGCGC
ACGAAGGCCAGAAAGACAGTATTGAACATACTTTCTAGGCTAGAGAGAGG
GGGAAAAACAGTCCCCAACTTCCAAAAG
GAAGCATTTGACAGATACGCAGAAATTCCTGAAGACGCTGTCTGCACAATA
CTTCCAGACATGCGAACATCAGCCGAT
GCAGCAATCAGGCAGAGTGTATGCCATTGCGCTACACAGCATATATGACATA
CCAGCCGATGAGTTCGGGGCGGCACTCT
TGAGGAAAAATGTCCATACGTGCTATGCCGCTTTCCACTTCTCCGAGAACCTG
CTTCTTGAAGATTCATGCGTCAATTTG
GACGAAATCAACGCGTGTTTTTCGCGCGATGGAGACAAGTTGACCTTTTCTTT
TGCATCAGAGAGTACTCTTAATTACTG
TCATAGTTATTCTAATATTCTTAAGTATGTGTGAAAACTTACTTCCCGGCCTC
TAATAGAGAGGTTTACATGAAGGAGT
TTTAGTCACCAGAGTTAATACCTGGTTTGTAAAGTTTCTAGAATAGATACTT
TTCTTTGTACAAAGGTGTGGCCCAT
AAAAGTGTAGATAGTGAGCAGTTTATATACTGCAATGGAAGACGCATGGCATT
ACAAAAAGACTCTTGCAATGTGCAACAG
CGAGAGAAATCCTCCTTGGGGATTTCATCATCAGTCAATTACTGGTTTCCAAAAA
TGAGGGATATGGTCATCGTACCATAT
TCGACATTTCTTTGGAGACTAGTAAGAGGACGCGCAAGGAAGTCTTAGTGTC
CAAGGATTTTCGTGTTACAGTGCCTAAC
CACATTCGAACATACCAGGCGAAAGCTCTTACATACGCAATGTTTGTCTT
CGTCGAATCGATTTCGATCGAGGGTAAT
CATTAAACGGTGTGACAGCGAGGTCCGAATGGGATGTGGACAAATCTTTGTTA
CAATCCTTGTCCATGACGTTTACCTGC
ATACTAAGCTTGGCGTTCTAAAGGATGACTTACTGATTAGCAAGTTTAGTCTC
GGTTCGAAAACGGTGTGCCAGCATGTG
TGGGATGAGATTTGCTGGCGTTTGGGAACGCATTTCCCTCCGTGAAAGAGA
GGCTCTTGAACAGGAACTTATCAGAGT
GGCAGGCGACGCATTAGAGATCAGGGTGCCTGATCTATATGTGACCTTCCAC
GACAGATTAGTGAAGTACAAGGCCT
CTGTGGACATGCCTGCGCTTGACATTAGGAAGAAGATGGAAGAAACGGAAGT
GATGTACAATGCACTTTCAGAATTATCG

Figure 6a

11/17

GTGTTAAGGGAGTCTGACAAATTCCGATGTTGATGTTTTTTCCAGATGTGCCA
ATCITTGGAAGTTGACCCAATGACGGC
AGCGAAGGTTATAGTCGCGGTCAAGCAATGAGAGCGGTCTGACTCTCACA
TTTGAACGACCTACTGAGGCGAATGTTG
CGCTAGCTTTACAGGATCAAGAGAAGGCTTCAGAAGGTGCATTGGTAGTTAC
CTCAAGAGAAGTTGAAGAACCGTCCATG
AAGGGTTCGATGGCCAGAGGAGAGTTACAATTAGCTGGTCTTGCTGGAGATC
ATCCGGAATCGTCTATTCTAAGAACGA
GGAGATAGAGTCTTAGAGCAGTTTCATATGGCGACGGCAGATTCGTTAATTC
GTAAGCAGATGAGCTCGATTGTGTACA
CGGGTCCGATTAAAGTTCAGCAAATGAAAACTTTATCGATAGCCTGGTAGC
ATCACTATCTGCTGCGGTGTCGAATCTC
GTCAAGATCCTCAAAGATACAGCTGCTATTGACCTTGAAACCCGTCAAAAGT
TTGGAGTCTTGATGTTGCATCTAGGAA
GTGGTTAATCAAACCAACGGCCAAGAGTCATGCATGGGGTGTGTTGAAACC
CACGCGAGGGAGTATCATGTGGCGCTTT
TGGAATATGATGAGCAGGGTGTGGTGACATGCGATGATTGGAGAAGAGTAGC
TGTTAGCTCTGAGTCTGTTGTTTATTC
GACATGGCGAAACTCAGAACTCTGCGCAGACTGCTTCGAAACGGAGAACCGC
ATGTCAGTAGCGCAAAGGTTGTTCTTGT
GGACGGAGTTCGGGGCTGTGGAAAAACCAAAGAAATCTTTCCAGGGTTAAT
TTTGATGAAGATCTAATTTTAGTACCTG
GGAAGCAAGCCGCGGAAATGATCAGAAGACGTGCGAATTCCTCAGGGATTAT
TGTGGCCACGAAGGACAACGTTAAAACC
GTTGATTCTTTCATGATGAATTTTGGGAAAAGCACACGCTGTGAGTTCAAGAG
GTTATTCAATTGATGAAGGGTTGATGTT
GCATACTGGTGTGTTAATTTTCTGTGGCGATGTCATTGTGCGAAATTGCAT
ATGTTTACGGAGACACAGCAGATTCT
CATACATCAATAGAGTTTCAGGATTCCCGTACCCCGCCCATTTTGCCAAATTG
GAAGTTGACGAGGTGGAGACACGCAGA
ACTACTCTCCGTTGTCCAGCGGATGTCACACATTATCTGAACAGGAGATATGA
GGGCTTTGTCATGAGCACTCTTCGGT
TAAAAAGTCTGTTTCGAGGAGATGGTCGGCGAGCCCGCTGATCAATCCG
ATCTCAAAACCCCTGTCATGGCAAGATCC
TGACTTTTACCAATCGGATAAAGAAGCTCTGCTTTCAAGAGGGTATTGAGAT
GTTACACTGTGATGAAGTGCAAGGC
GAGACATACTCTGATGTTTCACTAGTTAGGTTAACCCTACACCGGTCTCCAT
CATTGCAGGAGACAGCCCATGTTT
GGTCGCATTGTCAAGGCACACCTGTTGCTCAAGTACTACACTGTTGTTATGG
ATCCTTTAGTTAGTATCATTAGAGATC
TAGAGAAACTTAGCTCGTACTTGTAGATATGTATAAGGTGATGCAGGAAC
ACAATAGCAATTACAGATTGACTCGGTG
TTCAAAGGTTCCAATCTTTTGTGTCAGCGCCAAAGACTGGTGATATTTCTGA
TATGCAGTTTTACTATGATAAGTGTCT
CCCAGGCAACAGCACCATGATGAATAATTTTGATGCTGTTACCATGAGGTTG
ACTGACATTTCAATTGAATGTCAAAGATT

Figure 6b

12/17

GCATATTGGATATGTCTAAGTCTGTTGCTGCACCTAAGGATCAAATCAAACCA
CTAATACCTATGGTACGAACGGCGGCA
GAAATGCCACGCCAGACTGGACTATTGGAAAATTTAGTGGCGATGATTAAAA
GAAACTTTAACGCACCCGAGTTGTCTGG
CATCATTGATATTGAAAATACTGCATCTTTGGTTGTAGATAAGTTTTTTGATA
GTTATTTGCTTAAAGAAAAAGAAAAAC
CAAATAAAAAATGTTTCTTGTTCAGTAGAGAGTCTCTCAATAGATGGTTAGAA
AAGCAGGAACAGGTAACAATAGGCCAG
CTCGCAGATTTTGATTTTGTGGATTGCCAGCAGTTGATCAGTACAGACACAT
GATTAAAGCACACCCAAACAAAAGTT
GGACACTTCAATCCAAACGGAGTACCCGGCTTTCAGACGATTGTGTACCAT
TCAAAAAAGATCAATGCAATATTCGGCC
CGTTGTTTAGTGAGCTTACTAGGCAATTACTGGACAGTGTTGATTGAGCAGA
TTTTTGTTCACAAAGAAAGACACCA
GCGCAGATTGAGGATTTCTCGGAGATCTCGACAGTCATGTGCCGATGGATG
TCTTGGAGCTGGATATATCAAAATACGA
CAAATCTCAGAATGAATCCACTGTGCAGTAGAATACGAGATCTGGCGAAGA
TTGGGTTTCGAAGACTTCTTGGGAGAAG
TTTGAAACAGGGCATAGAAAGACCCTCAAGGATTATACCGCAGGTAT
AAAAACTTGCATCTGGTATCAAAGAAAAG
AGCGGGGACGTCACGACGTTTATTGGAAACACTGTGATCATTGCTGCATGTT
GGCCTCGATGCTTCGATGGAGAAAAT
AATCAAAGGAGCCTTTTGGCGTGACGATAGTCTGCTGTACTTTCCAAAGGGT
GTGAGTTTCCGGATGTGCAACACTCCG
CGAATCTTATGTGGAATTTTGAAGCAAAACTGTTTAAAAAACAGTATGGATA
CTTTTGGGAAGATATGTAATACATCAC
GACAGAGGATGCATTGTGTATTACGATCCCCTAAAGTTGATCTCGAAACTTG
GTGCTAAACACATCAAGGATTGGGAACA
CTTGGAGGAGTTCAGAAGGTCTCTTTGTGATGTTGCTGTTTCGTTGAACAATT
GTGCGTATTACACACAGTTGGACGACG
CTGTATGGGAGGTTTATAAGACCGCCCTCCAGGTTCTGTTTATAAAAGT
CTGGTGAAGTATTTGTCTGATAAAGTT
CTTTTAGAAGTTTGTATAGATGGCTCTAGTTGTTAAAGGAAAAGTGAATA
TCAATGAGTTTATCGACCTGACAAAAA
TGGAGAAGATCTTACCGTCGATGTTTACCCCTGTAAAGAGTGTATGTGTTCC
AAAGTTGATAAAATAATGGTTCATGAG
AATGAGTCAITGTCAGGGGTGAACCTTCTTAAAGGAGTTAAGCTTATTGATAG
TGGATACGTCTGTTTAGCCGGTTTGGT
CGTCACGGGCGAGTGGAACCTTGCCTGACAATTGCAGAGGAGGTGTGAGCGTG
TGTCTGGTGGACAAAAGGATGGAAAGAG
CCGACGAGGCCATTCTCGGATCTTACTACACAGCAGCTGCAAAGAAAAGATT
TCAGTTCAAGGTCGTTCCCAATTATGCT
ATAACCACCCAGGACGCGATGAGAAACGTCTGGCAAGTTTTAGTTAATATTA
GAAATGTGAAGATGTCAGCGGGTTTCTG
TCCGCTTCTCTGGAGTTTGTGTCGGTGTGTATGTTTATAGAAATAATATAA
AATTAGGTTTGAGAGAGAAGATTACAA

Figure 6c

13/17

ACGTGAGAGACGGAGGGCCCATGGAACCTACAGAAGAAGTCGTTGATGAGTT
CATGGAAGATGTCCCTATGTGATCAGG
CTTGCAAAGTTTCGATCTCGAACCGGAAAAAAGAGTGATGCCGCAAAGGGA
AAAATAGTAGTAGTGATCGGTCAAGTCC
GAACAAGAACTATAGAAATGTTAAGGATTTTGGAGGAATGAGTTTTAAAAAG
AATAATTTAATCGATGATGATTCGGAGG
CTACTGTCGCCGAATCGGATTCGTTTAAATAGATCTTACAGTATCACTACTC
CATCTCAGTTCGTGTTCTTGTCAtt
taaATGGCTAGCAAAGGAGAAGAAGCTTTCTACTGGAGTTGTCCCAATTCTTGTT
GAATTAGATGGTGATGTTAATGGGC
ACAAATTTTCTGTCACTGGAGAGGGTGAAGGTGATGCTACATACGGAAGCT
TACACTTAAATTTATTTGCACTACTGGA
AAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTCA
ATGCTTTTCCCGTTATCCGGATCATAT
GAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAA
CGCACTATATCTTTCAAAGATGACGGGA
ACTACAAGACCGGTGCTGAAGTCAAGTTTGAAGGTGATACCTTGTTAATCG
TATCGAGTTAAAAAGGTATTGATTTTAAA
GAAGATGGAAACATTCTCGGACACAAACTCGAGTACAACATAACTCACACA
ATGTATACATCACGGCAGACAAACAAAA
GAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCC
GTTCAACTAGCAGACCATTATCAACAAA
ATACTCCAATTGGCGATGGCCCTGTCTTTTACCAGACAACCATTACCTGTCTG
ACACAATCTGCCCTTTTCAAAGATCCC
AACGAAAAAGCGTGACCACATGGTCTCTTCTGAGTTTGTAACTGCTGCTGGGAT
TACACATGGCATGGATGAGCTCTACAA
ATAATGACACTCGAGGGGTAGTCAAGATGCATAATAAATAACGGATTGTGTC
CGTAATCACACGTGGTGCGTACGATAAC
GCATAGTGTTTTTCCCTCCACTTAAATCGAAGGGTTGTGTCTTGATCGCGCG
GGTCAAATGTATATGGTTCATATACAT
CCGCAGGCACGTAATAAAGCGAGGGGTTCCGGTTCGAGGTCCGGCTGTGAAACT
CGAAAAGGTTCCGGAAAAACAAAAAGAG
AGTGGTAGGTAATAGTGTTAATAATAAGAAAATAAATAGTGGAAGAAA
GGTTTGAAAGTTGAGGAAATTGAGGATA
ATGTAAGTGATGACGAGTCTATCGCGTCATCGAGTACGTTTAAATCAATATGC
CTTATACAATCAACTCTCCGAGCCAAT
TTGTTTACTTAAGTTCCGCTTATGCAGATCTGTGCAGCTGATCAATCTGTGT
ACAAATGCATTGGGTAACCAAGTTCAA
ACGCAACAAGCTAGGACAACAGTCCAACAGCAATTTCGGGATGCCTGGAAAC
CTGTGCCTAGTATGACAGTGAGATTTC
TGCATCGGATTCTATGTGTATAGATATAATTCGACGCTTGATCCGTTGATCA
CGGCGTTATTAAATAGCTTCGATACTA
GAAATAGAATAATAGAGGTTGATAATCAACCCGCACCGAATACTACTGAAAT
CGTTAACGCGACTCAGAGGGTAGACGAT
GCGACTGTAGCTATAAGGGCTCAATCAATAATTGGCTAATGAACGGTTTCG
TGGAAGTGGCATTGTTCAATCAAGCAAG

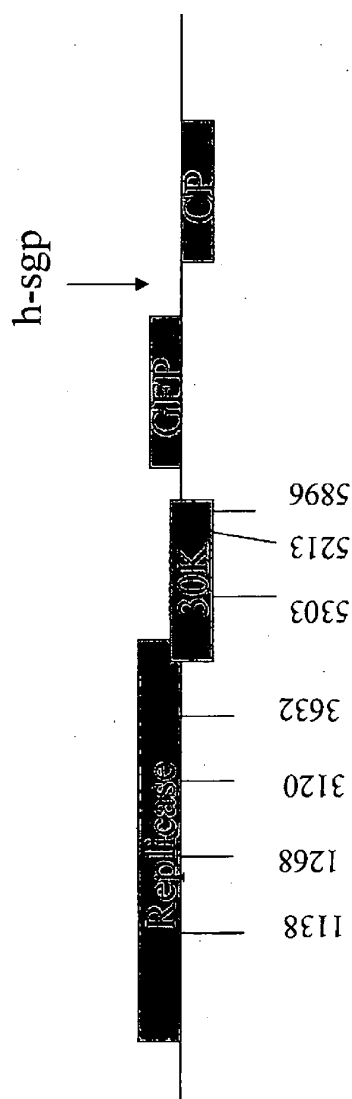
Figure 6d

14/17

CTTTGAGACTGCTAGTGGACTTGTCTGGACCACAACTCCGGCTACTTAGctattgtt
gtgagatttcctaaaataaagtc
actgaagacttaaaatcagggtggctgataccaaaatcagcagtggtgttcgtccacttaataaacgattgcata
tctggatccaacagttaaacatgtgatgggtatactgtggtatggcgtaaaacaacgaaaagtcgtgaagactfaa
aaatcagggtggctgataccaaaatcagcagtggtgttcgtccacttaaaaataacgattgtcatatctggatccaaca
gttaaacatgtgatgggtatactgtggtatggcgtaaaacaacggagaggttcgaatcccccctaaccgagggtagc
ggccca

Figure 6e

15/17



Schematic map of location of mutations in BSG 1057.
 30K = movement protein; GFP = green fluorescent protein;
 CP = coat protein. Nucleotide positions of BSG 1057 mutations
 are noted.

Figure 7



Figure 8

17/17